Characterization of the vasotocin neuropeptide hormone receptor system in the sea lamprey (*Petromyzon marinus*)

A Dissertation

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA

BY

Sally A. Mayasich

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Benjamin L. Clarke, Advisor

November 2015

Acknowledgements

There were several reasons for my return to school after many years as an environmental professional, including the usual financial and career improvement goals. But my main personal reason was my interest in understanding the evolution of behavior, which had led me to the oxytocin and vasopressin hormone systems. Without the aid, freedom and financial support I was afforded by my advisor, Dr. Ben Clarke, these studies would not have been possible on the Duluth campus (nor was any such opportunity available on the Twin Cities campuses). My deepest thanks to Ben for indulging my pursuit of new theories through a somewhat irregular animal model, as well as kindling my interest in the immune system. These studies were also made possible through the excellent visions of all those involved in the inception and operation of the Integrated Biosciences (IBS) graduate program, through which I received not only the knowledge and wisdom of numerous professors in coursework, seminars and hallway conversations, but also a one-year Graduate Research Assistantship, summer funding and travel support to present my research at two conferences.

I am deeply grateful to members of my committee: Drs. Grant Anderson, Jennifer Liang, Allen Mensinger, Teresa Rose-Hellekant, and Sara Zimmer. Special thanks to Grant for the Pharmacy School Teaching Assistantship; Al and the Biology Department for the Animal Physiology TA, facilities, and help with my "fish"; also fish help and molecular experience with Jennifer; Teri for lots conversations and decision help; and Sara for invaluable help reviewing my manuscripts and presentations.

Many others in the Biomedical Sciences Department provided much appreciated hands-on help, especially our Department Chair, Dr. Lynne Bemis. I also thank the labs of professors Jean Regal, Janet Fitzakerly, Pat Scott, Bob Cormier, Matt Slattery, Kevin Diebel, Kendra Nordgren, grad students Sara Prince, Kun Zhou; post-docs Katie Nemeth, Sarah Lacher and Aubie Shaw. Thanks to my lab mates for always being there to talk and help out: Techs Shannon RedBrook and Jeff Finlon; grad student Veronica Nelson Seifert; and fellow lamprey folk Terrence Wilcox and undergrad assistant Chris Osberg.

Financial support for the research in this dissertation was also provided by Minnesota Sea Grant, a Grant-in-Aid for Research, Artistry & Scholarship through the University of Minnesota Office of the Vice President for Research, and the Annette L. Boman Fellowship from the UMD Medical School. Travel grants came from the IBS program, the Department of Biomedical Sciences, the UMD Graduate School and the UMD Swenson College of Science and Engineering.

I write this paragraph less than 24 hours after the passing of Dr. Larry Wittmers, on April 25, 2015. Dr. Wittmers was invaluable to so many people over so many years at the

UMD Medical School through his gentle humor and incredibly patient teaching and mentoring. For me he was a “guardian angel” who always seemed to appear at the moment I needed something, whether large or small, from financial support, to a piece of equipment, or just a word of encouragement. He provided me with an unprecedented teaching assistantship in the Medical School. Between the two of us we were able to figure out how to keep the PowerLab working for the pulmonary physiology labs, although with his technology savvy he could easily have done it without my help. Dr. Wittmers also provided a great summer support and experience opportunity, employing me as an assistant in the High School All-Stars program. When I told him about my kids’ science projects he was always willing to help out with equipment, lab space and suggestions. These things show his commitment to science, research and education in many tangible ways, but I am most grateful for the conversations we had when I was frustrated with a failed experiment; it was the confidence he expressed through stories of his own research attempts, missteps and eventual success that kept me buoyed up and feeling poised to try again. I will miss and remember fondly his cheerful, comforting whistle as he walked down the hallways.

In spite of my age and assumed maturity, and my station as a professional, parent, and doctoral student, I have often during these last six years felt like a frustrated five-year-old learning to tie her shoelaces. I still remember my parents and grandparents patiently guiding my hands over, under around and through. When I could finally shout, “I did it myself!” I know they were proud, and understood that I wasn’t ungrateful for their help but that I was excited and maybe surprised at my new capacity to tie my laces solo. I thank not only family and all those mentioned during my time at UMD but many throughout my life who have gotten me through all the “solos” to this point: great teachers in my hometown of Virginia, MN, through high school and Mesabi Community College; Bemidji State profs and undergrad advisor the late Dr. Charles Fuchsman; University of Maryland Masters’ advisor Dr. Richard Smucker; mentors in the often intimidating world of work like Dr. Rich Henry and Dr. Gary Buchanan at Weston REAC in New Jersey, Tom Bader, Joe Renier and the late Dr. Linda Meyers-Schöne, at IT Corp and AMEC, and Dr. Mike Kierski at MWH; and the support and encouragement of many friends. They all gave me the best possible platform from which to launch from lace tying to cloning genes. Wow! All the help and mentorship I have received is immeasurable and un-repayable. I hope I have made them proud and have learned and am still learning to pay it forward.

Dedication

This dissertation is dedicated to my parents, Dick and Carol Okerstrom, my sister Sue and her family, my husband Joe and children Olivia and Spencer.

To Mom and Dad: Pretty simply, you made me what I am today. You gave me everything I needed and more, with love and patience.

To Sue: My big sister, you always led the way, thanks for your constant love and support!

To Joe: Thank you for your love and all your hard work. We have come a long way and have lots to be proud of, especially those two named below!

To Olivia and Spencer: You have already made us proud! I live for you as Grandma and Grandpa have lived for me. Take some hard roads—I'll always be right there with you.

In Memory,
Richard Aarre Okerstrom
April 1, 1929 - August 10, 2014

Abstract

The sea lamprey (*Petromyzon marinus*) is a jawless vertebrate at an evolutionary nexus between invertebrates and jawed vertebrates. Lampreys are known to possess the arginine vasotocin (AVT) hormone utilized by all non-mammalian vertebrates. We postulated that the lamprey would possess AVT receptor orthologs of predecessors to the arginine vasopressin (AVP)/oxytocin (OXT) family of G protein-coupled receptors found in mammals, providing insights into the early branching into the mammalian V1a, V1b, V2 and OXT receptors. We sequenced one partial and four full-length putative lamprey AVT receptor genes that are found on separate scaffolds in the *P. marinus* genome. Molecular phylogeny also utilizing the Japanese lamprey (*Lethenteron japonicum*) genome show that the lamprey receptors cluster with the larger V1a/V1b/OXTR and V2a/b/c clades but specific orthology is unclear. Synteny analysis supports the recently proposed one-round (1R) whole-genome duplication in the vertebrate lineage as the most likely scenario, but does not refute 2R or independent 3R scenarios.

The mRNA expression patterns were determined in 15 distinct tissues for these genes, showing transcription in tissues where function has been demonstrated in jawed vertebrates. The literature provides evidence of the expression of neuropeptide hormones and receptors in jawed vertebrate immune cells. For the first time, lamprey peripheral blood leukocytes were maintained in primary culture for periods of at least six days, in which mRNA transcription of a V1a/OXTR-like gene (lamprey AVT receptor Pm807) was demonstrated. These preliminary results also support the hypothesis that neuropeptide hormones may play a role in response to pathogenic challenge in all vertebrates.

The possibility of AVT involvement in mediating pheromone release from glandular cells in the gills of mature male lampreys was tested. The compound petromyzonamine disulfate (PADS) was detected at higher quantities after than before injection from several AVT and OXT injected males, but this was not true for the main sex pheromone component 3-keto petromyzonol sulfate (3kPZS).

The question of whether DNA methylation of cytosine-guanine (CpG) dinucleotides function to regulate lamprey gene transcription was addressed through analysis of CpG islands in the lamprey Pm807 V1a/OXT receptor gene promoter region. Using High Resolution Melt (HRM) PCR on bisulfite-converted DNA, we pinpointed a region with tissue-specific differences in DNA melt characteristics, indicating differences in methylation level. Sequencing revealed a pattern of methylation at specific CpGs at consistently higher levels in adult heart and larval liver, where Pm807 is transcribed to mRNA, than in adult liver where Pm807 is not transcribed. The methylated CpGs are associated with putative Krüppel-like factor (KLF) 4 transcription factor binding site sequences. In humans KLF4 binds to methylated DNA to initiate transcription. The results suggest that CpG methylation regulates lamprey gene transcription. Additional Pm807 putative promoter elements such as estrogen response element consensus binding sequences were found to be organized similarly to functional OXTR promoters in mammals. The results of my research support the hypothesis that, similarly to jawed vertebrates, differential mRNA expression and resultant functional pleitropy is generated through promoter region sequence and epigenetic regulatory elements in the jawless basal vertebrate lamprey.

Table of Contents

List of Tables

List of Figures

Chapter 1

General Introduction:
Background, Hypothesis and Aims

1.1 Sea lamprey biology

1.1.1 Life history and reproduction

Sea lampreys (*Petromyzon marinus*) are a natural species in ocean and estuarine ecosystems, however they are a severe problem in the Great Lakes fisheries, having been accidentally introduced when canals were opened to bypass Niagara Falls. They hatch in freshwater in both ocean and lake systems, and the filter-feeding ammocoete larvae (Fig. 1-1) live in stream sediment for approximately five to seven years (Sower et al., 2009). Metamorphosis to the parasitic juvenile stage takes approximately 10 to 11 months without feeding and is characterized by major physiological and morphological changes.

Following metamorphosis, lampreys migrating from streams into the Great Lakes spend 12 to 18 months in the parasitic phase, attaching to and extracting fluids from fish such as lake trout. Lampreys then return to streams to spawn and then die, with genetic assessment of parentage suggesting that the mating system includes both polyandry and polygyny in wild populations (Jones, 2007). Larval migratory pheromones in the form of bile acids attract spawning adults to these streams, acting as a signal of suitable spawning habitat (Polkinghorne et al., 2001; Sorensen et al., 2003). These bile acids, petromyzonol sulfate (PZS) and allocholic acid (ACA) are present in the livers and gall bladders of ammocoetes and are believed to be released to the environment through defecation (Polkinghorne et al., 2001). Larval anatomy is shown in Fig. 1-2.

Figure 1-1. A sea lamprey (*Petromyzon marinus*) ammocoete (larva). The tail of a burrowed larva is visible at center (photo by author).

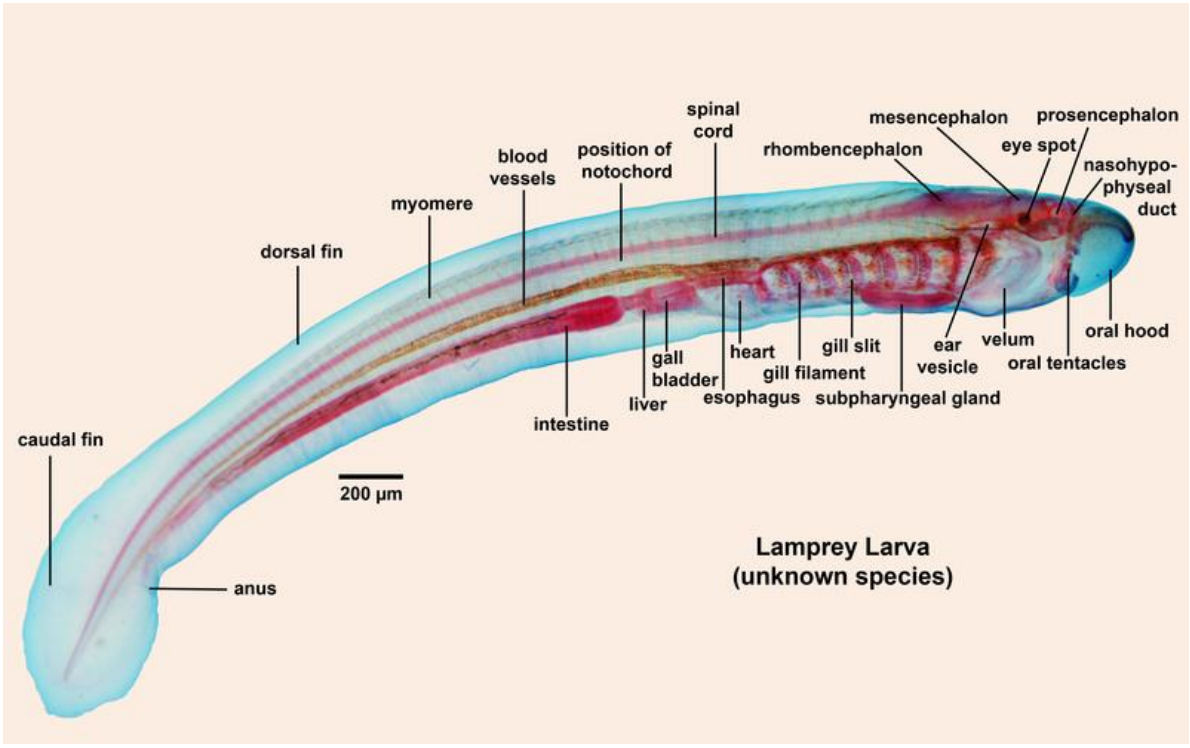


Figure 1-2. Microscopic whole mount larval specimen from an unknown lamprey (Petromyzontiformes) species, cleared and stained. From Wikimedia Commons, the free media repository. This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Author: Tracyanne, January 2012. https://commons.wikimedia.org/wiki/File:Lamprey_larva_labelled.png

Males build u-shaped rock nests, and when they begin spermiating release a mating pheromone from cells in their gills that induces search and preference behaviors in ovulating females (Siefkes et al., 2003). The male sex pheromone is also bile acid-based. The main component is known as 3-keto petromyzonol sulfate (3kPZS) and is related to the migratory pheromone released by ammocoetes (Siefkes et al., 2003). Female response to the male pheromone as the synthesized component 3kPZS when released into a stream triggers upstream swimming in ovulated females at concentrations as low as 10^{-14} M (moles/liter) (Johnson et al., 2009). A male secondary sexual characteristic develops in front of the anterior dorsal fin of mature males as a swollen ridge known as the 'rope' tissue. This tissue was found to be thermogenic brown adipose tissue that produces heat during sexual encounters (Chung-Davidson et al., 2013).

In the spawning choreography of lamprey mating pairs, a male attaches to the female's head, wraps his tail tightly around her trunk and both partners undulate rapidly. The motion and squeezing cause the ova to be expelled from the female cloaca. Fertilization takes place externally within the nest. Alternative mating tactics including smaller sneaker males invading the nest are known for several lamprey species (Hume et al., 2013). These reproductive behaviors appear to be the only adult conspecific social interactions in which the lampreys engage, and they appear to be mediated over long ranges only by olfactory sensing as eyesight plays no role in upstream migration (Binder and McDonald, 2007). Interestingly, lampreys' short-range sensing and retention on the nest during mating may involve response to weak electric fields, which have been shown to increase gonadotropin releasing hormone (GnRH-I) expression in the pre-optic area of

adult male lamprey brains (Chung-Davidson et al., 2008). Several endocrine hormones are involved in lamprey reproductive function and secondary sexual traits.

1.1.2 Steroid and neuroendocrine hormone systems

Many of the actions and behaviors attributed to steroid hormones actually occur through their role as transcription factors regulating neuropeptide genes. Estrogen, testosterone, progesterone and glucocorticoids are steroid hormones that when bound to their nuclear receptors are known to interact with the promoter DNA of the various neuroendocrine hormone genes and receptor genes in mammals, where they may act as either activators or repressors of transcription (Dhakar et al., 2013).

Lamprey steroid hormones have been reviewed (Bryan et al., 2008). The lamprey is the earliest vertebrate to diverge from the ancestral vertebrate lineage to utilize sex steroids and their receptors in a hormonal capacity similarly to jawed vertebrates. There are three known sex steroids believed to operate as hormones, including the classical estradiol used by both sexes, 15α-hydroxyprogesterone, and androstenedione as an androgen. Although 15α-hydroxytestosterone is produced, it is not clear whether this is a functional hormone (Bryan et al., 2008). The estradiol appears to activate a jawed-vertebrate-like nuclear estrogen receptor in lampreys and bind to consensus DNA estrogen response element (ERE) sequences (Paris et al., 2008).

Neuropeptide hormones are signaling molecules derived from pleiotropic genes, influencing phenotypic traits in multiple systems. These small proteins cleaved from larger precursors not only have important effects on homeostasis, reproductive function, secondary sexual traits and behavior but also on immune system function. There are 20

conservatively defined neuropeptide hormone families with 70 hormone genes in humans, and additional candidate neuropeptides (Burbach, 2010). Brain and neurohypophysial (pituitary-secreted neuropeptide) hormones studied in the lamprey include gonadotropin releasing hormone (GnRH), somatostatin-14, neuropeptide-Y, tachykinin, a melanocortin hormone system (Sower, 1998) and growth hormone/prolactin (GH/PRL) (Ellens et al., 2013) representing six families. The GnRH hormone receptor system is well characterized in lampreys (e.g., Silver et al., 2005; Kavanaugh et al., 2008; Joseph et al., 2012; Sower et al., 2012). The hypothalamic-pituitary-gonadal (HPG) and –thyroid (HPT) axes in the lamprey have also been reviewed (Sower et al., 2009).

Generally, jawed vertebrates (gnathostomes) have one or two GnRHs that act as hypothalamic hormones and two pituitary gonadotropins (luteinizing hormone and follicle-stimulating hormone) that stimulate the production of steroid sex hormones and ultimately the maturation of eggs and sperm. In the lamprey there are three hypothalamic GnRHs and three GnRH receptors, but only one known pituitary gonadotropin (gonadotropic hormone β) (Sower et al., 2009; Hall et al., 2013). *In situ* hybridization studies localizing estrogen receptor mRNA expression in the hypothalamus demonstrated its close proximity to GnRH expression areas in the lamprey brain, but the nature of the potential interaction has not been determined (Sower and Baron, 2011).

The hypothalamic-pituitary-adrenal/interrenal axis (HPA/HPI) has also been studied, with the findings that 11-deoxycortisol is the active stress hormone in lamprey plasma, released through a corticosteroid releasing hormone (CRH)-adrenocorticotropic hormone (ACTH) pathway. Another similar glucocorticoid compound in lamprey plasma

circulation, 11-deoxycorticosterone, was found to be non-functional. The functional 11-deoxycortisol binds a highly-specific nuclear corticoid receptor (Close et al., 2010).

Melanocortin receptors have been characterized but only two are present compared to the larger complement in jawed vertebrates. Also in contrast to jawed vertebrates, neither receptor is found centrally in the brain: the MC-a is found in skin and MC-b in skin, liver and heart. Therefore, the binding of these receptors to ACTH was weak, and the researchers suggest that the central nervous system functions evolved later, or there may be additional MC receptors that they were unable to isolate (Haitina et al., 2007).

Arginine vasotocin (AVT) is the non-mammalian vertebrate ortholog of the mammalian neurohypophysial hormone arginine vasopressin (AVP), described in more detail in the following section. Very little is known about AVT in lampreys. The AVT precursor gene has been sequenced in the Japanese lamprey (*Lampetra japonica*) (Gwee et al., 2009), but the receptor or receptors have not been determined or sequenced in jawless vertebrates. The only other study on the AVT system in lampreys is a report of AVT-dose-dependent vascular constriction leading to increased glomerular filtration rate and diuresis in kidneys of Japanese lampreys (Uchiyama and Murakami, 1994). The dose-response relationship was also measured for dorsal aortic blood pressure, and blocked by an AVT receptor antagonist that worked similarly on fish and bullfrog AVT receptors. The authors postulated that an AVT receptor in the smooth muscles and vasculature of the urinary tract mediated the response. My aim is to remedy this knowledge gap for the important and surprisingly neglected AVT hormone receptor system in the sea lamprey.

1.1.3 Genomics and molecular evolution

The sea lamprey genome was sequenced from the liver tissue of a wild-caught single female adult over a period extending from 2004 until publication in 2013 (University of Oklahoma, 2013; Smith et al., 2013). Starkly different structures and patterns were found in the lamprey compared with both invertebrate and vertebrate genomes. High guanine-cytosine (GC) content made sequencing difficult due to secondary structure formation, and a high degree of repetitive elements encumbered the construction of scaffolds, a precursor to assembly of whole chromosomes. The currently existing scaffolds (Ensembl, http://useast.ensembl.org/Petromyzon_marinus/Info/Index?db=core) often contain only one gene, making synteny analysis of neighboring genes, a method of determining gene relatedness (orthology) among species, impossible for some gene families. The genome is highly heterozygous (Smith et al., 2013) as originating from a wild specimen of a species that engages in multiple matings. More recently, new techniques have been developed that overcome the problems of assembling heterozygous genomes now that more wild species are being sequenced (Kajitani et al., 2014; Bodily et al., 2015). Additionally, lampreys undergo the unique programmed loss of ~20% of their germline DNA from somatic cells during early embryogenesis (Smith et al., 2012).

At the time of the genome publication in 2013, there was considerable support for the idea proposed by Ohno (1970) that vertebrates have undergone two rounds of whole-genome duplication (WGD), resulting often in a four-to-one ratio of copies of many genes present in vertebrates compared to invertebrates. The lamprey genome sequencing

results appeared to provide an answer to the prevailing question on the timing of these duplications in the vertebrate lineage, suggesting that two rounds of duplication (the "2R hypothesis") took place prior to the divergence of the ancestral lamprey from the jawed vertebrates. However, Smith and Keinath (2015) recently constructed the first lamprey meiotic map by crossing a single male with a single female wild-captured *P. marinus*. This procedure provided better resolution of chromosome-scale structure via parent-specific and parent-averaged linkage maps (Smith and Keinath, 2015). Their results through the linkage of more than 5,000 markers, in linkage groups containing at least 10 markers each, are significant in providing robust support for a single WGD in all vertebrate lineages, and only weak support for the 2R hypothesis.

Because of the lamprey's position as the earliest-diverged extant species in the vertebrate lineage, each family of genes characterized in lampreys provides more clarification of vertebrate evolution. My aim is to contribute to resolution of the WGD issue through characterizing the multiple paralogs of the AVT receptor family.

1.1.4 Immune system

All lampreys are equipped with proto-adaptive immunity consisting of "lymphocytes" (i.e. anticipatory immune cells) expressing higher vertebrate T- and B-cell-like properties and antigen-specific receptor molecules. The receptor molecules in lampreys are called variable lymphocyte receptors (VLR), and are not based on immunoglobulins as are the antibodies in jawed vertebrates, but have leucine-rich repeat domains similarly to toll-like receptors found in innate immune cells. The gene locus for variable lymphocyte receptors is arranged similarly to the antibody locus in jawed

vertebrates. The lamprey lymphocyte can select specific gene segments for assembling a functional antigen receptor using somatic recombination. Recombination of gene segments can produce an enormous repertoire of unique antigen-specific receptors (e.g. greater than 10^{14} recombinants) to recognize various pathogens (Alder et al., 2008).

The two germline VLR genes, VLRA and VLRB have been identified, producing separate lymphocyte populations. VLRA is a membrane-bound receptor similar to T cell receptors, and VLRB "antibodies" are excreted from lymphocytes similar to B cells (Guo, et al., 2009). Larval lampreys injected with *Bacillus anthracis* (anthrax) spores have shown a more rapid immune response after a second challenge, producing VLRs specific to an anthrax surface protein (Alder et al., 2008). This indicates a potential adaptive immune response.

Lamprey immune cells are clearly different in their immune-specific receptor composition from those of jawed vertebrates. A characteristic of jawed vertebrate immune systems is their interaction with neuroendocrine hormones, as discussed in the following section. My aim is to demonstrate the expression of neuropeptide hormones and receptors in lamprey immune cells, as a potential fitness-determining, evolutionarily conserved characteristic among vertebrates.

1.2 The vasopressin/oxytocin hormone systems in vertebrates

1.2.1 Hormone distribution among species

There are several reviews that discuss the origins of mammalian nine-amino-acid neuropeptide hormones arginine vasopressin (AVP) and oxytocin (OXT) (Donaldson, 2008; Goodson, 2005; Goodson and Kabelik, 2009). Homologous compounds exist in species from invertebrates through humans (Table 1-1). The OXT homolog in fish is isotocin (IT), and that in birds, reptiles and amphibians is called mesotocin (MT). The invertebrate homologs have existed for about 700 million years. Arginine vasopressin was first known as the antidiuretic hormone, or ADH, and has an essential role in the renal system to maintain osmotic balance (Sladek and Somponpun, 2008).

Table 1-1. Arginine vasopressin- and oxytocin-like hormone amino acid sequences (Modified from Stafflinger, et al., 2008).

Arginine vasopressin (AVP)	**CYFQNCPRG**	Mammals
Lys-Vasopressin	**CYFQNCPKG**	Pig, some marsupials
Phenypressin	**CFFQNCPRG**	Some marsupials
Insect oxytocin/vasopressin-like peptide (inotocin)	**CLITNCPRG**	*Locusta migratoria* (locust) *Tribolium castaneum* (beetle) *Nasonia vitripennis* (wasp)
Crustacean oxytocin/vasopressin-like peptide	**CFITNCPPG**	*Daphnia pulex* (water flea)
Arginine vasotocin (AVT)	**CYIQNCPRG**	Nonmammalian vertebrates
Arg-conopressin	**CIIRNCPRG**	*Conus geographicus* (sea snail)
Lys-conopressin	**CFIRNCPKG**	Leech, various mollusks
Oxytocin (OXT)	**CYIQNCPLG**	Mammals
Isotocin (IT)	**CYISNCPIG**	Fish
Mesotocin (MT)	**CYIQNCPIG**	Birds, reptiles, amphibians
Annetocin	**CFVRNCPTG**	Annelids
Cephalotocin	**CYFRNCPIG**	*Octopus vulgaris*
Octopressin	**CFWTSCPIG**	*Octopus vulgaris*

In fish not only the kidneys but also the gills express AVT and receptors and are a major site of osmoregulation (Balment et al., 2006). Balment et al. (2006) reviewed the AVT system in fish physiology, discussing its major roles in behavior, metabolism, stress

response, and cardiovascular responses as well as osmoregulation. Comparing to mammalian models, fish exhibited basic similarities of AVT function such as the stimulation of pituitary adrenocorticotropic hormone (ACTH) release in response to stress, glycogenolysis activation in the liver, induction of constriction in the gill, heart and branchial vasculature, and modulation of behaviors including courtship and vocalization. Balment et al. (2006) also pointed out that these responses in fish appear subject to circadian and seasonal rhythms due to AVT interaction with melatonin as has been established for AVP/melatonin in mammals.

Gwee et al. (2009) provide evidence that the AVP- and OXT-family of neurohypophysial hormones evolved in a common ancestor of jawed vertebrates through tandem duplication of the ancestral AVT gene. The AVP/AVT and OXT/MT/IT precursor molecules consist of a signal peptide, the nine-amino-acid (nonapeptide) hormone, a glycine-lysine-arginine processing signal, neurophysin and a copeptin or leucine-rich core segment (Searcy et al., 2009; Warne et al., 2000; and Stafflinger et al. 2008). The AVT nonapeptide differs by only one amino acid from mammalian AVP with isoleucine instead of phenylalanine in the third position and differs from mammalian OXT only in the eighth position with an arginine instead of leucine (Table 1-1). This confers a combination position between the AVP and OXT hormones which belies their duplication and overlapping receptor specificity. All known AVP- and OXT-like hormones have cysteines in the first and sixth positions creating a disulfide bridge and making the molecule somewhat cyclic (Table 1-1; Stafflinger et al., 2008).

1.2.2 Receptor types and expression

G-protein-coupled receptors (GPCRs) are proteins inserted in the cell membrane with an extracellular amino (N)-terminus, seven trans-membrane passages, and an intracellular carboxy (C)-terminal tail. In mammals, three AVP receptor types and one OXT receptor, members of the same subfamily of the Rhodopsin-like family of GPCRs, have been identified and characterized. The AVP receptors are classified as V1a, V1b (also known as V3), and V2 (Decaux et al., 2008). V1a-type receptors have been found in all vertebrates studied, including fish. V2-type receptors have been identified in ray- and lobe-finned fish (Lema, 2010; Konno et al., 2009 and 2010), providing knowledge that V2-type receptors involved in osmotic balance, thought previously to have arisen in tetrapods, are in fact present in teleost fish. A V1b receptor has been identified in amphibian newt species (Hasunuma et al., 2007), so is believed to have arisen in the tetrapods.

The pattern of expression in specific organs is a clue to predicting receptor functional properties and pathways. In mammals, the V1a receptor is wide-spread among several organs and mediates a variety of functions including vascular smooth muscle constriction; memory, stress adaptation, social recognition, and the regulation of temperature, blood pressure and heart rate in the brain; glycogenolysis in the liver; renal prostaglandin synthesis and blood flow regulation; and stimulation of aldosterone and cortisol secretion in the adrenal gland. The V1b receptor is more limited in its distribution, mediating ACTH and β-endorphin release in the anterior pituitary, and insulin release in the pancreas. The classical antidiuretic function in the renal collecting

tubules of the kidney is mediated by the V2 receptor via trafficking of aquaporin-2 (AQP2) water channel proteins; the V2 receptor also stimulates sodium resorption in lung cells and vasodilation in vascular smooth muscle (Decaux et al., 2008).

1.2.3 Intracellular signaling pathways

The ultimate function of the hormone-receptor interaction is brought about through several intermediary steps along intracellular pathways. V1a and V1b receptors couple with the Gq/11 protein to activate phospholipase C (PLC) which in turn hydrolyzes phosphatidylinositol to inositol triphosphate (IP3) then releasing intracellular calcium, and 1,2,-diacylglycerol (DAG), DAG then activating protein kinase C (PKC). The V2 receptor activates the stimulatory G protein (Gs) to stimulate adenylyl cyclase (AC) to increase cAMP, in turn activating protein kinase A (PKA) (Birnbaumer, 2000; Decaux et al., 2008). There is evidence that more than one pathway may be utilized by a receptor type. Thibonnier et al. (1997) showed that the human V1b (V3) receptor in the pituitary could utilize either the PLC/IP3/DAG/PKC or the AC/cAMP/ PKA pathway, depending on the density of receptor expression in the cell. This is evidence that the pathway and ultimate function these receptors mediate may be governed by level of expression in addition to receptor-specific conformation determining G-protein specificity and downstream events.

Oxytocin receptor activation by its cognate ligand results in coupling to either the Gq protein leading to calcium release (Gimpl and Fahrenholz, 2001) or the Gi/o protein, leading to MAP kinase pathway signaling through ERK 1/2 and p38 proteins (Mamrut et al., 2013).

1.2.4 Neuroendocrine-immune system interactions

Neuroendocrine hormones and receptors are expressed and function in immune cells and immune system tissues. Receptors for gonadotropin releasing hormone (GnRH) (Tanriverdi et al., 2004), OXT (Ndiaye et al., 2008), and AVP (Hu et al., 2003) have all been found on human lymphocytes and melanocortin-5 receptor has been isolated from rat lymphocytes (Akbulut et al., 2001). Studies in cows suggest that steroids in the reproductive tract regulate the expression of OXT receptors in lymphocytes, and the corpus luteum in the ovary secretes OXT that binds to the OXT receptors to regulate lymphocyte activity (Ndiaye et al., 2008). Gut microbes (lactic acid bacteria) have been found to up-regulate oxytocin circulating in plasma via communication through the vagus nerve to the brain. Oxytocin then facilitates wound healing by stimulating regulatory T-lymphocytes and trafficking these cells to the wound (Poutahidis et al., 2013). Hu et al. (2003) present evidence that AVP is expressed in immune tissues such as thymus and spleen, and AVP receptors have been detected in the rat, mouse, and human lymphocytes. AVP has long been known to be part of a complicated neuroendocrine-immune network, and has been implicated in lymphocyte activation, primary antibody responses, and stress-induced immune suppression. AVP receptors control signal transduction thresholds in lymphocytes, reflecting an important role for AVP in both humoral immune regulation and in stress-related immune responses (Hu et al., 2003). These studies show that OXT, AVP and other neuroendocrine hormones play important roles in regulating vertebrate immune systems.

1.2.5 Behavior and secondary sexual traits

Neuropeptides have been linked to the development of secondary sexual traits as well as behaviors that organisms engage in to choose mates and produce offspring in numerous species (Donaldson and Young, 2008; Goodson and Kabelik, 2009). In addition to ornamentation and competitive or other sexually-dimorphic structures, behavioral traits that include vocalizations, pheromone release or scent marking, dances, and aggressive displays are also considered secondary sexual traits. Some of these behaviors could more accurately be described as reflexive movements, such as penile erection, lordosis and copulation in mammals, copulation in birds, amplexis and spermatophore deposition in amphibians, and spawning reflex in fish (Goodson and Bass, 2001).

According to Goodson and Bass (2001), OXT and AVP are at the core of anatomical structure-behavior relationships across vertebrates. This is significant in indicating that specific patterns of behavior, regardless of species, occur due to the patterns and density of neuropeptide hormone receptor expression in specific brain areas. This patterning is dependent on a complex set of interacting factors including the actions of steroid hormones and other regulatory molecules that bind to the promoter, or "on-off switch" region of the gene.

1.2.6 Gene Regulation

Slight changes in the promoter regions, among individuals or between generations, as well as dynamic epigenetic conditions within an organism can change the location, timing and magnitude of neuropeptide hormone receptor expression.

Prior to the sequencing of the human genome in 2000, it was expected that humans had over 100,000 genes, each with a distinct function. In actuality, we have only about 25,000-30,000 genes found to be transcribed variously and produce proteins that perform differently in different parts of the body and at different stages of development. The key to this choreography is the mechanism by which expression of the genes is turned on and off. Studies have focused on the evolution of different body plans among species during embryonic development ("Evo Devo"; Carroll, 2005a).

Regulatory regions of non-coding DNA that act within a gene are called cis-operating elements, or from a distant gene are called trans-operating elements (Prud'homme et al. 2007). These elements are usually DNA sequences that provide binding sites for the on-off switch proteins, called transcription factors. Cis-regulation is more prevalent, thus the phenomenon of body-plan structural change over time due to mutations in the promoter DNA upstream of the transcription start site is generally referred to as cis-regulatory evolution. Cis-regulatory element mutations do not change the ultimate structure of the protein but only when, where and how much it is expressed, by altering transcription factor binding sites (Carroll, 2005a and 2005b).

An important objective of this dissertation is to provide evidence that cis-regulation not only applies to structural protein genes during development but also to signaling protein genes throughout the life of the organism. Therefore, while cis-regulatory element mutations in neuropeptide receptor genes may manifest in less obvious phenotypes, such as modifications in behavior and cellular function rather than body plan structure, the impact in evolutionary terms is no less profound.

Small differences in the promoter regions of neuropeptide and receptor genes, between species and individuals within species, are known to result in differences in expression. These cis-regulatory-associated changes include single-nucleotide polymorphisms and microsatellites, which are simple sequence repeats of often two or three nucleotides (Fondon et al., 2008). Neuropeptide receptors can be expressed in such a way as to control many conserved functions while allowing for evolution and diversity of behavior among species and individuals, as illustrated in Figure 1-3.

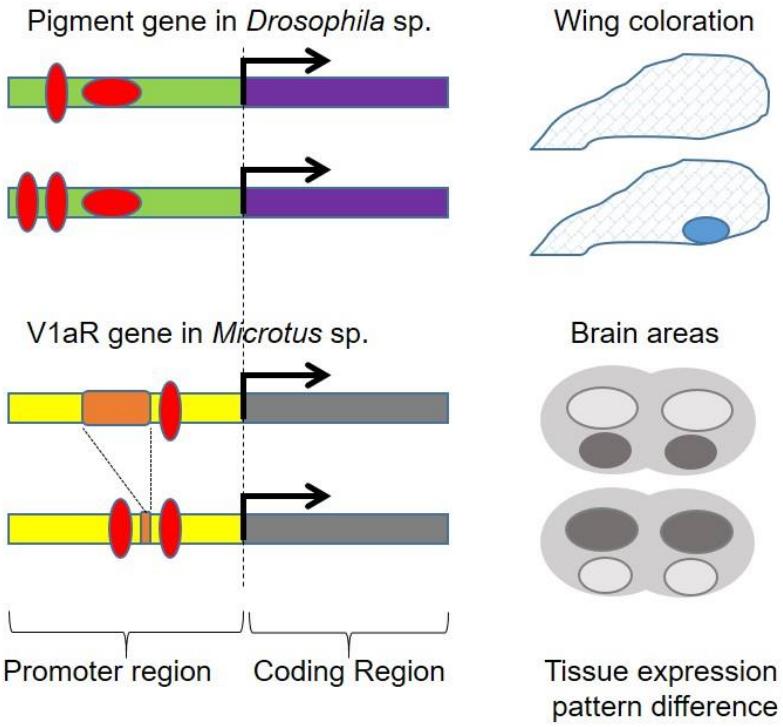


Figure 1-3. The arrangement of cis-regulatory elements in promoter region DNA affects how transcription factors bind (ovals depict binding of transcription factors to turn on a wing spot pigment gene in a *Drosophila* species wing, based on Carroll, 2005b). This results in visibly different expression patterns of a protein produced by the same gene in different species. Similarly, microsatellite sequences (rectangles) in the promoter of vole (*Microtus*) species may dictate variation in transcription factor (ovals) binding in the vasopressin V1a receptor promoter. This results in differential exression of the receptor in different brain areas, and ultimately differences in mating behavior (based on Nair and Young, 2006). The dotted line denotes the transcription start sites, with the promoter non-coding DNA to the left and protein coding region to the right.

1.3 Hypothesis and aims

1.3.1 Objective and central hypothesis

There have been no studies to characterize the AVP/OXT-family receptors in the earliest known, still extant species to diverge from the vertebrate lineage. This constitutes a gap in our evolutionary knowledge of an important hormone-receptor system tied to key physiological functions. Because the sea lamprey is a destructive invasive species in the Great Lakes fisheries, these studies could also provide a potential basis for developing a mechanism to control the invasive lamprey population. The ***objective*** of my research was to characterize the sea lamprey AVT hormone-receptor system, to provide a knowledge base for both vertebrate evolution and invasive species control.

My central hypothesis is that the molecular structures of genes encoding the AVT neuropeptide hormone and its receptors in the sea lamprey are evolutionarily conserved across vertebrate phyla, and are under the control of conserved cis-regulatory elements that allow selection of pleiotropically-linked traits. This concept, with the sea lamprey as a basal vertebrate model organism, is diagrammed in Figure 1-4. Each specific aim is addressed in the following chapters, and outlined below.

1.3.2 Specific aims

Aim 1 (Chapter 2): Determine the genetic molecular structure, phylogeny, synteny and tissue expression distribution patterns of all putative AVT hormone receptors in the sea lamprey, providing insights into the early branching into the mammalian V1a, V1b, V2 and OXT receptors.

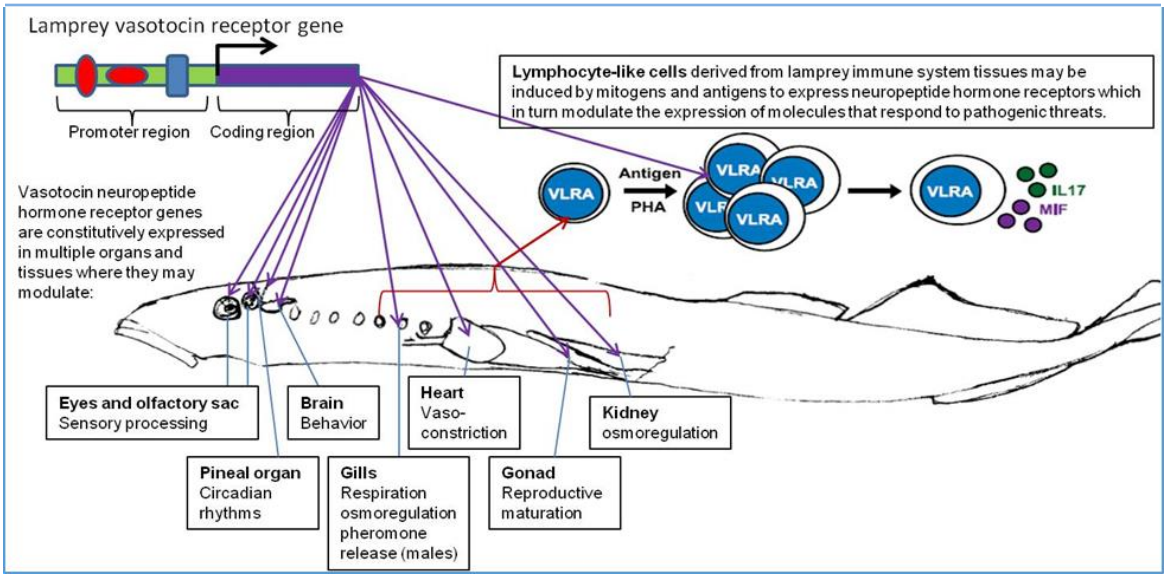


Figure 1-4. Multiple-system expression of a pleiotropic vasotocin receptor gene in the sea lamprey (*Petromyzon marinus*) predicts its role in fitness-determining traits homologous to those in higher vertebrates.

Aim 2 (Chapter 3): Clone the full-length coding sequences for expression of the receptor genes in mammalian cell lines. The ultimate goal is to demonstrate the ability of the proteins to trigger intracellular calcium and/or cAMP increases in a ligand concentration-dependent manner.

Aim 3 (Chapter 4): Determine whether lamprey immune cells express neuropeptide hormones and AVT receptors, supporting the probability that neuroendocrine-immune interaction is observed across vertebrate phyla.

Aim 4 (Chapter 5): Determine whether sex pheromone release as a secondary sexual trait or mating signal is mediated by the AVT hormone in sea lampreys and can be decreased or blocked by a mammalian AVP receptor antagonist.

Aim 5 (Chapter 6): Determine whether differential DNA methylation of cytosine-guanine (CpG) dinucleotides in the promoter region of a lamprey AVT receptor gene is

correlated with life-stage and tissue-specific differences in mRNA expression, potentially in conjunction with regulatory elements homologous to those known to control gene transcription in other species.

Genetic regulatory elements such as transcription factor binding sites or CpG sites, when in the vicinity of microsatellites or single nucleotide polymorphisms, provide the tinker toys for cis-regulatory evolution. Ultimately, small changes could fine-tune all of the systems illustrated in Figure 1-4 to result in selective advantages in reproductive success and fitness. In the research described in the following chapters I use the sea lamprey as a model organism and AVT receptors as a model system to explore the idea that neuropeptide hormones and receptors link the natural selection of important fitness traits. The knowledge gained may be of benefit in human medical applications as well as native wildlife and domesticated animal health and productivity, while exposing targetable vulnerabilities for the control of invasive species.

Chapter 2

The emergence of the vasopressin and oxytocin hormone receptor gene family lineage: clues from the characterization of vasotocin receptors in the sea lamprey (*Petromyzon marinus*)

This chapter has been submitted to and reviewed for publication in the journal *General and Comparative Endocrinology*, (S. Mayasich and B.L. Clarke, authors) and is currently in revision.

The sea lamprey (*Petromyzon marinus*) is a jawless vertebrate at an evolutionary nexus between invertebrates and jawed vertebrates. Lampreys are known to possess the arginine vasotocin (AVT) hormone utilized by all non-mammalian vertebrates. We postulated that the lamprey would possess AVT receptor orthologs of predecessors to the arginine vasopressin (AVP)/oxytocin (OXT) family of G protein-coupled receptors found in mammals, providing insights into the early branching into the mammalian V1a, V1b, V2 and OXT receptors. The earliest vertebrate in which these receptors are characterized is the jawed, cartilaginous elephant shark, which has genes orthologous to all four mammalian receptor types. Therefore, our work was aimed at helping resolve the critical gap concerning the number of large-scale (whole-genome) duplication (WGD) events in the vertebrate lineage. We sequenced one partial and four full-length putative lamprey AVT receptor genes. We also determined mRNA expression patterns in 15 distinct tissues for these genes. Three of the full-coding genes possess structural characteristics of the V1 branch containing the V1a, V1b and OXT receptors, two of which appear to be products of a lamprey-specific duplication at a V1/OXT sub-branch based on phylogenetic and syntenic analyses. Another full-length coding gene and the partial sequence are part of the V2 branch and appear to be most closely related to the newly established V2b and V2c types. Our synteny analysis also utilizing the Japanese lamprey (*Lethenteron japonicum*) genome supports the recently proposed one-round (1R) WGD in the vertebrate lineage as the most likely scenario, but does not refute 2R or independent 3R scenarios.

2.1 Introduction

The sea lamprey (*Petromyzon marinus*) is a jawless (agnathan) vertebrate that emerged approximately 500 million years ago and has become a key organism in studying the origin of jawed vertebrates from jawless ancestors. As demonstrated in studies of gonadotropin releasing hormone (GnRH) (Sower et al., 2012), melanocortin (Haitina et al., 2007) and growth hormone (GH)/prolactin (PRL) (Ellens et al., 2013) receptor families, the characterization of neuroendocrine hormone receptor groups in the sea lamprey has immense value in evolutionary studies. These findings are important in deciphering the time sequence of duplication and divergence within gene family lineages, as well as viewing the larger picture of whole-genome duplications (WGD) and speciation. Two rounds of WGD (1R and 2R) early in the vertebrate lineage have been proposed (Ohno, 1970) and apparently validated after the emergence of chordates (Putnam et al., 2008), but the timing of the events with respect to cyclostome-gnathostome divergence is an ongoing topic of study. A major question has been whether the agnathans diverged from the vertebrate lineage prior to 1R, between 1R and 2R, or after 2R (Kuraku et al., 2009; Smith et al., 2013). It has recently been proposed that three rounds (3R) of WGD took place in the lamprey lineage independently of the gnathostome 2R WGD scenario (Mehta et al., 2013; and Nah et al., 2014). The most recent studies, as discussed in Chapter 1, now contest the 2R WGD scenario (Smith and Keinath, 2015). A lamprey meiotic linkage map has been constructed to provide near chromosome-level evidence supporting an agnathan-gnathostome shared 1R WGD with subsequent segmental duplications, fusion and translocation (Smith and Keinath, 2015).

The arginine vasopressin (AVP) and vasotocin (AVT) hormones and their G-protein-coupled receptor (GPCR) families have been characterized in species representing the vertebrate classes Chondrichthyes, Osteichthyes, Amphibia, Reptilia, Aves and Mammalia, but they have not been characterized in Agnatha, the most primitive vertebrate class known to utilize the AVT hormone. Although the AVT preprohormone has been sequenced from the Arctic (Japanese) lamprey (*Lethenteron camtschaticum*, synonym *Lethenteron japonicum*) (Gwee et al., 2009), there have been no studies characterizing the AVT receptors in these agnathans. This constitutes a gap in our evolutionary knowledge of an important hormone-receptor system tied to key survival functions that could aid in resolving the WGD timing issue.

The AVT hormone is found in birds, non-avian reptiles, amphibians and fish, and is orthologous to AVP in mammals. Invertebrate homologs have existed for about 700 million years (Acher et al., 1995). Gwee et al. (2009) provide evidence that the AVP- and OXT-family of neurohypophysial hormones evolved in a common ancestor of jawed vertebrates through tandem duplication of the ancestral AVT gene. Thus, gnathostomes have both paralogs but agnathans have only AVT. Although AVT, AVP and OXT are nine-amino-acid peptides and structurally very similar, AVT differs from AVP by one amino acid in the third position and from OXT by one amino acid in the eighth position (Table 2-1).

Table 2-1. AVP/AVT/OXT receptor family members and their functional characteristics in jawed vertebrate (gnathostome) taxa.

Receptor type	Second messenger	Ligand type	Ligand amino acid sequence	Taxa where found[1]						Major functions
				M	B	R[2]	A[2]	F	S	
V1a	Ca^{++}	AVP	CYFQNCPRG	√						Vascular smooth muscle vasoconstriction; social behavior, blood pressure/heartrate regulation, memory (brain); circadian rhythms (pineal); glucogenolysis (liver); uterine contraction; glomular contraction (kidney); aldosterone/cortisol secretion (adrenal)[3]
		AVT	CYIQNCPRG		√	√	√	√	√	
V1b (V3)	Ca^{++}	AVP	CYFQNCPRG	√						Adrenocorticotropic hormone (ACTH)/β-endorphin release (pituitary); stress adaptation (brain and adrenal); insulin release (pancreas)[3]
		AVT	CYIQNCPRG		√	√	√		√	
Oxytocin	Ca^{++}	OXT	CYIQNCPLG	√					√[4]	Uterine contraction; milk ejection (mammary); maternal behavior, male and female sexual and pro-social behavior, memory and learning (brain); seminiferous tubule contraction (testis) and sperm ejaculation (prostate)[4]
Mesotocin		MT	CYIQNCPIG		√	√	√			
Isotocin		IT	CYISNCPLG					√		
V2a	cAMP	AVP	CYFQNCPRG	√						Resorption of water in renal collecting tubules via induction and membrane insertion of aquaporin-2 (kidney); vascular smooth muscle vasodilation; sodium resorption (lung)[3]
		AVT	CYIQNCPRG			√	√	√		
V2b	Ca^{++}	AVT	CYIQNCPRG					√	√	?
V2c	Ca^{++}	AVT	CYIQNCPRG		√	√	√	√	√?	?

[1]**Taxa**: M-Mammals, B-Birds, R-Non-avian Reptiles, A-Amphibians, F-Fish, S-Sharks. AVP - arginine vasopressin; AVT - arginine vasotocin. [2]Non-avian reptile receptor complement based on the anole lizard; amphibian receptor complement based on frogs and newts (see Lagman et al., 2013). [3]Decaux et al., 2008. [3]Some sharks may have other oxytocin-like ligand types; Gimpl and Fahrenholz, 2001.

Mammals express three AVP receptor types (V1a, V1b and V2) and one OXT receptor which are members of the same subfamily within the Rhodopsin-like family of GPCRs (Birnbaumer, 2000; Gimpl and Fahrenholz, 2001). V1a-type receptors have been found in all vertebrates studied, including teleost fish (Lema, 2010). V1b receptors have been identified in chickens (Cornett et al., 2003) and amphibian newt species (Hasunuma et al., 2007), but have not been found in fish. V2-type receptors involved in osmotic balance, thought previously to have arisen in tetrapods, have been identified in ray- and lobe-finned fish (Lema, 2010; Konno et al., 2009 and 2010). Subsequently, Ocampo Daza et al. (2012) and Yamaguchi et al. (2012) reported a V2b-type receptor that subdivided this lineage. Syntenic evidence supports a further split into a V2c subtype (Lagman et al., 2013), such as the formerly named VT1 receptor in the chicken (*Gallus gallus*) (Baeyens and Cornett, 2006). Table 2-1 provides a summary of the AVP/AVT/OXT family receptor types, their cognate ligands, functions and taxonomic distribution.

The functional result of the hormone-receptor interaction is realized through several intermediary steps along intracellular pathways. Both the V1a and V1b receptors couple with Gq/11 protein to activate phospholipase C, inducing hydrolysis of phosphatidylinositol to inositol triphosphate (IP3) and 1,2,-diacylglycerol which stimulate increased cytoplasmic Ca^{++} levels. What had been called simply the V2 receptor in mammals, now designated V2a (see Table 2-1), activates the stimulatory G protein (G_S) to stimulate adenylyl cyclase increasing cAMP levels (Birnbaumer, 2000; Decaux et al., 2008). Yamaguchi et al. (2012) found that the elephant shark V2b receptor

was structurally similar to the fish and mammalian V2a receptors. However, the shark receptor signals using the IP3 rather than cAMP pathway. Therefore, the phylogenetic split among V2 receptor genes is evident not only by sequence-based algorithms but also by divergence of function.

Jawed-vertebrate phylogenetic studies (Lagman et al., 2013; Yamaguchi et al., 2012; Ocampo Daza et al., 2012; Lema, 2010) have revealed a clear split between the V1/OXT-type receptors and the V2-types. In this study we sequenced and characterized five putative sea lamprey AVT receptors with the hypothesis that agnathan cyclostomes may provide clues to the structure of and orthological connections to ancestral forms within the V1/OXT and V2 receptor groups, providing additional information on the paths of duplication and divergence in the early vertebrate lineage. Our analyses provide evidence supporting the divergence of predecessor V1a/OXT receptor genes and V2b/V2c receptor genes into the OXTR-V2b and V1a-V2c chromosomal blocks found in gnathostomes under a shared 1R scenario with subsequent independent duplications. However, though less probable, our data does not refute the 2R and lamprey-specific 3R scenarios.

2.2 Materials and Methods

2.2.1 Animals

Sea lamprey adult upstream migrants for tissue distribution studies were obtained directly from US Fish and Wildlife Service (USFWS) personnel from barrier traps on the Middle River near Lake Superior in Wisconsin. Additional animals of larval, parasitic and adult life stages were obtained from USGS Hammond Bay Biological Station (HBBS) (Millersburg, MI). The larvae were maintained in sand-lined zebrafish breeder boxes within 10-gallon aquarium tanks at 15 ^{o}C in a temperature-controlled room at the University of Minnesota-Duluth (see Appendix A). A diet of moist cake yeast slurry in twice weekly feedings was provided according to the methods of Hansen et al. (1974). Parasitic-phase and adult animals were held without feeding in a 100-gallon stock tank at 15^{o}C for a maximum of two weeks after receipt. Animals were euthanized in tricaine methanesulfonate (MS-222, Sigma, St. Louis, MO), pH 7.4, at 1 g/L for larvae or 2 g/L for parasitic and adult animals. Animal treatment conformed to University of Minnesota animal care standards (IACUC protocol number 1305-30612A).

2.2.2 Lamprey AVT receptor RNA isolation and cDNA sequence determination

Brains were pooled from several larval lamprey and immediately processed or whole brains and gonadal tissue were collected from adult male and female lampreys and stored in RNAlater (Qiagen, Germantown, MD) at -20 ^{o}C. Total RNA was isolated using TriZol® (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Homogenized tissues were passed several times through a 0.22-gauge needle prior to a QiaShredder column (Qiagen) to shear cellular components. Isolated RNA was treated

with DNase I (DNA-*free*, Ambion, Austin, TX) to remove genomic DNA according to the manufacturer's instructions. The RNA was reverse transcribed to cDNA using ThermoScript™ RT (Invitrogen), with the addition of 1M betaine (Sigma) to eliminate secondary structure.

AVT and OXT-type receptor sequences of teleost fish, amphibian and bird species were used to query the basic local alignment search tool (BLAST) (http://useast.ensembl.org/Petromyzon_ marinus/Tools/Blast?db=core) to find homologous sequences within the lamprey whole genome scaffolds (Smith et al., 2013) deposited in the NCBI (GenBank) and Ensembl (current genome assembly Pmarinus_7.0) databases. The receptors were tentatively named based on original scaffold locations as Pm 807, Pm3133, Pm2017, Pm644 and Pm4769. Candidate sequences were used to design primers within the coding regions of putative lamprey AVT receptors. All primers used are listed in Tables 2-2A through 2-2C. The Phusion High-Fidelity PCR Kit (Thermo Scientific, Waltham, MA) was used in either 50 or 25 µl reactions, with DMSO added to all reactions, HF buffer used for Pm807 and GC buffer used for all other receptors in the proportions recommended by the manufacturer. Betaine at 1M final concentration was added to all reactions except β-actin and Pm807. Final primer concentrations were 0.5 µM.

The GeneRacer™ RACE kit (Invitrogen) was used to determine the 5'-UTR sequences of the Pm807 and the Pm2017 receptor genes, as well as the 5'- and 3'-UTRs of Pm4769. Isolated larval brain RNA was dephosphorylated, decapped and ligated to the RNA oligo following the manufacturer's instructions before reverse transcribing without

the addition of betaine. PCR and RACE products were visualized on 2% agarose gels with SYBR® Safe DNA Gel Stain (Invitrogen). The appropriate-sized bands were excised and purified with the QIAquick® Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The purified products were Sanger sequenced by the University of Minnesota Genomics Center (UMGC) using an Applied Biosystems (ABI) 3730xl sequencer.

2.2.3 Tissue distribution patterns of AVT receptors and preprohormone

Parasitic Phase. Brains, pineal organ, eye, olfactory sac, gill, heart, liver, kidney, gonad, and muscle tissue were dissected from one male and one female parasitic-phase sea lamprey. All tissues were stored in individual vials of RNAlater (Ambion) at -20 ºC. Total RNA was isolated using TriZol® and treated with DNase I as described in section 2.2. Seven hundred nanograms of each RNA was reverse transcribed to cDNA using Superscript III™ (Invitrogen). PCR was performed for Pm807, β-actin and AVT preprohormone in 50 µl reactions using Platinum PCR Supermix with 3 µl cDNA and a 0.5 µM final primer concentration. Primers used are listed in Table 2-2B. Thermal profile for Pm807 was: 2 min at 94 ºC followed by 34 cycles of 30 s at 94 ºC, 30 s at 60 ºC, and 60 s at 72 ºC, with a 5 min elongation step at 72 ºC. Thermal profile for β-actin was: 2 min at 94 ºC followed by 24 cycles of 30 s at 94 ºC, 30 s at 55 ºC, and 60 s at 72 ºC, with a 5 min elongation step at 72 ºC. Thermal profile for AVT preprohormone was: 2 min at 94 ºC followed by 35 cycles of 30 s at 94 ºC, 30 s at 60 ºC, and 60 s at 72 ºC, with a 7 min elongation step at 72 ºC. Products were visualized as in section 2.2.2.

Adult upstream migrants. Two pairs of male and female animals were collected in the migration season of separate years. Male and female pair one were collected in May 2012 and male and female pair two were collected in May 2013 from the Middle River in Wisconsin. Brains, pineal organ, eye, olfactory sac, gill, heart, liver, kidney, gonad, intestine, muscle and skin were dissected from all animals. Cloaca/penile tissues were collected from three animals of each gender obtained from HBBS in 2013 and pooled by gender. Brains collected from pair one were separated into forebrain comprised of the cerebellum, cerebrum, hypothalamus, olfactory and optic lobes; and hindbrain consisting of the medulla oblongata to approximately the start of the spinal cord (near the connection point with the spinal nerves). The brains from pair two were similarly separated into forebrain and hindbrain, but the hypothalami were removed by gripping with forceps and slicing off the protrusion with the tip of a scalpel (Fig. 2-1) and pooled by gender with those of two additional animals (see section 2.4). Pituitaries were not obtained for any of the animals. Tissue storage and RNA preparation was the same as for parasitic phase animals. For the first pair of animals: 700 ng of each RNA was reverse transcribed to cDNA using ThermoScript^{TM} RT (Invitrogen) in a 20 µl reaction; PCR was performed for Pm807, Pm3133, Pm2017, Pm644 and β-actin in a 50 µl reaction using GoTaq Master-Mix (Promega, Madison, WI), with 2 µl cDNA in the first-round reaction and 3 µl of the first-round product in the nested PCR reaction. PCR reactions for all genes except β-actin and Pm807 required a final concentration of 1M betaine.

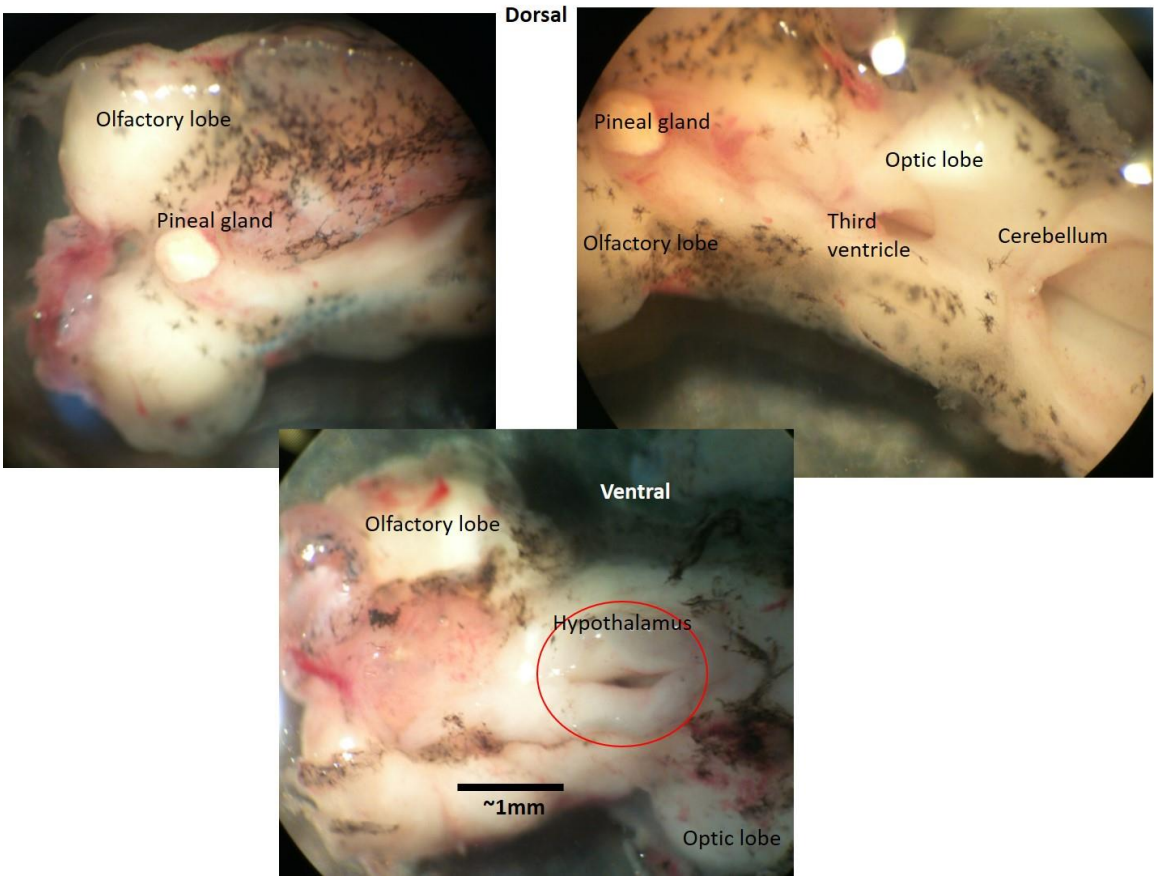


Figure 2-1. Lamprey brain showing the location of forebrain structures. Photographed through a dissecting microscope. Hypothalamus tissue was removed and pooled by gender for quantitative PCR.

Due to low expression in the initial PCR round for pair one, the procedure was modified for the second pair of animals as follows: 1µg of each RNA was reverse transcribed to cDNA as above but with the addition of 1M betaine (Sigma); PCR was performed in 25 µl reactions, with 2 µl cDNA in the first-round reaction and 1 µl of the first-round product in a nested PCR reaction. Thermal profiles were as above for the parasitic phase animals, with the addition of a modified thermal profile for receptor genes other than Pm807: 2 min at 95 ºC followed by 36 cycles of 30 s at 95 ºC, 30 s at 58 ºC, and 60 s at 72 ºC, with a 7 min elongation step at 72 ºC. PCR products were visualized as described in Section 2.2. Negative results were confirmed by retesting (data not shown).

2.2.4 Real-time quantitative PCR of forebrain and hypothalamus AVT receptor expression

RNA was isolated individually from forebrains collected from three male and three female adults, but hypothalami from these same animals were collected and pooled by gender. Primers are listed in Table 2-2C. Real-time qPCR was performed on a Rotor Gene Q (Qiagen) thermocycler profile of 15 min at 95ºC followed by 40 cycles of 95ºC for 10 s, 60ºC for 30 s and 72ºC for 60 s. QuantiTect® SYBR® Green I PCR Master Mix (Qiagen) was used in triplicate 25 µl reactions for each sample. The internal reference standard was β-actin. Assays were conducted for all genes with 0.5 µl of forebrain or hypothalamus cDNA (25 ng RNA) based on standard dilution curves conducted for Pm807 and Pm644 (Suppl. Fig. 2-1, Appendix B). Amplicon sizes ranged from 158 to 285 bp.

Table 2-2. PCR Primers. Length in base pairs and Guanine-Cytosine content (%GC) is given for each segment.

A. Coding sequence including start-stop codons.

Gene transcript	Primer name	Base pairs	%GC	Sequence (5'-3')
Pm807	Clone for	1326	59.6	CTGGAAGTGATTTTGGGAGGAAAAATGC
	Stop rev			AACGGCTCACGCAAAGTACGTCTT
Pm3133	Start for	1150	64.3	ATGGCCAACGGCACGGCGAAT
	Stop rev			ACCCCGCATCCTACTTCTGGCT
Pm2017	5TSS for	1440	70.8	TAATGCACTGCCCATAGGAGGAGG
	Stop rev			AGCGTTACCGTAGGGGCCATCGCA
Pm644	Start for	1274	67.4	CACGGGAACGATGAGTGACGATCCA
	Stop rev			TGGCTCAGGGCAGGCAGGACA
Pm4769	E1 For	1221	63.0	GTGCTGTTCGCCATGTTCGCAT
	Stop Rev			GATCCTTACCGATCACGCGCTAAT

B. Tissue expression.

Gene	1st round primers	Base pairs	%GC	Sequence (5'-3')	Nested primers	Base pairs	%GC	Sequence (5'-3')
Pm807	Ex2a For	991	58.0	AGTCGAGATCGCCTTACTCTCGAT	Ex2b For	285	55.8	AAGGCGTACATCACCTGGATGACA
	Ex3 Rev			ACTCACGAACGTGTGATTCCG	Ex2 Rev			GACTGACCACATCTGCACGAAGAA
Pm3133	Ex1 For	829	63.2	TTTATCATGCACCTGAGCATCGCC	Ex2 For	595	65.4	TTCCGCTTCAATGGCTCGGACATA
	Ex4 Rev			AGTCGTAAATGAGATGCCCGCTGA	Ex3 Rev			AGCAGAAGACGTACACCACGATGA
Pm2017	Ex1 For	796	67.7	TGTTCGCGTCCACCTACGTGCTCAT	Ex2 For	204	58.8	TCATCATCGTGGTCTACGTCGTCT
	Ex3b Rev			AGTGCGTACACCACCACGTGCTCT	Ex3a Rev			AGCAGGACATCACCTCCTCAA
Pm644	Ex1 For	996	66.4	ACGATCCATTCGTTGGGAACTCCT	Ex2 For	461	65.1	GATGTTTGCGTCCGCGTACATGAT
	Ex4a Rev			TGTTGAGGTTTCCCAGCAGCATGA	Ex2-3 Rev			TGTTGATCATGGCCCTGGAGAT
Pm4769	Ex1a For	370	61.1	ACATGCAGATCGTGGGCATGTTTG	Ex1b For	236	62.3	GTCACCTTCAAGAAGCGTATGGCT
	Ex1b Rev			CTGCGGCGTCTTGAAGTAGATGTT	Ex 1a Rev			TTGGTTTGGCACCACACAAGCA
β-actin	Ex1 For	721	61.7	ATCATGTTCGAGACCTTCAACACGC				
	Ex3 Rev			ACTCCTGCTTGCTGATCCACATCT				*NA*

C. Quantitative real-time PCR

Gene	Primer name	Base pairs	%GC	Sequence (5'-3')
Pm807	Ex2b For	285	55.8	AAGGCGTACATCACCTGGATGACA
	Ex2 Rev			GACTGACCACATCTGCACGAAGAA
Pm3133	Ex3 For	192	55.7	TGACCTTCGTCATCATCGTGGTGT
	Ex4 Rev			AGTCGTAAATGAGATGCCCGCTGA
Pm2017	Ex2 For	204	58.8	TCATCATCGTGGTCTACGTCGTCT
	Ex3a Rev			AGCAGGACATCACCTCCTCAA
Pm644	Ex3 For	158	60.8	CATGATCAACACGGTGCGCATGA
	Ex4A Rev			TGTTGAGGTTTCCCAGCAGCATGA
Pm4769	Ex1b For	236	62.3	GTCACCTTCAAGAAGCGTATGGCT
	Ex 1a Rev			TTGGTTTGGCACCACACAAGCA
β-actin	Ex1 For	280	63.2	ATCATGTTCGAGACCTTCAACACGC
	Ex1 Rev			TCTCCTTGATGTCACGCACGATCT

2.2.5 Sequence alignments and phylogenetic analysis

Intron-exon junctions were determined based on sequenced segments and from alignments with other species; all junctions followed the gt/ag rule (Breathnach and Chambon, 1981). Amino acid sequences were deduced from cDNA for the four complete lamprey receptor coding sequences (Pm807, Pm3133, Pm2017, Pm644) and partial sequence for the putative pseudogene (Pm4769). These were aligned with AVP/OXT-family receptors of other species obtained from GenBank and Ensembl, using ClustalW algorithm (Gonnet weight matrix, gap opening penalty 10.0, gap extension penalty 0.2). The alignment was used to construct a molecular phylogenetic tree using the Phylogenetic Maximum Likelihood (PhyML 3.0) method (Guindon et al., 2010) with the following settings: amino acid frequencies, proportion of invariable sites, and gamma shape parameters were estimated from the alignments, substitution rate categories was set to 8, starting tree was created using BIONJ, with NNI tree improvement, and tree topology and branch length optimization. Because we used a similar set of gene sequences to that of Lagman et al. (2013), we chose the same analysis settings, including the JTT model with 100-iteration non-parametric bootstrap tests. We constructed one tree with the sea lamprey vasotocin receptor genes and one tree without, to determine the effect of lamprey coding sequence structure on cluster confidence values. The trees were rooted at the three invertebrate octopus vasopressin-like receptors as an outgroup. The GenBank and Ensembl accession numbers for all sequences used in the analyses are listed in Table 2-3.

Table 2-3. Receptor and species identification for multiple species alignment and phylogenetic tree. Page 1 of 4.

Abbreviation	Description/Alternate Names	Accession^	Latin name	Common name
Pm2017	lamprey scaff 2017 V1a/b/OT-like receptor	KJ813003	*Petromyzon marinus*	Sea lamprey
Lja 2017	Japanese lamprey Pm2017 ortholog predicted	scaf KE993677	*Lethenteron japonicum*	Japanese lamprey
Aca V1b	vasotocin 1b receptor predicted	XP_003224056	*Anolis carolinensis*	Anole lizard
Cpy V1b	vasotocin 1b receptor	BAF38756	*Cynops Pyrrhogaster*	Firebelly newt
Lch V1b	vasotocin 1b-type receptor (fragment)	ENSLACP00000015539	*Latimeria chalumnae*	Coelocanth
Tgr V1b	vasotocin 1b receptor	ABQ23254	*Taricha granulosa*	Rough-skinned newt
Pan V1b	vasotocin 1b-type receptor	BAL70281	*Protopterus annectens*	West African lungfish
Cmi V1b	vasotocin 1b-type receptor, CM033	AB665984	*Callorhinchus milii*	Elephant shark
Gga V1b	vasotocin 1b receptor, VT2	AAG17937	*Gallus gallus*	Chicken
Hsa V1b	vasopressin 1b receptor	NP_000698	*Homo sapiens*	Human
Mmu V1b	vasopressin 1b receptor	NO_036054	*Mus musculus*	Mouse
Aoc V1a	Vasotocin 1a receptor	BAL70404	*Amphiprion ocellaris*	Clownfish
Cpy V1a	vasotocin 1a receptor	BAF38754	*Cynops Pyrrhogaster*	Firebelly newt
Lch V1b	vasotocin 1a-type receptor (fragment)	ENSLACP00000016482	*Latimeria chalumnae*	Coelocanth
Lpl V1a	vasotocin 1a receptor predicted	ENSLOCP00000020992	*Lepisosteus platyrhincus*	Spotted gar
Ola V1a1	vasotocin 1a receptor 1	BAL45623	*Oryzias latipes*	Medaka
Ola V1a2	vasotocin 1a receptor 2	BAL45624	*Oryzias latipes*	Medaka
Cmi V1a	vasotocin 1a receptor	AB665985	*Callorhinchus milii*	Elephant shark
Dre V1a	vasotocin 1a receptor predicted	XP_700287	*Danio rerio*	Zebrafish
Cne V1a2	vasotocin 1a receptor 2	GQ981413	*Cyprinodon nevadensis amargosae*	Amargosa pupfish
Cne V1a1	vasotocin 1a receptor 1	GQ981412	*Cyprinodon nevadensis amargosae*	Amargosa pupfish
Pan V1a	vasotocin 1a receptor	BAG66063	*Protopterus annectens*	West African lungfish
Xtr V1a	vasotocin 1a receptor predicted	XP_002931664	*Xenopus (Silurana) tropicalis*	African clawed frog
Tgr V1a	vasotocin 1a receptor	ADF30857	*Taricha granulosa*	Rough-skinned newt
Gga V1a	Chicken vasotocin 1a receptor, VT4	NP_001103908	*Gallus gallus*	Chicken

Table 2-3. Page 2 of 4.

Abbreviation	Description/Alternate Names	Accession^	Latin name	Common name
Mmu V1a	Mouse vasopressin 1a receptor	NP_058543	*Mus musculus*	Mouse
Hsa V1a	Human vasopressin 1a receptor	NP_000697	*Homo sapiens*	Human
Pm3133	lamprey scaffold 3133 V1a/OT-like receptor	KJ813004	*Petromyzon marinus*	Sea lamprey
Pm807	lamprey scaffold 807 V1a/OT-like receptor	KC731437	*Petromyzon marinus*	Sea lamprey
Lja 3133	Japanese lamprey Pm3133 ortholog predicted	scaf KE994228	*Lethenteron japonicum*	Japanese lamprey
Lja 807	Japanese lamprey Pm807 ortholog predicted	scaf KE993677	*Lethenteron japonicum*	Japanese lamprey
Mmu OTR	oxytocin receptor	NP_001074616	*Mus musculus*	Mouse
Hsa OTR	human oxytocin receptor	CAA46097	*Homo sapiens*	Human
Aoc ITR	isotocin receptor	BAL70407	*Amphiprion ocellaris*	Clownfish
Aca MTR	mesotocin receptor predicted	XP_003224939	*Anolis carolinensis*	Anole lizard
Lch OTR	isotocin receptor	ENSLACP00000020898	*Latimeria chalumnae*	Coelocanth
Lpl ITR	isotocin receptor putative	ENSDARP00000050113	*Lepisosteus platyrhincus*	Spotted gar
Ola OTR	oxytocin-like receptor	NP_001234561	*Oryzias latipes*	Medaka
Xma ITR1	isotocin receptor 1 predicted	XP_005799868	*Xiphophorus maculatus*	Southern platyfish
Xma ITR2	isotocin receptor 2 predicted	ENSXMAP00000016320	*Xiphophorus maculatus*	Southern platyfish
Tgr MTR	mesotocin receptor	ABA27136	*Taricha granulosa*	Rough-skinned newt
Xtr MTR	mesotocin receptor predicted	XP_002936297	*Xenopus (Silurana) tropicalis*	African clawed frog
Gga MTR	mesotocin receptor	AAX07197	*Gallus gallus*	Chicken
Pan MTR	mesotocin receptor	BAL70280	*Protopterus annectens*	West African lungfish
Dre ITR	isotocin receptor	ACT83513	*Danio rerio*	Zebrafish
Cne ITR	isotocin receptor	GQ981415	*Cyprinodon nevadensis amargosae*	Amargosa pupfish
Cmi OTR	oxytocin receptor	AB665983	*Callorhinchus milii*	Elephant shark
Gac V2b1	vasotocin 2b receptor 1 *Ocampo Daza 2012*	--	*Gasterosteus aculeatus*	Stickleback minnow
Dre V2b1	vasotocin ("oxytocin-like") predicted	XP_003198871	*Danio rerio*	Zebrafish

Table 2-3. Page 3 of 4.

Abbreviation	Description/Alternate Names	Accession^	Latin name	Common name
Cmi V2b1	vasotocin 2b receptor 1, V1c, CM198	AB665982	*Callorhinchus milii*	Elephant shark
Aoc V2b	vasotocin 2b receptor aka V1c	BAL70405	*Amphiprion ocellaris*	Clownfish
Xma V2b1	vasotocin V2b receptor 1 predicted	XP_005799961	*Xiphophorus maculatus*	Southern platyfish
Xma V2b2	vasotocin V2b receptor 2 predicted	ENSXMAP00000002229	*Xiphophorus maculatus*	Southern platyfish
Pm4769	lamprey scaffold 4769/1197 V2-like (partial)	KM015305	*Petromyzon marinus*	Sea lamprey
Pm644	lamprey scaffold 644 V2-like	KF031008	*Petromyzon marinus*	Sea lamprey
Lja 4769	Japanese lamprey Pm4769 ortholog	scaf KE993677	*Lethenteron japonicum*	Japanese lamprey
Lja 644	Japanese lamprey Pm644 ortholog	scaf KE993674	*Lethenteron japonicum*	Japanese lamprey
Lch V2c	vasotocin 2c receptor (fragment)	ENSLACP00000015306	*Latimeria chalumnae*	Coelocanth
Lpl V2c	vasotocin 2c receptor predicted	XP_006642964	*Lepisosteus platyrhincus*	Spotted gar
Xma V2c	vasotocin 2c receptor predicted	XP_005799151	*Xiphophorus maculatus*	Southern platyfish
Gga V2c	vasotocin 2c receptor, VT1	NP_001026650	*Gallus gallus*	Chicken
Xtr V2c	vasotocin 2c receptor predicted	XP_002932869	*Xenopus (Silurana) tropicalis*	African clawed frog
Cmi V2c	pseudogene vasotocin 2b receptor 2, CM900	AB671271	*Callorhinchus millii*	Elephant shark
Aoc V2a	vasotocin 2a receptor	BAL70406	*Amphiprion ocellaris*	Clownfish
Aca V2a	vasotocin 2a receptor predicted	XP_003216915	*Anolis carolinensis*	Anole lizard
Cpy V2a	vasotocin 2a receptor	BAF38755	*Cynops Pyrrhogaster*	Firebelly newt
Lch V2a	vasotocin 2a receptor	ENSLACP00000016913	*Latimeria chalumnae*	Coelocanth
Ola V2a1	vasotocin 2a receptor 1	BAL62980	*Oryzias latipes*	Medaka
Ola V2a2	vasotocin 2a receptor 2	BAJ04637	*Oryzias latipes*	Medaka
Xma V2a1	vasotocin 2a receptor predicted	XP_005801420	*Xiphophorus maculatus*	Southern platyfish
Pan V2aR	vasotocin 2a receptor	BAG66064	*Protopterus annectens*	West African lungfish
Xtr V2aR	vasotocin 2a receptor predicted	XP_002937279	*Xenopus (Silurana) tropicalis*	African clawed frog
Tgr V2aR	vasotocin 2a receptor	ABQ23253	*Taricha granulosa*	Rough-skinned newt

Table 2-3. Page 4 of 4.

Abbreviation	Description/Alternate Names	Accession^	Latin name	Common name
Dre V2aR2	vasotocin 2a receptor 2 predicted	XP_001346005	*Danio rerio*	Zebrafish
Dre V2aR1	vasotocin 2a receptor 1 predicted	XP_001922042	*Danio rerio*	Zebrafish
Cne V2aR	vasotocin 2a receptor	GQ981414	*Cyprinodon nevadensis amargosae*	Amargosa pupfish
Mmu V2aR	vasopressin 2a receptor	NP_062277	*Mus musculus*	Mouse
Hsa V2aR	vasopressin 2a receptor (isoform 1)	NP_000045	*Homo sapiens*	Human
Dre V2c	vasotocin 2-like, V2c receptor, V2bR2	NP_001103595	*Danio rerio*	Zebrafish
Gac V2c	vasotocin 2-like, V2c *Ocampo Daza 2012*, V2bR2	--	*Gasterosteus aculeatus*	Stickleback minnow
Ciona VPR	Invertebrate vasopressin receptor-like, Ci-VP-R	BAG72194	*Ciona intestinalis*	Vase tunicate (sea squirt)
Ovu CTR1*	octopus cephalotocin receptor 1	BAD67169	*Octopus vulgaris*	Common octopus
Ovu CTR2*	octopus cephalotocin receptor 2	BAD67167	*Octopus vulgaris*	Common octopus
Ovu OPR*	octopus octopressin receptor	BAD67168	*Octopus vulgaris*	Common octopus

^Accession numbers starting with ENS are from Ensembl database; scaffold numbers are from Japanese lamprey genome database; all other numbers are from GenBank. *Trees in Fig. 2-6A and B were rooted at the octopus receptors.

We also constructed a tree to help distinguish among orthologs of the Slit-Robo GTPase-activating proteins (srgap) 1, 2, and 3, and Rho GTPase-activating protein (arhgap) 4 for the synteny analysis, for which the faster non-parametric SH-like approximate likelihood ratio (aLRT) method was used. The sea squirt *Ciona intestinalis* srgap-like gene served as an outgroup.

2.2.6 Synteny analysis

An analysis of neighboring genes was performed comparing the chromosomal blocks containing AVP/OXT-family receptors in gnathostome representatives with sea lamprey (*P. marinus*) scaffold information in Ensembl and Japanese lamprey (*L. japonicum*) scaffold sequences in the Japanese Lamprey Genome Project (Jlamp) database (http://jlampreygenome.imcb.a-star.edu.sg/), assembly version LetJap1.0 (Mehta et al., 2013). Scaffolds identified in Smith et al. (2013) supplementary data and the Pmarinus_7.0 assembly in Ensembl as containing vasotocin receptor genes were examined for the existence of neighboring genes. Deduced amino acid sequences associated with these neighboring genes were used as queries in protein BLAST searches of the NCBI database and BLAST/BLAT searches of Ensembl to identify the potential orthologs in gnathostome species. The tblastn search tool was used to locate the sea lamprey vasotocin receptor orthologs and their neighbors in the Jlamp database. If necessary, as in the case of Pm3133 where neighboring genes were unknown, 10 kilobase (kb) nucleotide segments from the Jlamp scaffold that had been established to contain the ortholog were in turn used in blastn searches of the sea lamprey Pmarinus_7.0 assembly.

UniProt amino acid sequences of possible gene hits were blasted back against the NCBI database to determine or verify gene identity and in Jlamp to determine exact location within the scaffold. Locations in megabases (MB) and strand orientation in each species were verified with the chromosomal location information in the NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/mapview/) database and Ensembl Genome Browser (http://useast.ensembl.org/index.html). Genes included in the analysis in addition to the AVP/OXTR genes and their neighboring genes found on the sea lamprey or Japanese lamprey scaffolds were genes that are nearest neighbors in gnathostomes, those that show a pattern of connections where an AVP/OXTR gene is missing or rearranged, and marker genes established to be associated with the four AVP/OXTR chromosomal blocks (Lagman et al., 2013).

2.3 Results

2.3.1 Lamprey AVT receptor gene structure and sequences

Using BLAST searches and alignment among the predicted amino acid sequences of various species (Table 2-3), we amplified and sequenced the cDNA for the complete coding sequences of four putative AVT receptors and a partial sequence of a fifth receptor, confirming exact intron-exon junctions and alternative splicing. GenBank Accession numbers are listed in Table 2-4. The predicted amino acid sequences are aligned in Fig. 2-2, where seven transmembrane (TM) domains are indicated in all of the lamprey AVT receptor genes. Other characteristics typical of G protein-coupled receptors of higher vertebrates include the "DRY" motif in intracellular loop (ICL) 2 involved in G-protein coupling. Conserved residues involved in ligand binding specific to AVP family receptors (Lema, 2010) are also indicated in Fig. 2-2. Nucleotide and deduced amino acid sequences for the receptor genes are presented in Supplementary Figs. 2-2A through 2-2E (Appendix B). Alignments between the amino acid sequences of these receptors with their individual orthologs in the Japanese lamprey are shown in Supplementary Figs. 2-3A through 2-3E (Appendix B). A partial sequence for the AVT preprohormone found on scaffold 10824 was also determined, which is identical to that for *Lethenteron camtschaticum* (GenBank Accession number D31871) (Suppl. Fig. 2-2F, Appendix B).

Table 2-4. Vasotocin receptor gene accession numbers and locations within the *Petromyzon marinus* genome (Assembly 7.0).

Gene/scaffold	Type	Orient.	Gene Accession	Exon	Genomic Location	Start[1]	End[2]	Contig	Contig Accession
Pm807	V1	-	KC731437	I	GL477135	241614	241741	68529	AEFG01068454
				II	GL477135	201826	202823	68527	AEFG01068452
				III	GL477135	197234	197547	68527	AEFG01068452
Pm3133	V1	-	KJ813004	I	GL479461	14318	14536	38181	AEFG01038182
				II	GL479461	11543	12009	38181	AEFG01038182
				III	GL479461	11223	11440	38182	AEFG01038183
				IV	GL479461	3826	4061	38182	AEFG01038183
Pm2017	V1	+	KJ813003	I	GL478345	15683	15709	22350	AEFG01022351
				II	GL478345	24123	24974*	22350	AEFG01022351
				III	GL478345	27413	27636	22350	AEFG01022351
				IV	GL478345	33471	33823	22350	AEFG01022351
Pm644	V2	+	KF031008	I	GL476972	181639	181986	12030	AEFG01012031
				II	GL476972	182534	183027	12030	AEFG01012031
				III	GL476972	185794	185912	12031	AEFG01012032
				IV	GL476972	188734	189032	12032	AEFG01012033
Pm4769	V2	+	KM015305	I	GL481097	22590	23534	33746	AEFG01033747
(scf. 1197)				II	GL477525	13661	14010	17058	AEFG01017059

[1]Start locations are at the transcription start site for Pm807 and Pm2017, at the earliest in-frame position for Pm4769 and at the start codon for other genes. [2]End locations are at the stop codon for all genes. Orient.: relative gene orientation.
*Unstable trinucleotide repeat region resulted in the loss of 27bp. See Suppl. Fig. 2-2, A-E.

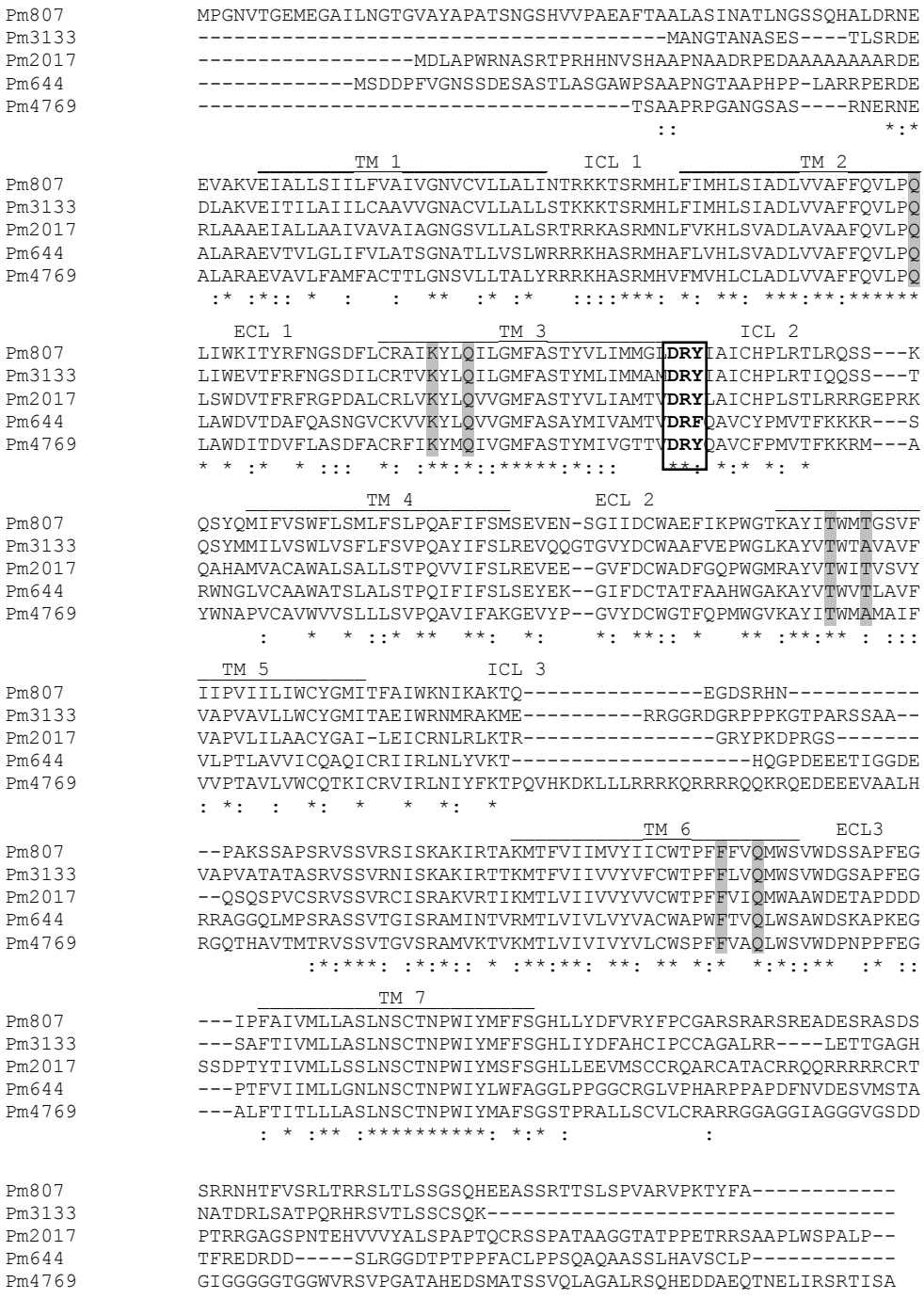


Figure 2-2. Alignment and transmembrane (TM) characteristics of the sea lamprey putative AVT receptor deduced amino acid sequences. Conserved residues in all receptors are indicated by an asterisk (*), identical residues in four of five receptors, or three receptors having one residue and two another are indicated by a semicolon (:). A line above a sequence indicates each approximate TM domain. Intracellular loops (ICL) and extracellular loops (ECL) are also indicated. The "DRY" motif in ICL 2 involved in $G_{q/11}$-protein coupling is boxed. Conserved residues involved in ligand binding to the receptors are highlighted in gray.

We selected the often-utilized β-actin gene as a positive reference gene. Because we found that no β-actin coding sequence has previously been characterized in lampreys, we sequenced and deposited a 721-bp segment in GenBank (Accession number KJ831069) (see Suppl. Fig. 2-2G, Appendix B). We determined its location and compared it to known β-actin sequences to verify its identity. The lamprey β-actin partial sequence has two introns and is found on scaffold 3188, contigs 27920 (AEFG01027921) and 27921 (AEFG01027922); the adjacent upstream contig (AEFG01027920) contains a gene similar to slc29a4 which is also near the β-actin gene (ACTB) on human chromosome 7. The amino acid translation of this β-actin partial sequence shares 100% identity with that of the human β-actin cytoplasmic 2 protein.

Genomic studies have provided new support regarding the significance of intron position in determining duplication events, orthology and paralogy of genes (Henricson et al., 2010; Putnam et al., 2008). A schematic diagram depicting each of the five lamprey vasotocin receptor genes and the human AVPR1A, AVPR1B, OXTR and AVPR2A genes for comparison is presented in Fig. 2-3. The intron lengths were not empirically determined in this study but are based on the current genome assembly. Within the coding regions, all of the lamprey AVT receptors had an intron between the sixth and seventh TM domains, as is the case for all of the human receptors. This was the only intron in the Pm807 and Pm4769 coding regions, while Pm2017 had one additional intron between the fifth and sixth TMs, and Pm3133 and Pm644 had an intron in that location and another intron between the second and third TM domains. The human AVPR2 gene also has an additional intron in the coding region.

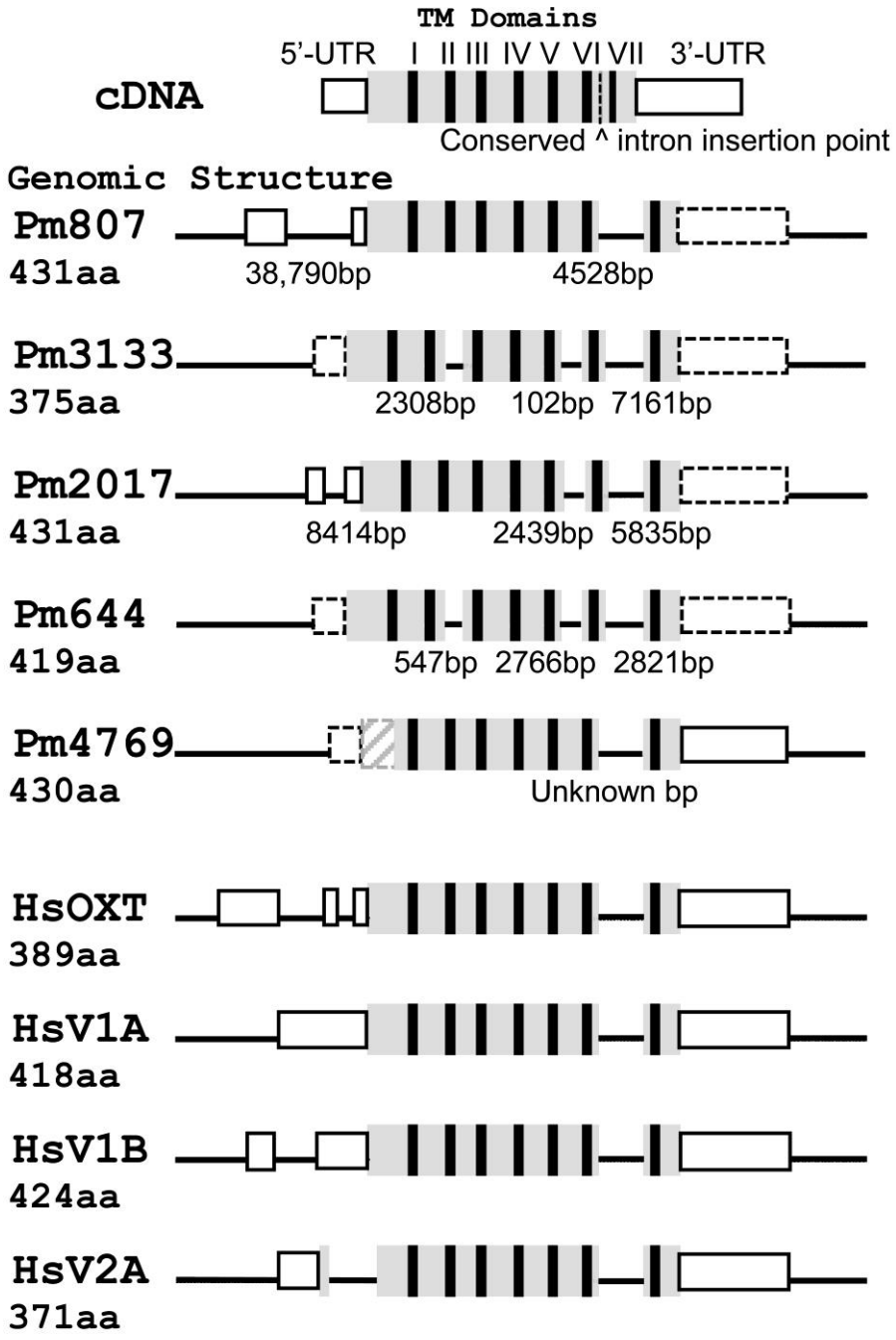


Figure 2-3. Structural schematic comparing intron-exon structure (5'-3') of the putative sea lamprey (Pm) vasotocin receptor genes to human (Hs) vasopressin and oxytocin receptor genes. Exons are indicated by boxes, grey for coding region, white for untranslated region (UTR), vertical black bars are transmembrane domains (TM), connecting lines are introns or inter-genic regions. The structures are not to scale. Sequences of UTRs with broken lines have not been determined. Complete coding sequences were determined for all receptors except the putative pseudogene Pm4769, shown with grey hatching. Alternative isoforms are not shown (See Suppl. Fig. 2-2, A-E).

RACE was conducted to identify features within the 5'- UTRs of the most highly expressed (Pm807) and the most specifically expressed (Pm2017) receptors, as well as the 5'- and 3'-ends of Pm4769. The 5'-UTRs in both Pm807 (Suppl. Fig 2-2A) and Pm2017 (Suppl. Fig. 2-2C) were determined to contain introns. Introns are also found in the 5' UTRs of human OXTR and AVPR1B genes (Fig. 2-3).

Pm807 was found to have an alternative transcription start site (TSS) 51 bp downstream of the first site, and a potential alternative TIS 24 bp downstream of the first TIS. A region of instability with a trinucleotide (CGT) repeat motif resulted in the deletion of 27 bp in the Pm2017 5' UTR of our larval lamprey brains compared with the sequence in the genome database. The deleted sequence (CGTCGCCGTCATCGTCGTCATCGTCGC) is flanked by $(CGT)_4$ and $(CGT)_2$ which become $(CGT)_6$ in the deletion variant. No translation initiation site (TIS) could be found in the genome database for Pm4769 and attempts to find a TIS using 5' RACE were unsuccessful. Scaffold 4769 (GL481097) contains an in-frame stop codon immediately preceding the predicted open reading frame (ORF) comparative to the other receptors. Based on the predicted translation of Pm4769 and the sequence of the ortholog of this gene found in the Jlamp database, the N-terminus is estimated to be 26 amino acids in length, consistent with the other lamprey receptor N-termini (Fig. 2-2). The potential N-terminus is shown in an alignment between the deduced amino acid sequences of the Pm4769 receptor and its Japanese lamprey ortholog in Supplementary Fig. 2-3E, indicating a possible error in the Pmarinus_7.0 sequence construction around this gene.

The last exon of Pm4769 was found on scaffold 1197 (GL477525), preventing determination of the length of the intervening intron although 3' RACE confirmed the junction between the two scaffolds. An alternative isoform of the Pm4769 mRNA was found in which a 110 bp intron is created in the first exon resulting in a frame shift (Suppl. Fig. 2-2E, Appendix B).

2.3.2 *Tissue distribution patterns of AVT receptors and preprohormone*

We determined relative tissue-, life-stage-, and gender-specific expression patterns to provide insight into each receptor's possible physiological function, as specific receptor types within this family exhibit similar expression patterns across numerous phyla. Our initial work in parasitic phase animals showed abundant mRNA for Pm807 from most tissues after one round of conventional reverse-transcription (RT)-PCR (Fig. 2-4A). Due to low expression of the other receptors in the initial round of PCR for the first pair of adult animals (2012; Suppl. Fig. 2-4A), the procedure was modified for all genes in a second pair of adults collected in 2013 to include nested PCR, which also gives an accurate indication of low mRNA abundance versus complete lack of expression. Similarly, quantitative PCR could not be comparably conducted among all receptors and tissues with multiple order-of-magnitude differences in transcript levels. All receptors and the AVT preprohormone were strongly expressed in the forebrain of all animals tested (adult and parasitic) and more strongly in the hypothalamus (tested only in adults) after one round of conventional RT-PCR (Fig. 2-4A and B). Therefore, for orthological comparisons, conventional RT-PCR was chosen as the most informative method to provide relative receptor mRNA expression pattern data in most tissues,

augmented with qPCR in the tissues with high mRNA expression (forebrain and hypothalamus).

Conventional RT-PCR products visualized on agarose gels are presented in Fig. 2-4A - 2-4D and Supplementary Fig. 2-4A. No bands were observed in no-template and non-reverse-transcribed negative controls in any tissue, and ribosomal bands confirmed quality of isolated RNA (Suppl. Fig. 2-4B). In summary, the vasotocin preprohormone gene was transcribed in all adult tissues (Fig. 2-4C); tissues expressing all receptors after nested PCR were the gill, eye, and gonad; and among all receptors, tissues that most often lacked any expression after nested PCR were the liver, heart and skin (Fig. 2-4D). Quantitative real-time PCR was used to assess AVT receptor mRNA levels in the hypothalamus compared with the rest of the forebrain where in other vertebrates each receptor has its own transcriptional specificity (Vaccari et al., 1998; Hasunuma et al., 2010). Relative expression was determined in the forebrains of three males and three females individually, and hypothalami were pooled by gender from the same animals. Forebrain reactions for Pm3133 and Pm2017 yielded C_t values >40, so relative abundances of these mRNAs were considered below the limit of reliable quantitation. Levels were similar between genders for the other three receptors in forebrains (Fig. 2-5). Hypothalamus expression was substantially higher than in the rest of the forebrain for all receptors, and all receptors were detected at C_t values of 28 cycles or fewer. Pm807, Pm4769 and Pm2017 were somewhat more strongly expressed in the female than male pooled hypothalamus samples. The level of expression of Pm644 in hypothalamus was more than 20 times its expression in forebrain (Fig. 2-5).

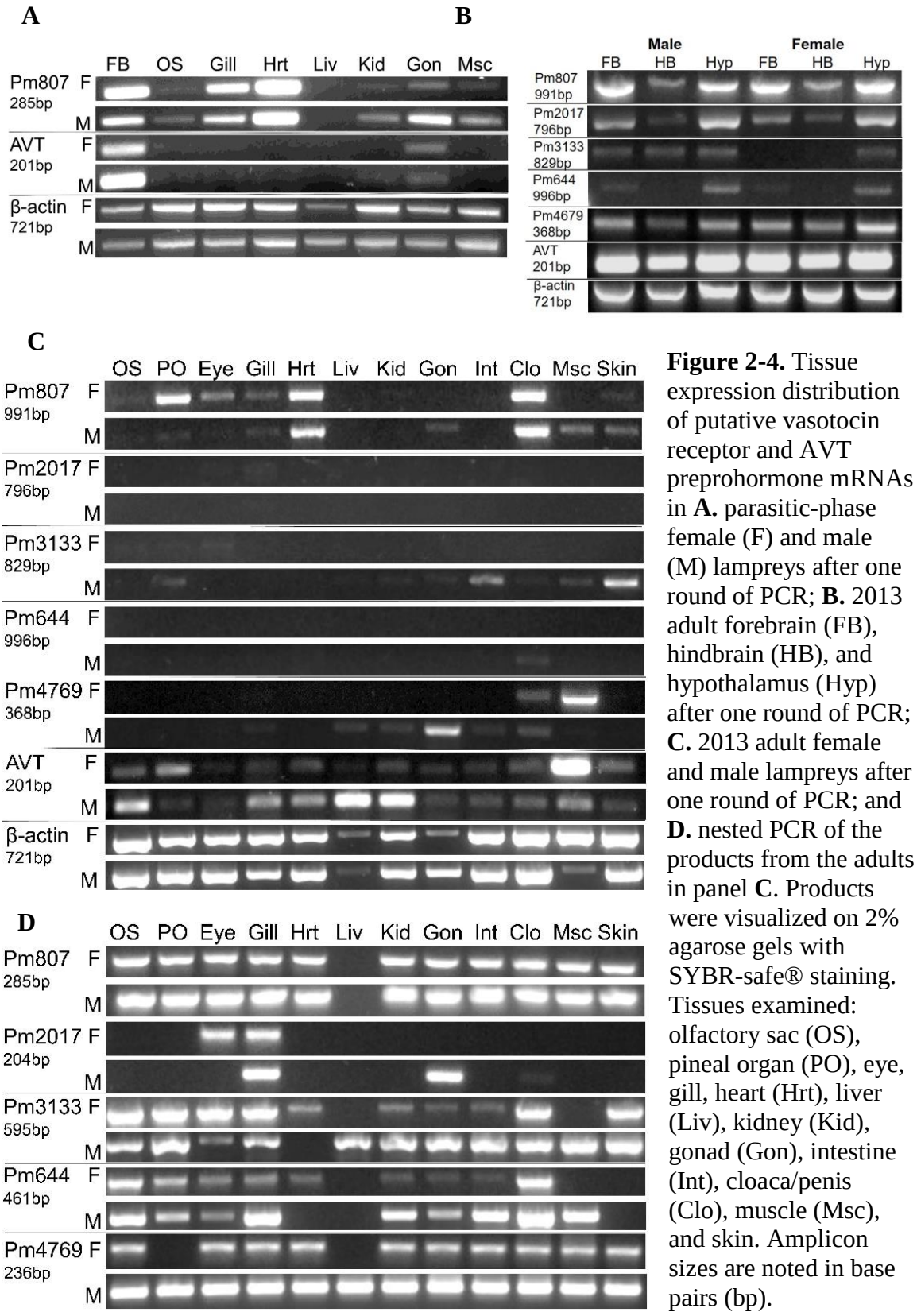


Figure 2-4. Tissue expression distribution of putative vasotocin receptor and AVT preprohormone mRNAs in **A.** parasitic-phase female (F) and male (M) lampreys after one round of PCR; **B.** 2013 adult forebrain (FB), hindbrain (HB), and hypothalamus (Hyp) after one round of PCR; **C.** 2013 adult female and male lampreys after one round of PCR; and **D.** nested PCR of the products from the adults in panel **C**. Products were visualized on 2% agarose gels with SYBR-safe® staining. Tissues examined: olfactory sac (OS), pineal organ (PO), eye, gill, heart (Hrt), liver (Liv), kidney (Kid), gonad (Gon), intestine (Int), cloaca/penis (Clo), muscle (Msc), and skin. Amplicon sizes are noted in base pairs (bp).

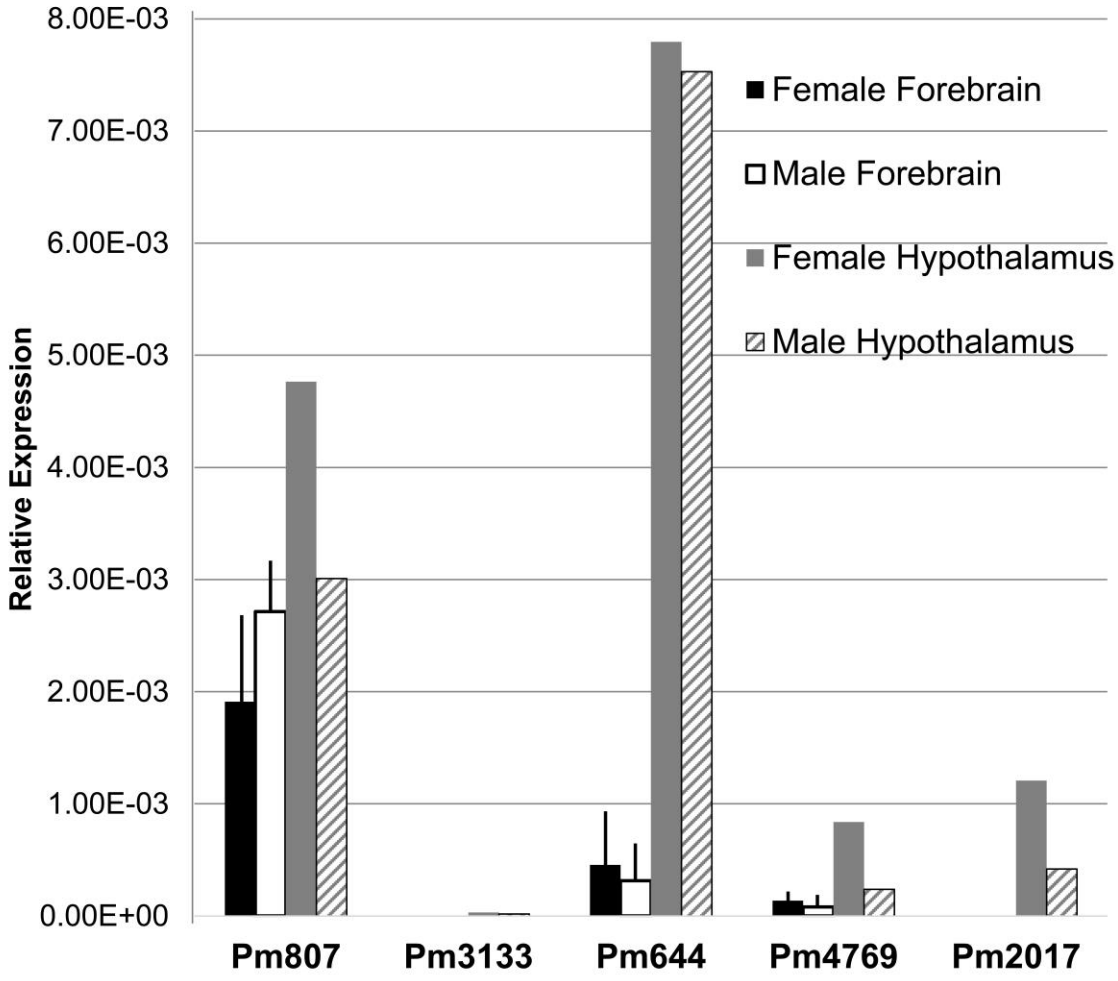


Figure 2-5. Real-time quantitative PCR measurement of adult lamprey AVT receptor mRNA relative expression in forebrains and hypothalami. Results are normalized to β-actin. Forebrain samples were tested separately (n=3 per gender, mean+SD) and hypothalami were pooled by gender from the same animals. Forebrain reactions for Pm3133 and Pm2017 yielded C_t values >40.

2.3.3 Phylogenetic tree analysis

Phylogeny was analyzed through the Phylogenetic Maximum Likelihood method to determine evolutionary relationships among the agnathan lamprey putative AVT receptors and those of jawed vertebrates. The amino acid sequence alignment among all receptors used to create the tree in this study is presented in Supplementary Fig. 2-5. In our phylogenetic tree without the lamprey receptors (Fig. 2-6A), gnathostome receptors cluster into well-defined clades except for the V2c genes, consistent with trees published by other researchers (Yamaguchi et al., 2012, Ocampo Daza et al., 2012 and Lagman et al., 2013). When the sea lamprey receptors are included (Fig. 2-6B), branch support values are substantially lower and the lamprey receptors do not align within the clades of V1a, V1b, OXT, V2a, V2b and V2c receptors established for gnathostomes. Branch support (bootstrap) values drop from 96% to 87% at the major branch between the V2a and V1/OXT receptor clades, from 90% to 69% at the V1a/V1b node and from 98% to 77% at the V1ab/OXTR node when the sea lamprey receptors are added to the analysis. In the gnathostome tree (Fig. 2-6A) the outer V2c cluster forms a node (100%) outside of the V2a/V2b node (65%), with an inner V2b/V2c node (73%). In the lamprey tree (Fig 2-6B) however, these node values are 100%, 49% and 33%, respectively, with the cluster containing the sea and Japanese lamprey Pm644 and Pm4769 sequences forming a node between V2b and inner V2c clusters with values less than 50%. The positions of Pm644 and Pm4769 are not likely influenced by a small number of missing residues in the Pm4769 N-terminus (Suppl. Fig. 2-3E). Pm807 and Pm3133 form a node with the OXTR cluster (61%) while the Pm2017 orthologs form a node with the V1B cluster (70%).

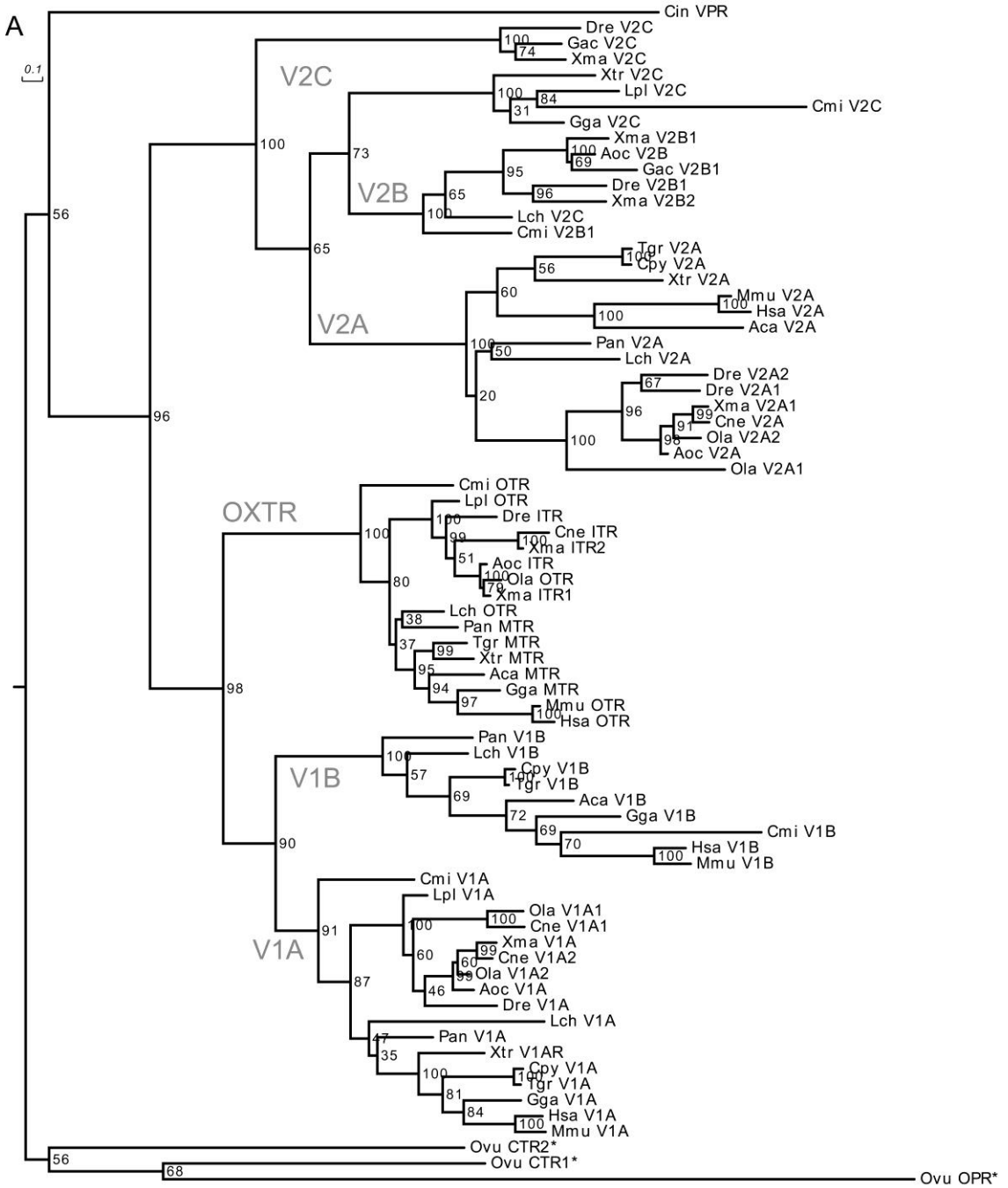


Figure 2-6. Molecular phylogenetic trees of the oxytocin/vasopressin receptor family. **(A) Jawed vertebrate (gnathostome) receptor sequences only.** Tree was constructed using the Phylogenetic Maximum Likelihood (PhyML 3.0) method from a ClustalW amino acid sequence alignment and is rooted at the three invertebrate common octopus (Ovu) sequences as an outgroup. Bootstrap values based on 100 replicates are shown as percentages at the branch points (values <50% are not considered informative). The scale bar represents a genetic distance of 0.1 substitutions per site. Abbrieviations for organism scientific and common names, and gene accession numbers are included in Table 2-3.

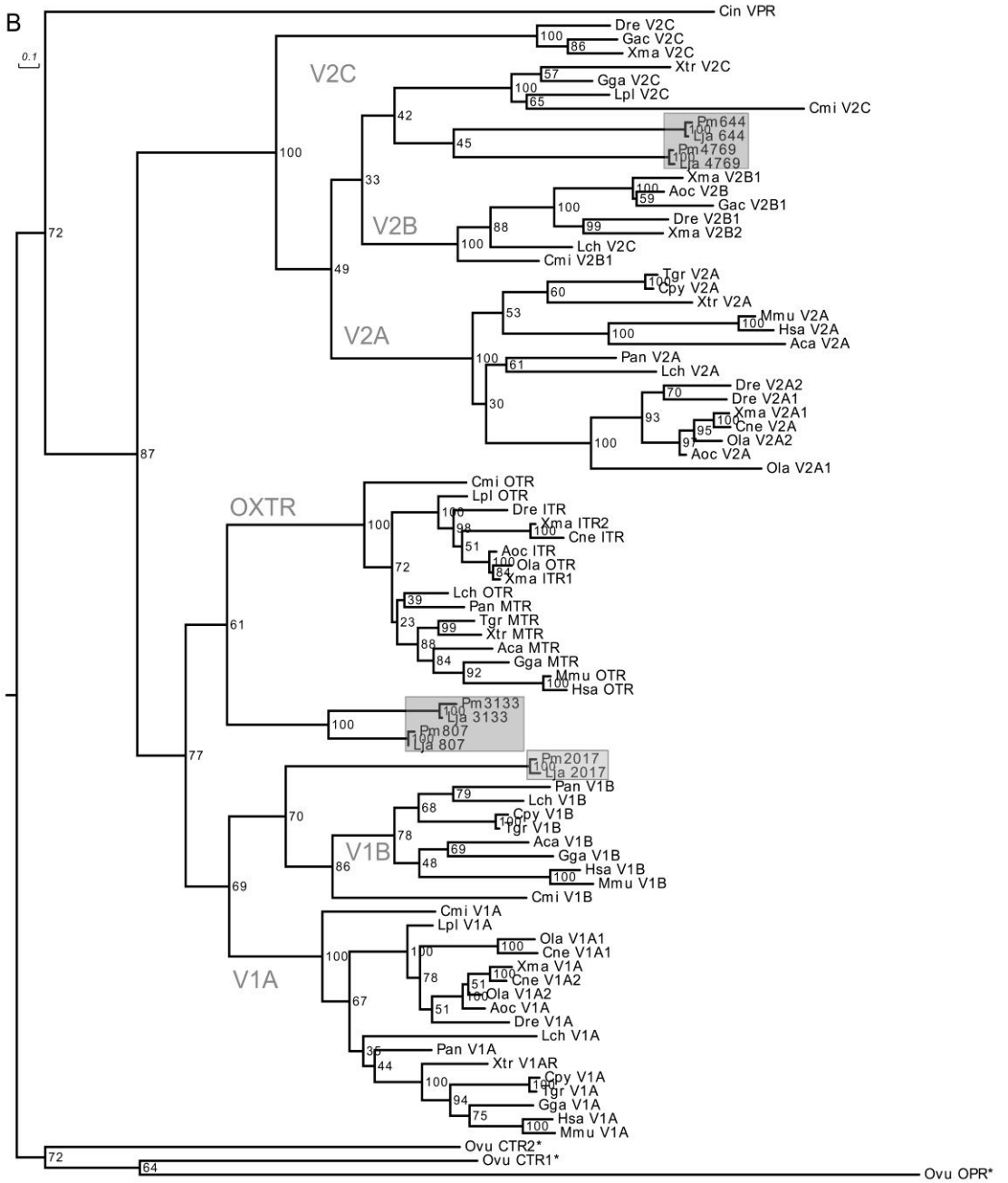


Figure 2-6. Molecular phylogenetic trees of the oxytocin/vasopressin receptor family. **(B) Sea lamprey (Pm) and Japanese lamprey (Lja) vasotocin receptors (shaded)** in relation to the jawed vertebrate (gnathostome) receptor sequences. Tree was constructed using the Phylogenetic Maximum Likelihood (PhyML 3.0) method from a ClustalW amino acid sequence alignment and is rooted at the three invertebrate common octopus (Ovu) sequences as an outgroup. Bootstrap values based on 100 replicates are shown as percentages at the branch points (values <50% are not considered informative). The scale bar represents a genetic distance of 0.1 substitutions per site. Abbrieviations for organism scientific and common names, and gene accession numbers are included in Table 2-3.

2.3.4 Gene synteny analysis

A comparison of AVP/OXT-family receptors on gnathostome chromosomes in relation to adjacent genes on both sea lamprey and Japanese lamprey scaffolds is depicted in Fig. 2-7. Pm3133, Pm2017, and Pm4769 are the sole protein-coding genes on their respective sea lamprey scaffolds, however genes neighboring their orthologs were found on the Japanese lamprey scaffolds so that syntenic relationships could be established (Fig. 2-7A). Neighboring genes comprising sea lamprey scaffold 807 (GL477135) are ssuh2, Pm807 AVT receptor, fbxo32, scala5, tmem5, srgap1, C12orf66, and thumpd3, which except for fbxo32 and scala5 are interdigitated members of gnathostome chromosomes adjacent to either V1a and OXTR genes. The Japanese lamprey genome assembly places these same genes except for fbxo32 and scala5 on the same scaffold (KE993674) with both the Pm807 and Pm644 orthologs. The fbxo32 and scala5 genes appear to be orthologous to the FBXO32 and SCALA5 genes found on human chromosome 8 (which bears no synteny with AVP/OXT receptor genes) and were excluded from the analysis as uninformative and possibly misplaced. Strikingly, the Pm2017 and Pm4769 orthologs also share the same scaffold (KE993677) in the Jlamp assembly, also bearing interdigitated genes appearing on the V1a and OXTR chromosomes, none of which are duplicated within the Pm807-Pm644 scaffold. We identified two neighboring genes near the Pm3133 gene ortholog on Japanese lamprey scaffold KE994228. These are yaf2 and pzdrn4, adjacent on the V1a-V2c chromosome block. In chicken chromosome 1 they are downstream of both receptor genes (Fig. 2-7A). The KE994228 scaffold was not exhaustively searched so may contain additional genes.

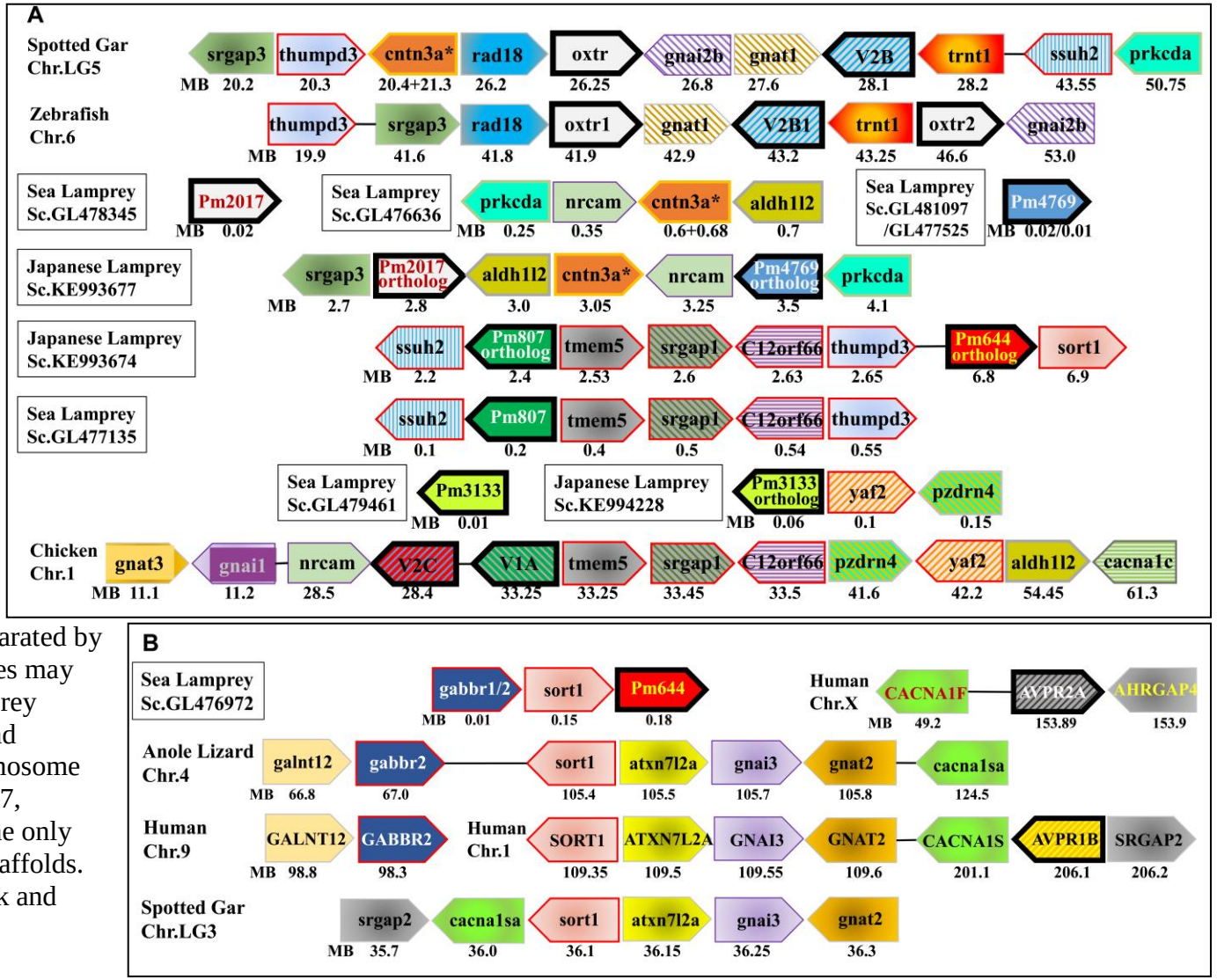


Figure 2-7. Synteny analysis comparing genes neighboring the oxytocin/vasopression receptor family loci (outlined in black) on sea lamprey (*Petromyzon marinus*) and Japanese lamprey (*Lethenteron japonicum*) scaffolds (Sc.) with those on jawed vertebrate species chromosomes (Chr.). Location of the gene starting near the 5'-end is given in megabases (MB), and genes oriented on the plus strand point right, minus strand left. Non-adjacent genes are separated by a black line; additional genes may also exist on Japanese lamprey scaffolds. **(A)** V1A-V2C and OXTR-V2B receptor chromosome blocks; sea lamprey Pm2017, Pm3133 and Pm4769 are the only genes on their respective scaffolds. **(B)** V1B chromosome block and V2A chromosome block.

Because it has been difficult to resolve orthology for many lamprey gene families with several paralogs (Qiu et al., 2011), genes may be misnamed in the databases. One example is the srgap/arhgap family. As shown in Fig. 2-7A, in gnathostomes srgap1 is reliably found near the V1aR gene (as exemplified in the chicken), srgap3 near the OXTR gene (as in the spotted gar and zebrafish), and arhgap4 near the V2aR gene (as in humans); Fig. 2-7B shows that srgap2 is near the V1bR gene (as shown in human chromosome 1). The ortholog near Pm807 in the Ensemble database is called arhgap4b, and there is a paralog called srgap1 found on scaffold GL476653. We constructed a limited phylogeny toward identifying these orthologs. Results were inconclusive, with very low support at the nodes where the lamprey arhgap4b and the other lamprey srgap gene separated the gnathostome srgap1 and srgap2 from the srgap3 and the arghap4 clusters (Suppl. Fig. 2-6, Appendix B).

The tmem5 gene has no paralogs in the lamprey, and has a one-to-one orthology with mammalian species in the Ensembl database. The tmem5 gene is found adjacent to srgap1 in all mammalian, avian, and several fish species included in the Genomicus v80.01 synteny database (http://www.genomicus.biologie.ens.fr/genomicus-80.01/cgi-bin/search.pl). Therefore, a srgap gene next to tmem5 is most likely a srgap1 ortholog. The C12orf66 gene also has no paralogs and is also commonly linked with tmem5 and srgap1.

Sea lamprey scaffold 644 (GL476972) contains a gabbr1 or 2 ortholog, sort1 and the Pm644 AVT receptor (Fig. 2-7B). The sort1 gene is often found adjacent to genes on the same chromosome carrying the V1bR gene. This and the gabbr 1 or 2 gene would

place Pm644 in a position that might have been occupied by the missing “V2X” gene (Fig. 2-8). However, since the Pm644 ortholog and the sort1 gene are found on the same scaffold with the Pm807 ortholog in the Japanese lamprey assembly, sort1 may be a different paralog than predicted or is misplaced. Additionally, the gabbr1/2 gene found on the sea lamprey Pm644 scaffold is not present on the orthologous Japanese lamprey KE993674 scaffold.

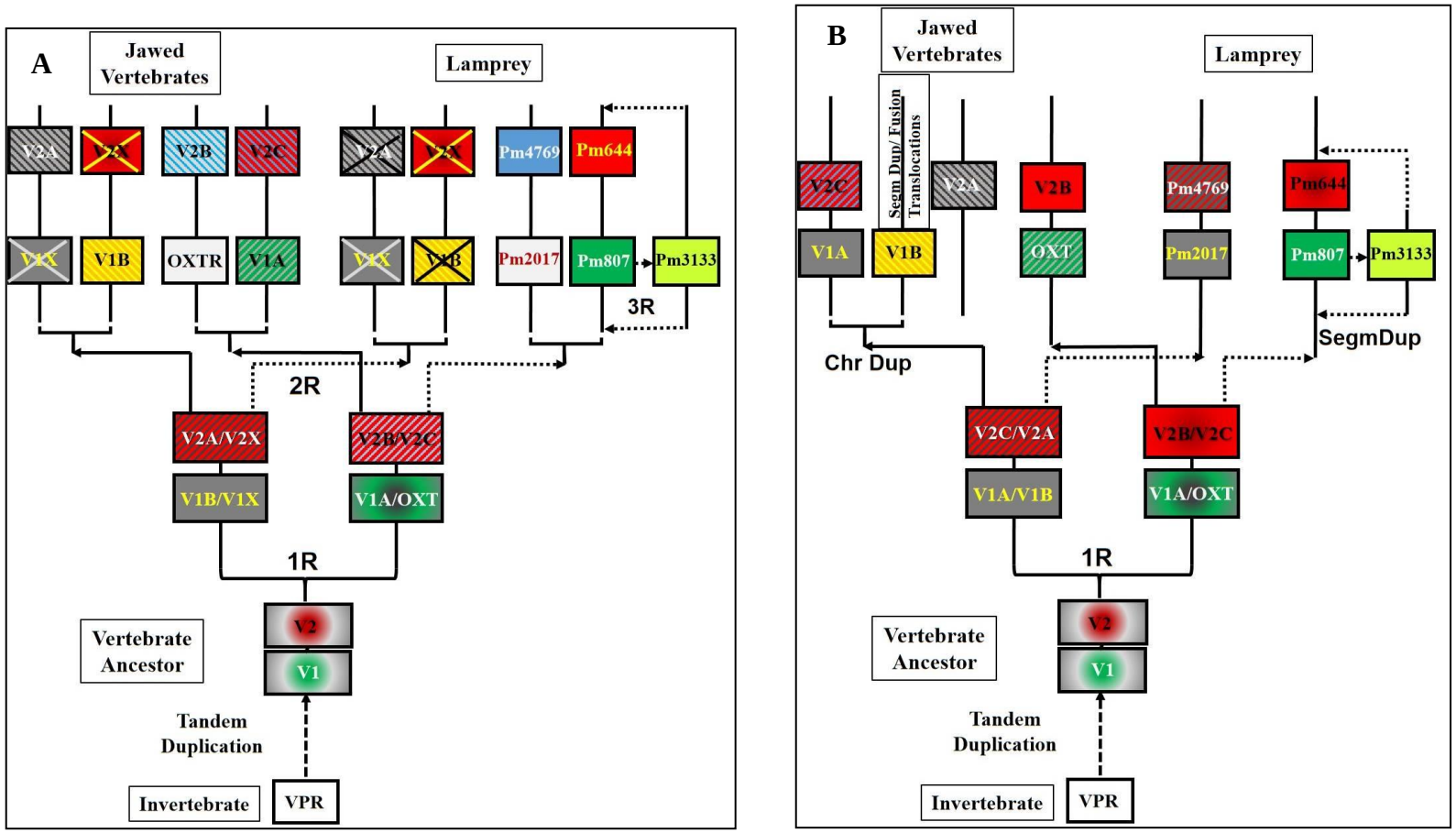


Figure 2-8. Postulated whole-genome duplication (WGD) evolutionary history scenarios for the AVT/AVP/OXT receptor family and possible orthology of the lamprey AVT receptor genes. The vertebrate ancestor carried a V1 and a V2 gene resulting from an invertebrate receptor gene tandem duplication before the first WGD (1R). **Scenario A:** lampreys diverged from the vertebrate lineage soon after 2R, and phylogeny and synteny suggest that the Pm3133 gene is the 3R duplicate of Pm807 with all other 3R genes lost (not shown). **Scenario B:** lampreys diverged after 1R of WGD and segmental duplications and translocations occurred independently.

2.4 Discussion

2.4.1 Lamprey AVT receptor gene structure and expression

The jawless vertebrates including the sea lamprey are key species in resolving WGD scenarios and the resulting orthological and paralogical relationships generated among the members of gene families conserved from invertebrates through the jawed vertebrate lineages. We have identified and sequenced one partial and four full-length putative AVT receptor genes in the sea lamprey and utilized mRNA tissue distribution, phylogenetic maximum likelood, and synteny analysis methods to aid in understanding AVP/OXT receptor family evolution.

The intron-exon structure of the sea lamprey putative AVT receptors is shown in Fig. 2-2 and may provide another clue to orthology, considering that amphioxus and humans share 85% of their introns (Putnam et al., 2008). The elephant shark AVT and OXT receptors each had only the one intron between the sixth and seventh TM domains (Yamaguchi et al., 2012), as is the case for the human AVPR1B, AVPR1A and OXT receptor genes (Fig. 2-2) and those of all types including V2b and V2c in most species (Ocampo Daza et al., 2012). The human AVPR2 gene has an additional intron in the coding region, and many fish studied have five exons and four introns in the coding regions of their V2a receptor genes. Introns in the 5'-UTRs of both Pm807 (Suppl. Fig. 2-2A) and Pm2017 (Suppl. Fig. 2-2C) parallel those found in the 5'-UTRs of human OXTR and AVPR1B genes. Thus, both receptor genes appear to have a conserved intron structure orthologous to the V1 and OXT receptor types, while the other receptor genes indicate a more divergent intron-exon structure. However, the conservation of the sixth-

to-seventh TM intron position in all receptors within this gene family from invertebrates (Kanda et al., 2005) through humans (Ocampo Daza et al., 2012) likely indicates the presence of regulatory elements under selection pressure (Henricson et al., 2010).

Although mRNA expression does not necessarily correspond to protein expression level and function due to post-transcription regulation and protein degradation activity (Vogel and Marcotte, 2012), gene transcription patterns themselves may be conserved among gene orthologs. In rats, OXT receptor mRNA was found to be widely distributed throughout the forebrain and more extensive than the V1b receptor which was restricted to discrete areas, especially the olfactory bulbs (Vaccari et al., 1998). The newt *Cynops pyrrhogaster* showed high expression of the V1a receptor in the hypothalamic nuclei and throughout the forebrain and midbrain, the V2a was not expressed in the hypothalamus but was moderately expressed in other forebrain and midbrain areas, and the V1b showed moderate expression in only the dorsal hypothalamic nucleus and parts of the midbrain (Hasunuma et al., 2010). Our analysis detected varying receptor expression from modest in the hindbrain, high in the forebrain, to most highly expressed in the hypothalamus for all receptors (Fig. 2-4B, 2-5; Suppl. Fig. 2-4A), and the AVT preprohormone was very highly expressed in all three areas (Fig. 2-4B).

Specifically, Pm644 mRNA was highly expressed in both males and females in the hypothalamus and much less in the forebrain (Fig. 2-5). As discussed for the newt and shown for the pupfish (Lema, 2010), V2a receptor expression is limited in all brain areas while V2b receptor expression was relatively high in the hypothalamus of the elephant shark (Yamaguchi et al., 2012). The Pm644 pattern seems therefore more V2b like, but

V2b and V2c expression has been studied in very few species. The other V2b/V2c-like gene, Pm4769, displayed a similar brain-hypothalamus pattern but was more generally expressed in peripheral tissues. The Pm807 gene was highly expressed in all brain areas but with less difference between the forebrain and hypothalamus than for the other receptors (Fig. 2-4B, 2-5; Suppl. Fig. 2-4A). Pm807 was also consistently highly expressed in the heart, pineal organ, gills and gonads of most or all animals (Fig. 2-4A and C; Suppl. Fig. 2-4A), and in the cloaca of the 2013 adults (Fig. 2-4C), with moderate expression in the kidneys and no expression in liver, consistent with known functions of both V1aR and OXTR (Table 2-1). Overall, Pm807 was the more highly expressed of the five lamprey receptors, the others showing little expression in the first round of PCR (Fig. 2-4C). In comparison, the V1a receptor was the more highly expressed of the elephant shark genes (Yamaguchi et al., 2012). The Pm3133 gene was expressed in similar tissues to Pm807 but visualization required two rounds of PCR. The Pm2017 V1a/OXT gene was expressed in the forebrain, hypothalamus, gonad, eye, gill, and weakly in the cloaca with no expression in the other tissues after nested PCR (Fig. 2-4B through D; Suppl. Fig. 2-4A). This limited pattern is similar to that of the isotocin receptor in the pupfish (Lema, 2010) but also reminiscent of the V1b gene in the newt (Hasunuma et al., 2007) and the chicken (Tan et al., 2000).

2.4.2 AVT receptor evolutionary history

The sea lamprey AVT receptor genes provide new information on the pathways of the AVP/OXT receptor family paralogs through one or two rounds of WGD (Fig. 2-6, 2-7 and 2-8). Ocampo Daza et al. (2012) and Lagman et al. (2013) proposed that a vertebrate

ancestral tandem gene duplication resulted in the divergence between the V1 and V2 receptor groups, and that two rounds of WGD resulted in six receptors on four chromosomes in the gnathostome ancestor: (1) V1b; (2) V2a; (3) V1a and V2c; and (4) OXT and V2b. The 2R WGD could have produced eight possible receptors on four chromosomes: V1b and V2X; V1X and V2a; V1a and V2c; and OXT and V2b. The "X" loci losses could have occurred between 1R and 2R or after the second WGD, because genes of this paralog are present neither in the lamprey nor any gnathostome species yet investigated. However, as noted in Chapter 1, new support has been found for a 1R hypothesis based on a sea lamprey meiotic linkage map (Smith and Keinath, 2015). Our hypothesized evolutionary history of the AVP/OXT receptor family attempts to discern the plausibility of 2R (Fig. 2-8A) or 1R (Fig. 2-8B) scenarios.

Only the V1a-V2c and the OXT-V2b chromosomes seem to remain in the lamprey, based on the neighboring genes occupying the orthologous Japanese lamprey Pm807-Pm644 scaffold and the Pm2017-Pm4769 scaffold, respectively. These Japanese lamprey scaffolds (and the sea lamprey Pm807 scaffold) encompass a mix of protein-coding genes homologous to those adjacent to gnathostome V1a-V2c genes exemplified by chicken chromosome 1 and the OXTR-V2b genes exemplified by spotted gar chromosome LG5 and zebrafish chromosome 6 (Fig. 2-7A). The neighboring genes on one scaffold do not duplicate those on the other. This suggests that, employing the rationale of Smith et al. (2013), a duplication of the vertebrate ancestor V1a/OXT-V2b/V2c chromosome led to paralogous formation of V1a and OXT receptor types, and that the losses of duplicate copies of neighboring genes from either chromosome differed between the lampreys and

gnathostomes. The lamprey could have lost the remaining V2a gene as well as the V1b gene, although Pm2017 phylogeny and synteny are contradictory. Based on synteny our findings suggest that the first WGD round in lampreys, or both lampreys and gnathostomes, resulted in one chromosomal block containing the V1b/V1X and V2a/V2X precursors, and one chromosomal block containing the V1a/OXT and V2b/V2c precursors (Fig. 2-8A). However, the orthology of possible post-2R WGD chromosomal blocks between lampreys and gnathostomes is uncertain, as exemplified by the decreased confidence values from the gnathostome-only phylogenetic tree (Fig. 2-6A) to the tree with lamprey receptor sequences added (Fig. 2-6B).

The neighboring genes also tend to be reshuffled or oriented in the opposite direction, as found for the lamprey Hox gene clusters by Mehta et al. (2013) who also note that the lampreys experienced a different history of secondary gene loss from that of humans. Adjacent to Pm807 are three V1a-linked genes in a row (tmem5, srgap1 and C12orf66), and because the Pm3133 gene ortholog has two V1a-chromosome neighbors (yaf2 and pzdrn4) (Fig. 2-7A) and is closely linked phylogenetically to Pm807 (Fig. 2-7B), perhaps this sways the argument toward Pm807 and Pm3133 being duplicate V1a paralog products of the theorized lamprey-specific third-round (3R) WGD (Mehta et al., 2013; Nah et al., 2014) (Fig. 2-8A).

Our schematic diagram (Fig. 2-8A) for the AVT receptor family in a 2R scenario is similar to the hypothesis for the GnRH receptor family (Sower et al., 2012) postulating that paralogs formed on two of the four sets of chromosomes were lost as agnathans diverged from the vertebrate lineage, creating a hidden paralogy (Kuraku, 2010).

Included in this hypothesis is a lamprey-specific duplication of the GnRH-2 and -3 receptors (Sower et al., 2012). Nah et al. (2014) postulate a Runx gene family duplication scenario in which three of the four possible paralogs after the 2R WGD were retained but all of the duplicates produced in the lamprey-specific 3R were lost. Similarly, the Pm3133 resulting from the lamprey-specific 3R chromosome duplication could have created a separate chromosome analogous to the V1a duplication arrangement on zebrafish chromosomes 4 and 25 (Lagman et al., 2013). A tandem duplication is also possible, analogous to the OXTR gene duplication arrangement seen in zebrafish chromosome 6 (Fig. 2-7A; Lagman et al., 2013). The Pm3133 gene appears to be the only duplicate retained of the eight additional AVT receptor genes that could have been produced through a 3R event. Convincing evidence from the Japanese lamprey Hox and Runx gene clusters has been put forth that agnathans not only experienced a lineage-specific 3R WGD but possibly that all three rounds of WGD in agnathans were independent of the 2R WGD that occurred in gnathostomes (Mehta et al., 2013; Nah et al., 2014). The results of our study do not refute this hypothesis; the dim orthology between lamprey and gnathostome AVP/OXT family receptors make it a plausible alternative.

An independent 2R scenario has similar aspects to a 1R scenario. In light of the lamprey meiotic map, and evidence indicating that any additional duplications could be lineage-specific (Smith and Keinath, 2015), the more likely events are proposed to be segmental duplications and translocations as depicted in Fig. 2-8B.The gene losses and syntenic arrangement comparisons between the lampreys and jawed vertebrates in several

families including AVT receptors tend to cast doubt on the possibility of 2R occurring in all vertebrates. Smith and Keinath (2015) provide near chromosome-scale support for the premise that lamprey and gnathostome chromosomes were derived from the same ancestral chromosomes, and subsequent divergence occurred through segmental duplications, fusions and translocations.

Phylogenetic tree topology shows that nucleotide substitutions have occurred at different points in the agnathan and gnathostome sequences. According to the shorter phylogenetic branch lengths, somewhat fewer substitutions have occurred in the lamprey V1a/OXT-like Pm807 and Pm3133 receptors than most other genes in the family (Fig. 2-6B). Also bearing on the trajectories of the V1a and OXT receptors is the fact that tandem duplication of the AVT hormone gene to create two hormone genes (AVT and OXT) occurred after agnathan divergence and is first seen in the cartilaginous vertebrates (Gwee et al., 2009). Therefore the new ligand specificity that ultimately selected for functionally separate V1a and OXT receptors in gnathostomes was not driving the fixation of nucleotide substitutions in agnathans. The most enigmatic of the genes is the Pm2017, with phylogeny leaning toward V1b, tissue expression V1b or OXTR-like, and synteny supporting a V1a/OXT orthology. Based on these three methods, the Pm807-Pm644 and Pm2017-Pm4769 chromosomes reflect both conserved clues to common origins, and a deeply divergent agnathan evolutionary trajectory from the rest of the vertebrate lineages.

Although we have not yet demonstrated that the sea lamprey putative AVT receptors are expressed as proteins and function in intracellular signaling, Uchiyama and

Murakami (1994) showed AVT-dose-dependent vascular constriction leading to increased glomerular filtration rate and diuresis in Japanese lamprey kidneys. A radioimmunoassay showed that circulating AVT levels matched the minimum effective concentrations. The dose-response relationship was also measured for dorsal aortic blood pressure, and the responses to AVT were abolished by an AVT antagonist that worked similarly on fish and bullfrog AVT receptors (Uchiyama and Murakami, 1994). The putative AVT receptors are nearly identical between the two lamprey species (Suppl. Fig. 2-3A through 2-3E). The receptors contain residues known to be involved in ligand binding and all but Pm644 possess the "DRY" motif known to facilitate G-protein coupling. It would seem that at least one of the putative AVT receptor genes is functional in the kidneys and vasculature in these species. The tasks remain to determine which of the receptors is responsible for these responses, and in what other systems the receptors may function.

2.5 Conclusions

Through mRNA expression patterning, phylogenetic evidence and gene synteny analysis we add agnathan branches to the AVP/OXT family of neuropeptide hormone receptors in vertebrate animals. Our data supports a one-round WGD scenario but does not refute a two-round WGD scenario in common between lampreys (with hidden paralogy) and gnathostomes, with the possibility of a lamprey-specific third round. Due to uncertain orthology however, three rounds of WGD in lampreys independent of the gnathostome 2R scenario remains a plausible hypothesis. Work is currently underway to demonstrate receptor function and the expression of sea lamprey receptor proteins in tissues as well as mechanisms of gene transcription regulation.

Chapter 3

Cloning and preparation for functional assay of putative sea lamprey vasotocin hormone receptors

3.1 Introduction

An important step in characterizing new-found hormone receptor genes is to determine whether the gene produces a protein capable of functioning in intracellular signaling. As described in the previous chapters, neuropeptide G protein-coupled receptors (GPCR) such as those of the oxytocin-vasopressin family change conformation when bound to their cognate hormone ligand, triggering the eventual release of second-messenger signaling components such as calcium or cyclic AMP within the cell. To perform the standard functional test, the full coding sequence of the gene from start codon to stop codon must first be inserted into a plasmid vector and cloned. The bacterial plasmid with the lamprey receptor gene insert, with the proper sequence confirmed, is then ready to be transiently expressed in a mammalian cell line. In this phase of my research I attempted to clone four of the sea lamprey putative vasotocin receptor genes characterized in Chapter 2: Pm807, Pm2017, Pm3133 and Pm644. Cloning was not attempted for the fifth receptor, Pm4769, for which the N-terminal region containing the start codon could not be found.

Cloning vectors can be chosen for specific purposes; a simple transcription vector is designed only to replicate the gene for sequencing, while an expression vector contains a promoter construct allowing for translation of the gene into a protein when transfected into the target mammalian cells. The pcDNA™ 3.1/V5-His TOPO® plasmid expression vector for topoisomerase (TA) cloning was chosen for this project, with features listed by the manufacturer (Invitrogen, Catalog no. K4800-01, 2009 user manual), and briefly described as follows. The vector contains a human cytomegalovirus (CMV) immediate-

early promoter/enhancer permitting expression of the protein in mammalian cells (for expression, the gene must contain the ATG start codon which is not included in the vector construct). Priming sites [T7 forward and bovine growth hormone (BGH) reverse] allow sequencing through the inserted gene. If the stop codon is not included in the inserted gene sequence, a V5-epitope (in-frame) is fused to the expressed protein to allow detection using an anti-V5 antibody. Also without the stop codon, the protein is tagged with C-terminal polyhistidines (H_6) to permit protein purification on a metal-chelating resin. However, in order to express the native protein for functional assays, the stop codon must be present and the V5 epitope and H_6 tag will not be included. An ampicillin resistance gene allows for positive selection of plasmid-containing *E. coli* colonies on agar plates containing the antibiotic. The bacterial colonies must then be tested using PCR to determine that the inserted gene is oriented in the 5'-to-3' direction.

This chapter describes the cloning process resulting in a successful clone of the Pm807 gene into the plasmid expression vector. However, difficulties most likely due to high guanine-cytosine (GC) content were encountered for the other receptor sequences. Therefore, expression and functional assays are not included here but research is currently ongoing in our lab.

3.2 Methods

3.2.1 RNA isolation and RT-PCR

Three large larval lampreys were euthanized in tricaine methanesulfonate (MS-222) at 0.1 g/L, pH 7.4. Brains were pooled and RNA was isolated using the TriZol® method according to the manufacturer's instructions and as previously described (Chapter 2). The RNA was treated with DNase I (DNA-*free,* Ambion) to remove genomic DNA. The treated RNA was reverse transcribed to cDNA using the Superscript III Supermix kit (Invitrogen, no. 18080-400) in 20 µl reactions according to the manufacturer's instructions.

3.2.2 PCR amplification of full coding sequences

The cDNA was PCR amplified using the Phusion® High-Fidelity DNA Polymerase kit (New England BioLabs) with either HF buffer for Pm807, or GC buffer and 1M betaine for the other receptors, with primer concentrations of 0.5 µM in 50 µl reactions. The thermocycler profile was 98 ºC for 1 min followed by 34 cycles of denaturation at 98 ºC for 10 sec, annealing at 60 ºC for 20 sec, and elongation at 72 ºC for 30 sec, followed by a final elongation step at 72 ºC for 10 min. Forward primers were designed to include the translation start codon and reverse primers to include the stop codon (Table 3-1).

Table 3-1. Primers for coding sequences of lamprey putative vasotocin receptor genes.

Gene transcript	Primer name	Base pairs	Amplicon % GC	Sequence (5'-3')
Pm807	Clone for	1326	59.6	CTGGAAGTGATTTTGGGAGGAAAAATGC
	Stop rev			AACGGCTCACGCAAAGTACGTCTT
Pm3133	Start for	1150	64.3	ATGGCCAACGGCACGGCGAAT
	Stop rev			ACCCCGCATCCTACTTCTGGCT
Pm2017	5TSS for	1440	70.8	TAATGCACTGCCCATAGGAGGAGG
	Stop rev			AGCGTTACCGTAGGGGCCATCGCA
Pm644	Start for	1274	67.4	CACGGGAACGATGAGTGACGATCCA
	Stop rev			TGGCTCAGGGCAGGCAGGACA

The PCR products were visualized on 1.5% agarose gels to determine proper size, and the gel bands were excised and DNA extracted using the QIAquick gel extraction kit (Qiagen). Because the Phusion polymerase product does not have the 3'-adenine overhang required for insertion into the plasmid vector used in this study, the overhangs were added by incubating 25 µl of extracted DNA with 25 µl GoTaq Green polymerase mastermix (Promega) at 72 ºC for 25 min. This 50 µl mix was than diluted to 100 µl with nuclease-free water and mixed with 100 µl phenol:chloroform reagent, vortexed and centrifuged to re-extract and concentrate the DNA. The aqueous phase was mixed with 10 µl 3M sodium acetate, pH 5.2, and precipitated with 95% ethanol. After centrifugation and brief drying the pellet was resuspended in 10 µl nuclease-free water.

3.2.3 Cloning

Cloning was carried out as follows for each receptor. The pcDNA™ 3.1/V5-His TOPO® TA expression kit was used according to the manufacturer's instructions, with 4 µl of the re-extracted DNA, 1 µl salt solution, and 1 µl TOPO® vector in the 6 µl cloning reaction. Two µl of the cloning reaction mixture was added to a vial of One Shot® TOP10 chemically competent *E. coli* cells and incubated on ice for 10 min. The cells were heat-shocked for 30 sec at 42 ºC and put back on ice. Room temperature SOC medium (250 µl) was added and the tube was placed in a shaker bath at 200rpm, 37 ºC for 1 hour. The cells were spread on pre-warmed LB-agar plates containing carbenicillin antibiotic and incubated overnight at 37 ºC. To determine correct orientation of the inserted gene, approximately one-half colony of each of 10 colonies were picked with a sterile transfer loop and placed in a PCR tube with a 25 µl reaction mixture of GoTaq

Green PCR mastermix containing T7 primer (5'-TAATACGACTCACTATAGGG-3') for the forward priming site on the TOPO® vector. After PCR and gel visualization, the other half of colonies with appropriate PCR bands were placed in a 15-ml screw-capped tube with 3 ml LB broth containing 1 µl/ml carbenicillin and incubated in a shaker bath at 37 ºC for 18 hours. Cells were harvested and TOPO® vectors were purified using the QIAprep Spin MiniPrep kit (Qiagen). Plasmids were Sanger sequenced at the University of Minnesota Genomics Center. Custom sequencing for the Pm3133, Pm2017 and Pm3133 genes included 1M betaine and reactions run for 40 cycles.

3.3 Results

3.3.1 Establishment of conditions for PCR amplification of lamprey cDNA

Before each gene could be inserted into the plasmid vector, a major hurdle concerning lamprey genes was that PCR amplification of the cDNA was difficult to achieve. Primers had to be designed to amplify segments of each gene in order to "walk" through the full coding sequence, confirm the intron locations and determine whether the sequences matched those found in the lamprey genome database. This was not difficult for Pm807, the first receptor gene tried. However, initial attempts to amplify any of the other receptor genes failed using the same conditions. Several sets of primers were tried and the PCR conditions were adjusted, modifying the annealing temperature or using a "touchdown" PCR profile with only moderate success for Pm2017 and no amplification of Pm3133 or Pm644. My realization that high guanine-cytosine (GC) content may be the cause of the difficulty came in early 2012, prior to the reference to high GC content published in the lamprey genome paper (Smith et al., 2013). The first research on the GC-rich nature of lamprey genes was published in 2011 (Qui et al., 2011), but did not contain any information on PCR technique. Troubleshooting techniques suggested by PCR kit manufacturers include dimethylsulfoxide (DMSO) which I tried at various concentrations with no improvement. A PCR kit for high GC template cDNA was purchased (Amplitaq Gold, Applied Biosystems) with an "optimized" buffer containing DMSO and glycerol, which was also unsuccessful. Other compounds tested unsuccessfully included formamide (Sarkar et al., 1990) and trehalose (Spiess et al., 2004). Finally, a proprietary enhancer yielded positive results (Q-solution, Qiagen; later found to be 5M betaine, also

available from Sigma). Addition of betaine to the PCR mix allowed the successful amplification of the Pm2017, Pm3133 and Pm644 genes for visualization of clear bands on agarose gels, gene sequencing, tissue expression distribution analysis and synthesis of the full coding sequences for cloning attempts.

3.3.2 Sequencing of the cloned plasmids

Sequencing of the inserted genes in the plasmids was necessary to confirm the correct receptor sequences. The Pm807 sequence matched the sequence we deposited in the GenBank database except for two bases, which were synonymous base substitutions that did not change the predicted amino acid composition of the protein (data not shown). However, after two additional rounds of cloning attempts, only five colonies for Pm2017, two for Pm3133 and one for Pm644 displayed bands appropriate for sequencing. All of these clones contain deletions or residue changes in the amino acid sequences. For Pm3133, three amino acid changes were predicted (Fig. 3-1) based on the nucleotide sequence, and for Pm644, five amino acid changes were predicted (Fig. 3-2). The nucleotide sequences of all five Pm2017 clones were similar to each other but contained five multiple-base deletions dominated by G- and/or C-rich sequences, and several single-base substitutions, nearly all involving a guanine or cytosine (Fig. 3-3). Because of these sequence anomalies, no attempts were made to transfect the Pm3133, Pm2017 or Pm644 plasmid clones into mammalian cells, however the Pm807 plasmid clone is ready for transfection.

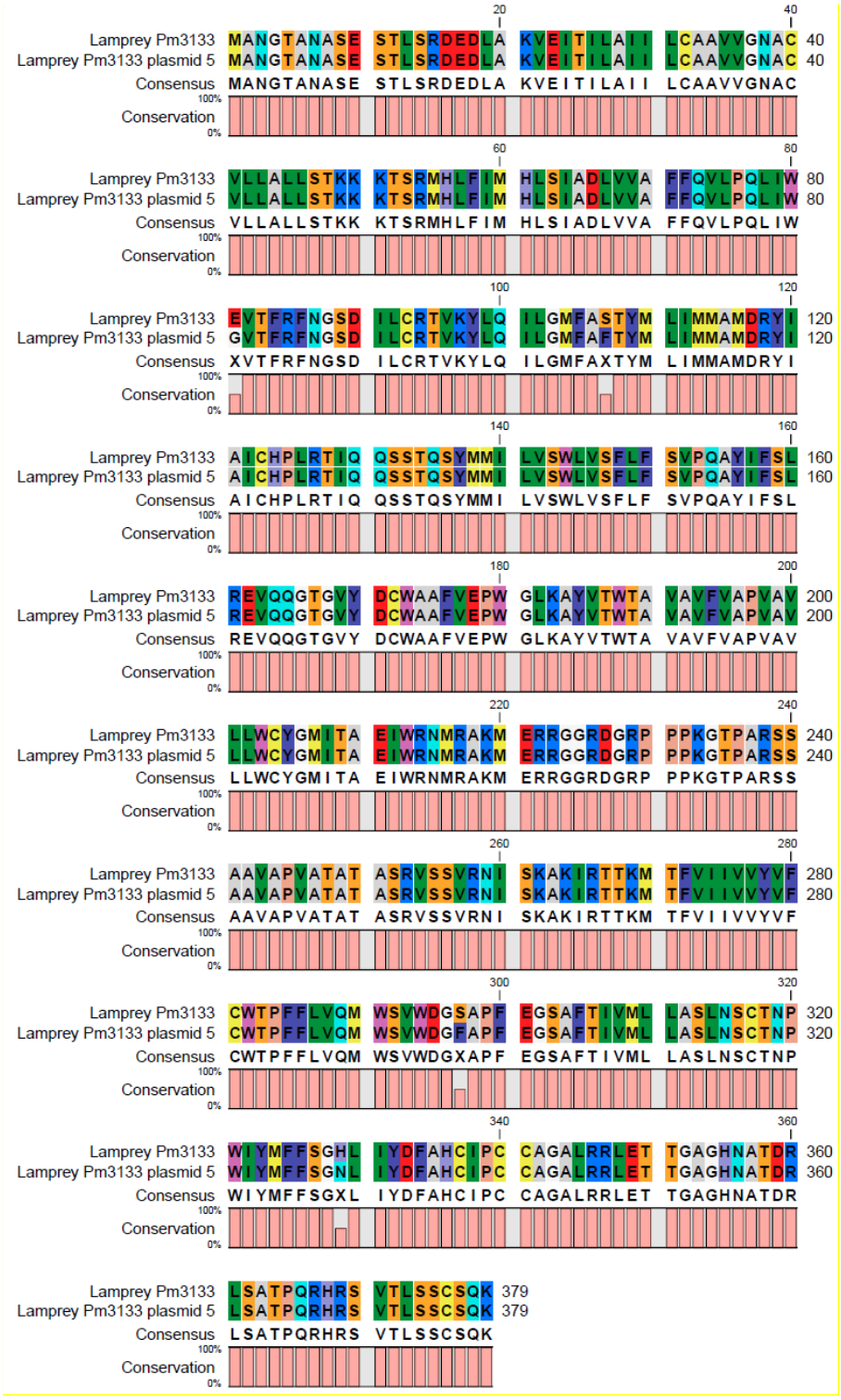


Figure 3-1. Predicted amino acid sequence of the most similar Pm3133 plasmid (5) to the translation of the full coding sequence deposited in the GenBank database (KJ813004).

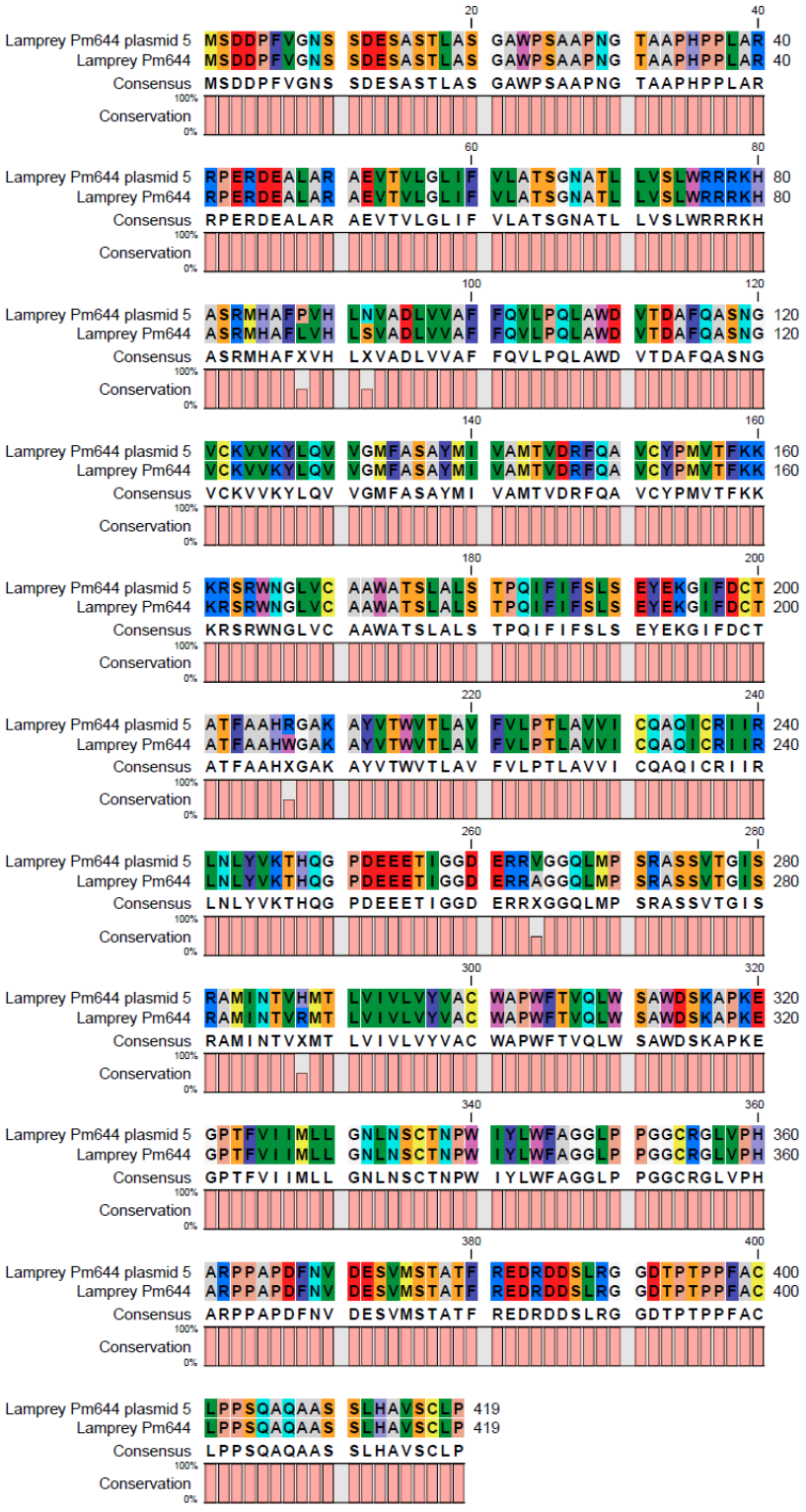


Figure 3-2. Predicted amino acid sequence of the single Pm644 clone (plasmid 5) to the translation of the full coding sequence deposited in the GenBank database (KF031008).

Figure 3-3. Nucleotide sequence of Pm2017 clone (continued on next page).

```
Pm2017_C1           ATGGATCTCGCGCCGTGGAGGAACGCCAGCCGGACGCCCCGCCACGACAACATTTCCCAC
Pm2017_KJ813003     ATGGACCTCGCGCCGTGGAGGAACGCCAGCCGGACGCCCCGCCACCACAACGTCTCCCAC
                    ***** ************************************** ***** * ******

Pm2017_C1           GCGGCTCCCAACGACACCGATCGCCACGAGGGCACGGCGGCGGCGGC---------ACGC
Pm2017_KJ813003     GCGGCTCCCAACGCCGCCGACCGGCCCGAGGACGCGGCGGCGGCGGCGGCGGCGGCACGC
                    ************* * **** ** * ***** * *************         ****

Pm2017_C1           GACGAGTGTCTCGCCGCGGCGGAGATCGCCCTGCTCGCCGCCATCGTCGCCGTCGCAATC
Pm2017_KJ813003     GACGAGCGCCTCGCCGCGGCGGAGATCGCCCTGCTGGCCGCCATCGTCGCCGTCGCCATC
                    ****** * ************************** ******************** ***

Pm2017_C1           GCCGGCAACGGTTCGGTGCTGCTAGCGCTGAGCCGCACGCGCTGCAAGGCGTCGCGCATG
Pm2017_KJ813003     GCCGGCAACGGCTCGGTGCTGCTGGCGCTGAGCCGCACGCGCCGCAAGGCGTCGCGCATG
                    *********** *********** ****************** *****************

Pm2017_C1           AACCTCTTCGTCAAGCACCTGAGCGTCGCCGACCTGGCCGTGGCCGCCTTCCAGGTTCTG
Pm2017_KJ813003     AACCTCTTCGTCAAGCACCTGAGCGTCGCCGACCTGGCCGTGGCCGCCTTCCAGGTCCTG
                    ************************************************************ ***

Pm2017_C1           CCGCAGCTCTCGTGGGACGTGACCTTCCGCTTCCACGGGCCTGACGCGCTCTGCCGCCTC
Pm2017_KJ813003     CCGCAGCTGTCGTGGGACGTGACCTTCCGCTTCCGCGGGCCCGACGCGCTCTGCCGCCTC
                    ******** ************************* ****** ******************

Pm2017_C1           GTCAAGTACCTGCAGGTGGTGGGCATGTTCGCGTCCACCTACGTGCTCATCGCGATGACC
Pm2017_KJ813003     GTCAAGTACCTGCAGGTGGTGGGCATGTTCGCGTCCACCTACGTGCTCATCGCCATGACC
                    ************************************************** ******

Pm2017_C1           GTCGACCGTTACCTCGCCATCTGCCACCCGCTGAGCACGCTGCGTCGACGGGGCGAGCCG
Pm2017_KJ813003     GTCGACCGCTACCTCGCCATCTGCCACCCGCTGAGCACGCTGCGCCGGCGGGGCGAGCCG
                    ******** ********************************** ** ************

Pm2017_C1           AGGAAGCAGGCGCACGTCATGGTGGCCTGCGCGTGGGTGCTGAGCGCCCTGCTCAGCACG
Pm2017_KJ813003     AGGAAGCAGGCGCACGCCATGGTGGCCTGCGCGTGGGCGCTGAGCGCCCTGCTCAGCACG
                    ***************** ******************** ***********************

Pm2017_C1           CCGCAGGTCGTCATCTTCTCGCTGCAGGAGGTGGAGGAGGGGCGTCTTCGACTGCTGGGCC
Pm2017_KJ813003     CCGCAGGTGGTCATCTTCTCGCTGCGGGAGGTGGAGGAGGGGCGTCTTCGACTGCTGGGCC
                    ******** ***************** ***********************************

Pm2017_C1           GACTTCGGGCAGCCGTGGGGCATGTGGGCGTACGTGACGTGGATCACGGTGTCCGTGTAC
Pm2017_KJ813003     GACTTCGGGCAGCCGTGGGGCATGCGGGCGTACGTGACGTGGATCACGGTGTCCGTGTAC
                    ************************ ***********************************

Pm2017_C1           GTGGCGCCCGTGCTCATCCTGGCGGCGTGCTACGGGGCCATTCTGCTCGAGATATGCAGG
Pm2017_KJ813003     GTGGCGCCCGTGCTCATCCTGGCGGCCTGCTACGGGGCCATTCTGCTCGAGATTTGCAGG
                    ************************** ************************** ******

Pm2017_C1           AACCTGCGCCTCAAGACGCGTGGCCGATACCCCAAGGACCCCCGGGGCTCGCAGTCTCAA
Pm2017_KJ813003     AACCTGCGCCTCAAGACGCGTGGCCGATACCCCAAGGACCCCCGGGGCTCGCAGTCTCAA
                    ************************************************************

Pm2017_C1           TCGCCCGTGTGCTCGCGCGTGAGCAGCGTGCGCTGCATCTCCCGCGCCAAGGTGCGCACC
Pm2017_KJ813003     TCGCCCGTGTGCTCGCGCGTGAGCAGCGTGCGCTGCATCTCCCGCGCCAAGGTGCGCACC
                    ************************************************************

Pm2017_C1           ATCAAGATGACGCTCGTCATCATCGTGGTCTACGTCGTCTGCTGGACTCCCTTCTTCGTC
Pm2017_KJ813003     ATCAAGATGACTCTCGTCATCATCGTGGTCTACGTCGTCTGCTGGACTCCCTTCTTCGTC
                    *********** ************************************************

Pm2017_C1           ATCAGGATGTGGGCGGCCTGG-ACGAGACCGCACCAGACGACGACTCATCCGACCCCACC
Pm2017_KJ813003     ATCCAGATGTGGGCGGCCTGGGACGAGACCGCACCAGACGACGACTCATCCGACCCCACC
                    ***  *************** ************************************
```

```
Pm2017_C1           TACACCATCGTGATGGTCCTTTCGAGCCTCCACAGCTCCCCCAACCCGTGGATCTACATG
Pm2017_KJ813003     TACACCATCGTGATGCTCCTTTCGAGCCTCAACAGCTGCACCAACCCGTGGATCTACATG
                    *************** ************** ****** * ********************

Pm2017_C1           TCGTTCAGCGGGCACCTCCTCGAGGAGGTTCTGGCGTGCTGCCGCCAGGTGCGGTGCCCT
Pm2017_KJ813003     TCGTTCAGCGGGCACCTCCTTGAGGAGGTGATGTCCTGCTGCCGCCAGGCGCGGTGCGCC
                    ******************** ********  ** * ************* ******* *

Pm2017_C1           GCCACGTGCCGCCAGCAGC------GGCGGCGGCGGTGCCGAACGCCGACGAGGCGCGGG
Pm2017_KJ813003     ACCGCGTGCCGCCGGCAGCAGCGGCGGCGGCGGCGGTGCCGAACGCCGACGAGGCGCGGG
                     ** ********* *****      **********************************

Pm2017_C1           GCGGGCTCGCCCAACACGGAGCACGTGGTC---TACGCGCTGTCGCCGGCACCTACGCAG
Pm2017_KJ813003     GCGGGCTCCCCCAACACAGAGCACGTGGTGGTGTACGCACTGTCGCCGGCACCCACGCAG
                    ******** ******** ***********    ***** ************** ******

Pm2017_C1           TGCCGCAGCTCACCCGCCGCTGTCGC------CACCGCCACCCCC---GAGACGCGAAGA
Pm2017_KJ813003     TGCCGCAGCTCCCCCGCCACCGCCGCAGGTGGCACCGCCACCCCCCCCGAGACGCGGAGA
                    *********** ****** * * ***      *************   ******** ***

Pm2017_C1           TCCGCAGCGCCGCTGTGGTCGCCAGCGTTACCGTAGG
Pm2017_KJ813003     TCCGCAGCGCCGCTGTGGTCGCCAGCGTTACCGTAGG
                    ************************************
```

Figure 3-3. Nucleotide sequence of the most similar Pm2017 clone (C1) to the full coding sequence deposited in the GenBank database (KJ813003). Start codon **ATG** is in bold, stop codon *TAG* in italics. Asterisks (*) indicate identical bases. The differing bases involve a guanine or cytosine and the missing segments are dominated by G- and/or C-rich sequences.

3.4 Discussion

Cloning of four of the five putative AVT receptors characterized in Chapter 2 was attempted for the purpose of determining intracellular signaling function. The Pm807 putative V1A/OXT receptor gene was successfully cloned with a sequence matching that determined previously (Chapter 2) and deposited in GenBank. However, all of the other three putative receptor clone sequences had predicted amino acid substitutions (Fig. 3-1 and 3-2). Therefore, expression of these clones in target cells may not yield proteins that would be properly inserted in the cell membrane, bind with the ligand or interact with the G-protein complex. The cloning problem seems to stem from the high-GC sequences, especially Pm2017 (Fig. 3-3). Polynucleotide G or C sequences, or simple repeats can cause polymerase slippage. The accuracy of DNA transcription has been found to be roughly correlated with GC-content, and errors occur at the highest frequency in sequence motifs beginning with triplet combinations of G and C (Mellenius and Ehrenberg, 2015). Although betaine has worked in PCR amplification of shorter sequences, apparently additional measures need to be taken for the full-length coding sequences, possibly at the level of reverse transcription of the mRNA. With the new information gained from the Japanese lamprey genome database, primers could also be designed to find the start codon for the Pm4769 gene so that this V2b/c-like gene could then be cloned.

In addition to resolving the clone sequence issues, future completion of the functional studies will require a test to confirm the efficiency of transfection and translation in the target cells. We now have specific antibodies that can be used to detect

the predicted Pm807, Pm2017 and Pm644 proteins by extraction from the cells and subsequent Western blotting. If the antibodies are not effective, and for Pm3133 and Pm4769, primers could be designed to utilize the V5-epitope and H6 features of the expression vector noted in section 3.3.1. Also, transfection efficiency can be monitored by co-transfection with a green fluorescent protein marker. Mammalian cell lines and also oocytes from African clawed frogs (*Xenopus* sp.) have been used for non-mammalian AVT-family GPCR studies, including chicken (Cornett et al., 2003), frog (Kohno et al., 2003), teleost fish (Konno et al., 2010), lungfish (Konno et al., 2009), shark (Yamaguchi et al., 2012), octopus (Kanda et al., 2005) and nematode (Beets et al., 2012). Mammalian cells have also been used to study lamprey gonadotropin releasing hormone (GnRH) (Joseph et al., 2012) and melanocortin receptors (Haitina et al., 2007).

Although types of functional assay methods vary, often employing either fluorescence, luciferase or radioactive ligands, the ligand concentrations for the above cited AVT/OXT receptor family studies ranged from 0.001 nM to 100 or 1000 nM. For the native ligand-receptor combinations, in the chicken study activity increased at each level including the 1000 nM level; in the octopus, nematode and frog studies activity leveled off between 100 nM and 1000 nM; in the shark study the highest ligand concentration tested was 100 nM, which also showed leveling resulting in sigmoid curves; and in the fish and lungfish studies, receptor activation at 1000 nM dropped below the peak levels observed at 100 nM.

Many of the studies cited use Fluo-4 dye in a FLIPR® Tetra (Molecular Devices) instrument which is a plate reader with simultaneous pipet and read capability, important

in calcium mobilization assays where signals occur quickly and transiently. Plate-reader methods are high-throughput but with significant accuracy limitations, so a trade-off is that while single-cell fluorescent microscopy methods are slow, they are highly accurate (Heusingveld and Westerink, 2011). Flow cytometry methods are a multi-cell but single-sample alternative currently being explored in our lab. Also, because the Pm644 and Pm4769 genes are in the V2-receptor-like clade (Chapter 2), a cyclic AMP assay is needed to determine which second messenger system is used by these receptors.

In summary, further cloning attempts and techniques are needed to overcome the GC-content problems, especially for the 70%-GC lamprey Pm2017 gene; the lamprey antibodies need to be validated for transfection verification; and calcium mobilization and cAMP assay methods must be established to complete functional studies of these putative receptors. In addition to testing AVT, OXT and AVP ligand specificity, antagonists such as the non-peptide Vaptans (Decaux et al., 2008) can also be tested for their ability to block receptor function.

Chapter 4

Primary culture of sea lamprey immune cells and preliminary demonstration of vasotocin hormone and receptor gene expression

4.1 Introduction

Neuroendocrine-immune interactions and parallel functions are important in understanding behavior, disease and evolution. The expression and function of neuroendocrine hormone systems have been demonstrated in mammalian immune cells (e.g., Akbulut et al., 2001; Chen et al., 2002; Baurenhofer et al., 2003; Hansenne et al., 2004; Ndiaye et al., 2008; Loser et al., 2010; Koike and Seki, 2013; Poutahidis et al., 2013). Few similar studies have been conducted in lower vertebrates (Shaik Mohamed and Khan, 2006; Yada et al., 2004) and invertebrates (Zhang et al., 2000). Therefore, more knowledge of the neuroendocrine-immune system interactions in non-mammals is needed to understand their broader impacts on overall fitness. Methods that could be used to study both systems in concert would benefit not only an evolutionary perspective but also practical applications in animal care and husbandry.

Factors such as disease (Reversi et al., 2006), stress (Kuenzal et al., 2013), exposure to toxic chemicals (Wolstenholme et al., 2012; Rosenfeld, 2015), genetic variation (Donaldson and Young, 2013) and epigenetic modifications (Mamrut et al., 2013) affect neuropeptide hormone and receptor expression both peripherally and centrally. Immune cell sampling and primary culture can be used as a cost-effective surrogate for more invasive methods, to potentially monitor the effects of stress, disease and toxins on both immune and neuroendocrine endpoints. In the case of the lamprey, neuropeptide receptor ability to initiate intracellular signaling could be studied in the lamprey's own cells rather than a mammalian cell line, and antagonists to block the system could be tested.

Challenges have been encountered when studying immune cells for the purposes of understanding the function and characteristics of the immune system itself in species like the lamprey that do not have immortalized cell lines. Lamprey immune cell studies have required that the cells be stimulated by a series of antigen injections into ammocoete larva over a period of weeks (Pancer et al., 2004; Guo et al., 2009). The blood volume yield per animal is small, on the order of 50 microliters (µl), and cells must be immediately analyzed. Adult and parasitic-phase lampreys afford much larger blood volumes of one to more than four milliliters (ml), depending on the size of the animal. Blood can be collected shortly after receiving the animals from the supplier, without need for high-maintenance larval cultivation in aquariums for long periods.

Primary immune cell culture methods for economically important fish species (Miller and Clem, 1988; DeKoning and Kaattari, 1992; Yada et al., 2004) have been developed, demonstrating that teleost lymphocytes can be stimulated and proliferate *in vitro*. I modified the fish techniques to attempt the primary culture of lamprey immune cells, which have never before been cultured *in vitro*. Here I present a preliminary demonstration of neuropeptide hormone and receptor gene expression in freshly isolated and mitogen-treated cultured lamprey immune cells.

4.2 Materials and Methods

4.2.1 Animals

Sea lamprey adult upstream migrants were obtained directly from US Fish and Wildlife Service (USFWS) personnel from barrier traps on the Middle River near Lake Superior in Wisconsin or from USGS Hammond Bay Biological Station (HBBS) (Millersburg, MI). Larvae were also obtained from HBBS and were maintained in sand-lined zebrafish breeder boxes within 10-gallon aquarium tanks at 15 ^{o}C in a temperature-controlled room at the University of Minnesota-Duluth. A diet of moist cake yeast slurry in twice weekly feedings was provided according to the methods of Hansen et al. (1974). Details of the larval cultivation method are included in Appendix A. Adult lampreys used in the experiments were held without feeding in a 100-gallon stock tank at 15 ^{o}C for a maximum of two weeks after receipt. Animals were euthanized in tricaine methanesulfonate (MS-222, Sigma, St. Louis, MO), pH 7.4, at 2 g/L for adults or 1 g/L for larvae. Animal treatment conformed to University of Minnesota animal care standards (IACUC protocol number 1305-30612A).

4.2.2 Larval blood and tissue collection and RNA isolation

In two separate trials, three larval lampreys were euthanized and blood and positive control tissues were collected. Blood and each tissue type were separately pooled. Blood was collected by severing the tail and draining the blood into 1 ml 0.6X Dulbecco's phosphate buffered saline (PBS) with 20 mM ethylene diamine tetraacetic acid (PBS-EDTA), pH 7.4. Peripheral blood mononuclear cells (PBMC) were isolated similarly to the method of Shaik Mohamed and Khan (2006). Briefly, blood was diluted to 5 ml with

PBS-EDTA, carefully layered onto 5 ml Histopaque 1077 in a 15 ml nuclease-free centrifuge tube and centrifuged at room temperature at 400 g for 40 min. The interface containing lymphocytes was aspirated and washed with PBS-EDTA by centrifugation at 260 g for 10 min at 4 ^{o}C. The cell pellet was re-suspended in Trizol reagent (Invitrogen) for total RNA isolation. Larval cells were not cultured due to the small blood volume obtained per animal. In *Trial 1* the cranial area including the brain was excised with a sterile scalpel and placed in 1ml Trizol. In *Trial 2* the cranial, spinal cord and gill tissues were similarly collected. Brain and gill tissues are expected to display AVT and Pm807 V1a/OXT-type putative vasotocin receptor expression (Chapter 2). GnRH was expected in brains and has been identified in spinal cord of other species. Tissue was homogenized to isolate RNA according to manufacturer's instructions and as described in Chapter 2.

4.2.3 Culture medium and mitogen preparation

Culture medium was formulated based on the methods of DeKoning and Kaattari (1992) using RPMI-1640 tissue culture medium with L-glutamine (Sigma, Cat. No. R6504), supplemented with 2 g/L Na_2HCO_3, 10 ml/L penicillin-streptomycin solution (Sigma, Cat. No. P0781), and 5% fetal calf serum (FCS). Because no CO_2 incubator with low-temperature control was available, the cultures were incubated in the 15 ^{o}C temperature-controlled room, and pH was maintained by adding 20 mM HEPES buffer. This "incomplete" medium was filter sterilized and provided the base formula used for diluting, isolating and washing the cells, and dissolving the mitogens. Each mitogen was separately dissolved in 1 ml incomplete medium: 1.5 mg phytohemagglutinin-P (PHA-P), 1 mg pokeweed mitogen (PWM), 2 mg lipopolysaccharide (LPS; *E. coli* 055:B5 strain).

These mitogen stock solutions were heat pasteurized at 70^{o}C for 1 hour.

4.2.4 Adult PBMC isolation and culture

After an adult lamprey was euthanized for 15 min in MS-222, blood was drawn via heart puncture with a syringe containing anti-coagulant, either PBS-EDTA or acid citrate-dextrose (ACD) buffer, resulting in a blood:buffer dilution of 2:1. Total blood volume was measured, the blood was layered on an equal volume of Histopaque 1077, and centrifuged at 18 ^{o}C for 30 min at 400 g. The serum was drawn off, further diluted to 20% in incomplete culture medium and heat pasteurized for 30 min at 60 ^{o}C. Cells were aspirated and washed with incomplete culture medium. The stimulation medium for culture was made in 1-ml aliquots for culture in 12-well plates. All treatments for all experiments received pasteurized lamprey serum to a final concentration of 2%.

Final concentrations of mitogens were 15 µg/ml PHA-P, 100 µg/ml LPS, and 10 µg/ml PWM, distributed in different treatment combinations as shown in Table 4-1. An aliquot of 10^{6} cells was distributed in each treatment after counting and viability assessment by trypan blue exclusion (viability>95%) in a hemacytometer. Control aliquots were placed directly in TriZol® reagent for total RNA isolation as described in section 4.2.2.

In a separately incubated set of cells from a male lamprey after six days of incubation, duplicate control and mitogen-treated cells (PHA+LPS+PWM) were counted on a hemacytometer using trypan blue exclusion to determine the percentage of live and dead cells (Strober, 1997). Trypan blue dye is excluded from cells with intact membranes (live), but enters cells with disrupted membranes (dead), staining them blue.

Table 4-1. Lamprey peripheral blood mononuclear cell (PBMC) treatments for gene expression experiments.

Sample	Treatment	Gender	Time (days)
CM0	Control	Male	0
SM3	PHA/LPS	Male	3
SM6	PHA/LPS	Male	6
SM6d	PHA/LPS dup	Male	6
CF0	Control	Female	0
SF1	PHA/LPS/PWM	Female	1
SF6	PHA/LPS/PWM	Female	6
SF6-L	PHA/PWM	Female	6

C=Control; S=Stimulated; M=Male; F=Female; d=duplicate; -L=no LPS used; 0, 1, 3, 6=time of incubation in days. PHA-P=phytohemagglutinin-P, PWM=pokeweed mitogen, LPS= lipopolysaccharide (*E. coli* 055:B5 strain).

4.2.5 Reverse transcription - polymerase chain reaction (RT-PCR)

For larval *Trial 1*, 1400 ng cranial tissue mRNA or 300 ng PBMC mRNA were reverse transcribed to complimentary DNA using ThermoScript^{TM} RT (Invitrogen) in a 20 µl reaction according to the manufacturer's instructions. For *Trial 2*, 300 ng of mRNA from each tissue were reverse transcribed. PCR was performed in 50 µl reactions using Platinum PCR Supermix with 2 µl cDNA and a 0.5 µM final primer concentration. Thermal profile was: 2 min at 94º followed by 35 cycles of 30 s at 94 ºC, 30 s at 60 ºC, and 60 s at 72 ºC, with a 7 min elongation step at 72 ºC.

For adults, total RNA was treated with DNase I (DNA-*free*, Ambion, Austin, TX) to remove genomic DNA according to the manufacturer's instructions. For samples SM6, SM6d, CF0, SF1 and SF6-L, approximately 500 ng mRNA were reverse transcribed to cDNA, and for sample CM0 800 ng were reverse transcribed using SuperScriptIII First-

Strand Synthesis Supermix (Invitrogen) in 20 µl reactions. Low concentrations of mRNA in samples SM3 and SF6 allowed only approximately 50 ng to be reverse transcribed. PCR was performed in 50 µl reactions using Platinum PCR Supermix with 3 µl cDNA and a 0.5 µM final primer concentration. Negative controls consisted of the PCR mix with water instead of cDNA. Thermal profile for Pm807 was: 2 min at 95 ºC followed by 36 cycles of 30 s at 95 ºC, 30 s at 58 ºC, and 60 s at 72 ºC, with a 5 min elongation step at 72 ºC Thermal profile for β-actin was: 2 min at 94 ºC followed by 24 cycles of 30 s at 94 ºC, 30 s at 55 ºC, and 60 s at 72 ºC, with a 5 min elongation step at 72 ºC. Primers are listed in Table 4-2.

Table 4-2. Primers for neuropeptide hormone and receptor gene expression. β-Actin was used as a positive reference gene.

Gene		Primer sequences 5'-3'	Size (bp)
AVT	For	TGCGGAAGCGATGCTGGAACTGAT	201
	Rev	AACTCATTATACCAACGGGACGCC	
AVT	For	TGTGCATACTGTTCTAGTCGGCGT	668
	Rev	CAACCATGCACCTGGTTCTTAGCA	
GnRH-1	For	AGCTCTGAAGTAAGACCCACGCAT	280
	Rev	ACAGACAGGACAGCCACTCTGAAA	
Pm807	For	AAGGCGTACATCACCTGGATGACA	285
	Rev	GACTGACCACATCTGCACGAAGAA	
Pm807	For	AGTCGAGATCGCCTTACTCTCGAT	991
	Rev	ACTCACGAACGTGTGATTCCG	
β-Actin	For	ATCATGTTCGAGACCTTCAACACGC	721
	Rev	ACTCCTGCTTGCTGATCCACATCT	

Products were visualized on 1% agarose gels and purified using the QIAquick® PCR Purification Kit (Qiagen), eluted with nuclease-free water. The purified products were sequenced at the University of Minnesota Genomics Center using an Applied Biosystems (ABI) 3730xl sequencer (St. Paul, MN).

4.3 Results

4.3.1 Neuropeptide hormone expression in larval PBMCs

Freshly collected larval PBMCs were analyzed using RT-PCR to show that both AVT and GnRH-1 mRNA were expressed without *in vitro* mitogen treatment. Bands visualized on agarose gels were of the expected size for AVT and GnRH-1 amplicons in both the positive control cranial tissues and the PBMCs (Fig. 4-1). Neither the cranial tissue, spinal cord tissue, nor PBMC cDNA showed amplification of the Pm807 V1a/OXT receptor in Trial 2, though gill cDNA produced a strong Pm807 band. Pm807 receptor and AVT amplification may have been reduced in Trial 2 by the use of the longer, less efficient 991-bp and 668-bp amplicon sizes, respectively.

Interestingly, I found an indication of differing allele expression for the AVT gene. The sequence amplified is in the 3'-untranslated region of the gene (see Supplementary Fig. 2-3F). The AVT sequence of the *P. marinus* larval PBMC purified PCR product matched sequences from the *Lethenteron japonicum* genome and the GenBank AVT hormone precursor sequence for *Lethenteron camtschaticum (japonicum)*, which all differed slightly from the *P. marinus* expressed sequence tag (EST) found in GenBank and the sequence in the *P. marinus* genome. The *L. japonicum* and *P. marinus* sequences obtained using the blastn and BLAT searches, respectively, produced only the sequences shown in Fig. 4-2A. Sanger sequence trace is shown in Figure 4-2B. The primer sequences exactly match all sequences for both species and the number of base pairs between primers are equal. The GnRH-1 sequence for the larval PBMCs matched *P. marinus* GenBank sequences (alternative splicing did not affect the amplified segment

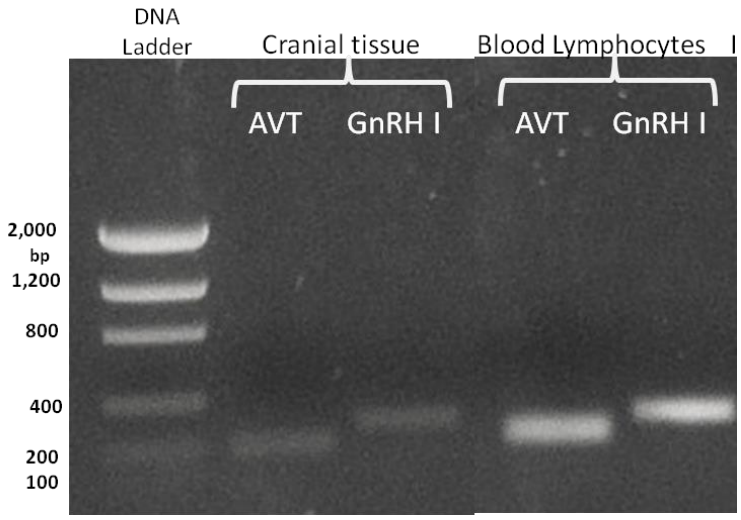
used in this study) and differed from the *L. japonicum* genome sequence (Fig. 4-3).

4.3.2 AVT hormone receptor expression in adult lamprey PBMCs

The control (freshly isolated) PBMCs from both the male and female adult animals showed actin expression indicating transcriptionally active populations of cells (Fig. 4-4). Even though transcriptional activity is observed in these cells, no Pm807 receptor cDNA could be amplified from the female and a weak band was observed from the male, which had a higher starting concentration of mRNA than for any of the other samples. After 6 days of incubation actin mRNA was still present, and Pm807 receptor cDNA amplification was observed in 6-day (SM6 and SF6L) but not in 1-day or 3-day incubations. It is not known why the duplicate male 6-day incubation (sample SM6d) did not produce Pm807 mRNA since the actin band indicates that mRNA was still present in cells within the culture. In the case of the female 6-day incubation with all mitogens (SF6), mRNA concentration that was likely too low or degraded to detect either actin or Pm807 amplification. RNA was not extracted from control cells cultured in media without mitogen stimulation, so it is unknown whether the culture medium alone would have resulted in expression of Pm807 mRNA. The purified PCR product (generated using the 285-bp Pm807 primers) from the stimulated female 6-day culture sample SF6-L was sequenced. Except for two bases, the resulting sequence matched the putative Pm807 receptor sequence characterized in Chapter 2 and deposited in the GenBank database (Accession number KC731437) (Fig. 4-5).

Trypan blue exclusion live-dead cell counts on the separate 6-day incubated cells showed 38% live cells in the control and 45% live cells in the mitogen treatment.

A



B

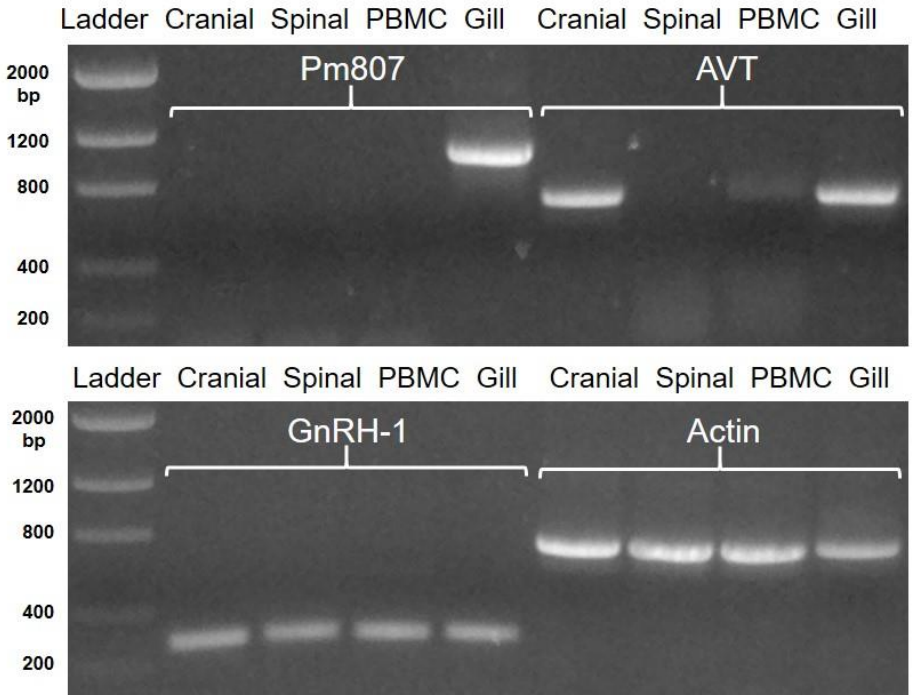


Figure 4-1. Neuropeptide hormone expression in larval lamprey peripheral blood mononuclear cells (PBMC). **A.** *Trial 1*. Messenger RNA was extracted from freshly isolated PBMC, and cranial tissue as positive control. Three larvae were pooled. 201-base-pair (bp) vasotocin (AVT) hormone and 280-bp gonadotropin releasing hormone I (GnRH I) gene segments were amplified through reverse transcription-polymerase chain reaction (RT-PCR) and the DNA was electrophoresed on a 1% agarose gel. **B.** *Trial 2*. Tissues from another three larvae were pooled, also including spinal cord and gill tissues. Actin (721 bp) and Pm807 V1A/OXT receptor (991 bp) segments were also PCR amplified, and the AVT hormone gene was amplified with primers for a 668-bp segment.

A

```
Lj_KE993680      CTGCGAGACTGACACGTGTCCACGCGCTGTGCGGCATGGTTTTATTTGCAGCGTCTAAAC
Pmarinus_PBMC    CTGCGAGACTGACACGTGTCCACGCGCTGTGCGGCATGGTTTTATTTGCAGCGTCTAAAC
Lj_D31871        CTGCGAGACTGACACGTGTCCACGCGCTGTGCGGCATGGTTTTATTTGCAGCGTCTAAAC
Pm_GL487152      CTGCGAGACTGACACGTGTACACGCGCTGTACGGAATGGTTTTACTTGCAGCGTCTTAAC
Pm_EST_DW585247 CTGCGAGACTGACACGTGTACACGCGCTGTACGGAATGGTTTTACTTGCAGCGTCTTAAC
                ****************** ********** *** ********* *********** ***

Lj_KE993680      GTGTTAAACAATACGGATAGGGTCGCATTGTCTTAATGTAATCACGAGCCCCCGTAATGC
Pmarinus_PBMC    GTGTTAAACAATACGGATAGGGTCGCATTGTCTTAATGTAATCACGAGCCCCCGTAATGC
Lj_D31871        GTGTTAAACAATACGGATAGGGTCGCATTGTCTTAATGTAATCACGAGCCCCCGTAATGC
Pm_GL487152      GTGTTAAACAATACGGATAGGGTCGCATTGTCTTAATGTACTCACGAGCCCCCGTAGTGC
Pm_EST_DW585247 GTGTTAAACAATACGGATAGGGTCGCATTGTCTTAATGTACTCACGAGCCCCCGTAGTGC
                ************************************ *************** ***

Lj_KE993680      CTGTGCATACTGTTCTAGTCGGCGTCCCGTTGGTATAATGAGTT
Pmarinus_PBMC    CTGTGCATACTGTTCTAGTCGGCGTCCCGTTGGTATAATGAGTT
Lj_D31871        CTGTGCATACTGTTCTAGTCGGCGTCCCGTTGGTATAATGAGTT
Pm_GL487152      CTGTGCGTACTGTTTTAGTCGGCGTCCCGTTGGTATAATGAGTT
Pm_EST_DW585247 CTGTGCGTACTGTTTTAGTCGGCGTCCCGTTGGTATAATGAGTT
                ****** ******* ****************************
```

B

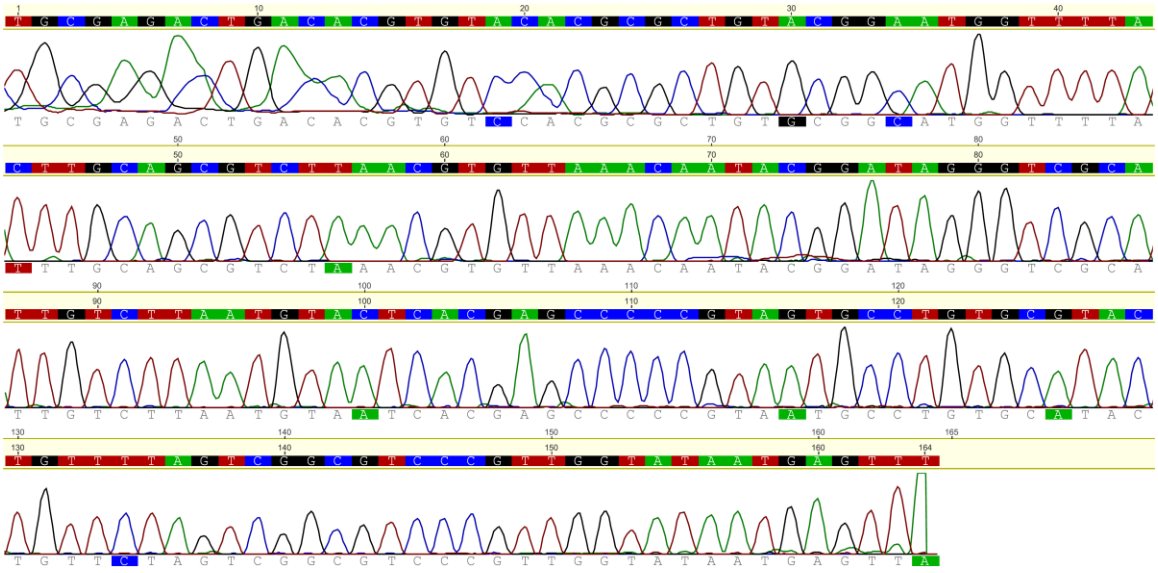


Figure 4-2. Sequenced PCR product amplified from freshly obtained larval peripheral blood mononuclear cells (PBMC) using vasotocin (AVT) hormone precursor-specific primers. **A.** Alignment of PBMC sequence (bold/italics) with *Lethenteron japonicum* (Lj) genomic scaffold KE993680, the GenBank vasotocin precursor sequence (Accession no. D31871), *Petromyzon marinus* (Pm) genomic scaffold GL487152 and GenBank expressed sequence tag (EST) DW585247 for vasotosin precursor. Non-consensus bases are highlighted. **B.** Sanger sequencing trace for larval PBMC PCR product showing very low background and no heterozygous peaks at polymorphic locations (highlighted bases), aligned with Pm_EST_DW585247 (all-colored sequence).

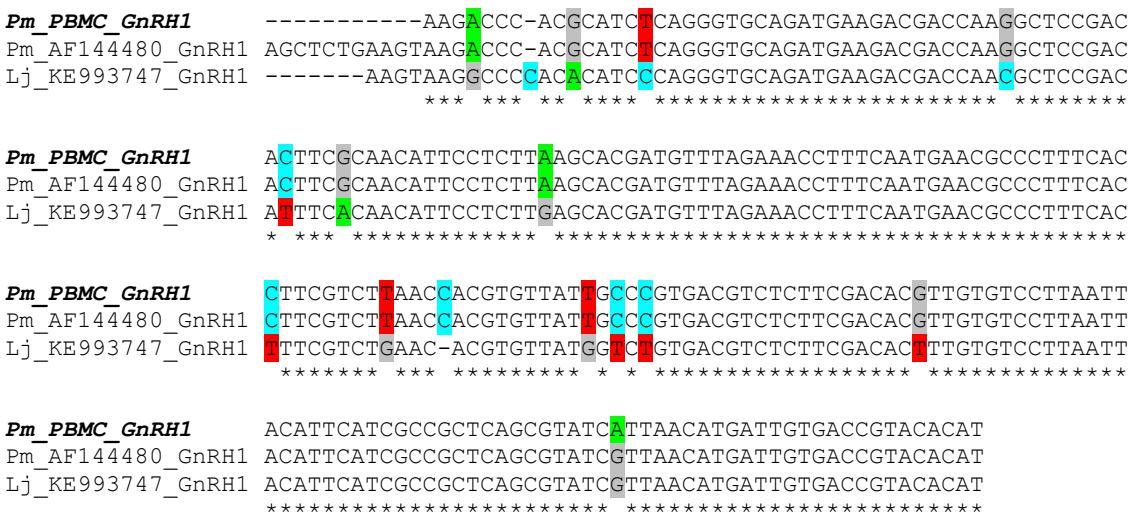


Figure 4-3. Sequenced PCR product amplified from freshly obtained larval peripheral blood mononuclear cells (PBMC; bold/italics) using gonapotropin releasing hormone-1 (GnRH-1)-specific primers. Alignment with *Lethenteron japonicum* (Lj) genomic scaffold KE993747 and *Petromyzon marinus* (Pm) GenBank sequence AF144480 for GnRH-1 hormone precursor. Non-consensus bases are highlighted.

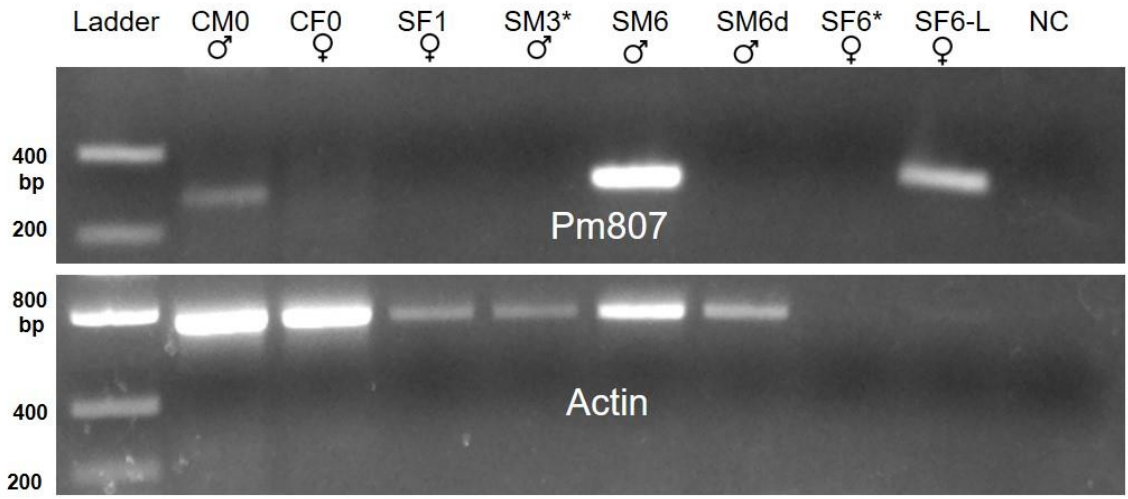


Figure 4-4. Adult peripheral blood mononuclear cell (PBMC) cDNA from freshly isolated cells (CM0 and CF0) or cultured cells stimulated (S) with varying combinations of phytohemagglutinin-P (PHA-P), lipopolysaccharide (LPS) and pokeweed mitogen (PWM) (see Table 4-1). Pm807 V1A/OXT receptor (285 bp) and actin (721 bp) segments were PCR amplified. *Sample CM0 started with 800 ng mRNA, samples SM3 and SF6 represent approximately 50 ng starting mRNA, all other samples started with approximately 500 ng mRNA in the reverse transcription reaction. NC=Negative Control.

```
Pm807_PBMC      AAGGCGTACA-CACCTGGATGACAGGCTCAGTCTTCATCATCCCGGTGATCATCCTCATC
Pm807_KC731437  AAGGCGTACATCACCTGGATGACAGGCTCAGTCTTCATCATCCCGGTGATCATCCTCATC
                ********** **************************************************

Pm807_PBMC      TGGTGCTACGGCATGATCACCTTCGCCATCTGGAAAAACATCAAGGCCAAGACGCAGGAA
Pm807_KC731437  TGGTGCTACGGCATGATCACCTTCGCCATCTGGAAAAACATCAAGGCCAAGACGCAGGAA
                ************************************************************

Pm807_PBMC      GGTGACAGCAGACACAACCCGGCCAAGAGCAGCGCCCCTTCCAGGGTCAGCTCGGTGCGC
Pm807_KC731437  GGTGACAGCAGACACAACCCGGCCAAGAGCAGCGCCCCTTCCAGGGTCAGCTCGGTGCGC
                ************************************************************

Pm807_PBMC      AGCATCTCCAAGGCGAAGATCCGCACCGCCAAGATGACGTTCGTCATCATCATGGT-TAC
Pm807_KC731437  AGCATCTCCAAGGCGAAGATCCGCACCGCCAAGATGACGTTCGTCATCATCATGGTGTAC
                ************************************************** ***

Pm807_PBMC      ATCATCTGCT
Pm807_KC731437  ATCATCTGCT
                **********
```

Figure 4-5. Sequenced PCR product amplified from female adult peripheral blood mononuclear cells (PBMC; bold/italics) after primary culture for six days at 15^{o}C (sample SF6-L, see Table 4-1 and Fig. 4-4), alignment with *Petromyzon marinus* (Pm) GenBank sequence KC731437 for Pm807 vasotocin V1A/OXT-type hormone receptor. Non-consensus bases are highlighted.

4.4 Discussion

This chapter describes a preliminary effort to establish a method for primary culture of lamprey immune cells. The ultimate goal is to utilize the cell cultures to study the role of neuroendocrine hormones and their receptors in immune cell function. Immune cells isolated from adult lampreys were maintained in an *in vitro* culture medium containing 2% lamprey serum for periods of at least 6 days, after which the presence of live cells were demonstrated. In control cultures 38% of the cells were alive and in mitogen-treated cultures 45% were alive after six days based on trypan blue dye exclusion live-dead cell counts. Weak expression of Pm807 mRNA was observed in the freshly isolated cells from the adult male lamprey and no visible band was observed for the female. However, after six days of incubation at 15 ^{o}C, apparent evidence of Pm807 mRNA transcription was observed in samples SF6-L and SM6 at a high level compared to actin. These results seem to indicate that Pm807 mRNA expression increased after exposure to mitogens (Fig. 4-4). Freshly isolated larval PBMCs appear to constitutively express the AVT and GnRH-1 hormone precursors (Fig. 4-1).

The cells in this study were isolated from peripheral blood using a density centrifugation technique that in mammals results in a population of mononuclear cells (thus, PBMCs) including mostly lymphocytes and some monocytes. The population of lamprey leukocytes that have been isolated with this technique (Pancer et al., 2004) are the lymphocyte-like cells, and also what are termed myeloid cells. It is unclear whether these are stem-type cells or differentiated monocyte-like cells, but their derivation cannot be from bone marrow as in mammals.

In this preliminary study it is not known whether the cells were actually stimulated by incubation with mitogens. Stimulation of immune cells involves a complex set of events. Mammalian lymphocytes that have not been exposed to specific antigens or polyclonal mitogens are in the quiescent (G_0) phase of the cell cycle. Polyclonal mitogens like those used in this study result in obligate stimulation toward mitosis, proliferation of lymphocytes and other cell types by interaction with constitutive cell surface pattern-recognition receptors like Toll-like receptors (TLRs). Although it is generally accepted (Janeway, 2008, p. 769) that exposure to polyclonal mitogens results in entry into the G_1 phase, the first step of the stimulation process leading to mitosis, this progression depends on having the appropriate growth factors in the culture medium. Activation can be tested in mammalian cells using specific antibodies to detect known activation markers such as CD3 or CD95. Activation markers are not well established for lamprey immune cells, nor are there antibodies available to detect any putative activation markers. Parameters that can be empirically measured in studies of both mammalian and non-mammalian immune cells in assessing whether a culture has been stimulated to proliferate are those that actually count live and dead cells; that measure metabolic activity; that measure incorporation of labeled nucleic acids into DNA, indicating passage from G_1 to the S-phase of the cycle; and that detect markers of apoptosis.

The antigen/mitogen cocktail injected into larval lampreys by Pancer et al., (2004) to stimulate the immune cells *in vivo* contained live *Escherichia coli* cells, sheep erythrocytes, PHA and PWM. The mitogens PHA-P (15 µg/ml), PWM (10 µg/ml) and LPS (100 µg/ml) were used in the present *in vitro* study of adult lamprey immune cells at

higher concentrations than were proven to induce responses in other studies, as discussed below. Additionally, most primary culture experiments used only one mitogen. All three mitogens combined could likely produce a more powerful effect in a mixed cell population than each mitogen individually.

Pokeweed (*Phytolacca americana*) mitogen (PWM) is a lectin that was found to induce DNA synthesis in B-lymphocytes but not T-cells at a low dose of 1 µg/ml (Mellstedt, 1975). A dose of 10 µg/ml was found in humans to induce DNA synthesis in both cell types, but B-cells to a greater degree than T-cells (Mellstedt, 1975).

Phytohemagglutinin-P (PHA-P, purified) and PHA-M (crude extract) are a lectin protein from the red kidney bean *Phaseolus vulgaris*, that cause proliferation of T-lymphocytes (Yachnin and Svenson, 1972; Mellstedt, 1975). In a study on human PBMCs, oxytocin (OXT) was found to increase proliferative responses to 5 µg/ml PHA-M (Maccio et al., 2010). As little as 10^{-12} M OXT significantly increased the number of PBMCs that entered the S-phase of the cell cycle, and significantly increased expression of the interleukin-2 receptor (IL-2R or CD25) and CD95 compared to unstimulated cells or PHA alone. An OXT receptor antagonist blocked the effects, confirming that the OXT hormone was working through its native receptor (Maccio et al., 2010).

Lipopolysaccharide (LPS) is the major component of the outer cell membrane of Gram-negative bacteria such as *E. coli*, from which the LPS in this study was derived. Myeloid lineage cells are the primary sentinel cells reacting to LPS in mammals, but also monocytes and B-cells (Alexander and Rietschel, 2001). DeKoning and Kaattari (1992) used only LPS at an optimal concentration of 200 µg/ml in a medium containing 2%

salmon serum and 5% FCS to achieve cell proliferation in salmonid lymphocytes. LPS was the only mitogen used in a study on neuropeptide hormone effects on proliferation and survival of cultured rainbow trout (*Oncorhynchus mykiss*) peripheral blood leukocytes (Yada et al., 2004). Trout leukocyte survival was decreased between two days and four days, and further decreased after eight days in culture medium with 0.5% trout serum and no LPS. Survival and proliferation were significantly increased in 4-day cultures with 5 µg/ml LPS, or neuropeptide hormones prolactin (5 nM) or growth hormone (GH; 5 nM). These results were corroborated by four different assays: total cell number, tetrazolium [MTT] reduction showing metabolic activity, propidium iodide distinguishing live and dead cells, and DNA incorporation of bromo-deoxy-uridine. Additionally, the Annexin V assay was used to show that prolactin alone and with LPS produced a significant decrease in apoptosis compared to control (Yada et al., 2004). These assays would be appropriate measures in future lamprey immune cell studies to demonstrate stimulation and proliferation in primary culture and to assess the functional role of AVT and other neuroendocrine hormones.

Also of note, the larval AVT PBMC sequence was a near exact match of the *Lethenteron* rather than the *Petromyzon* sequences. No other AVT-like hormone precursor is known in lampreys (see Chapter 2 and Gwee et al., 2009), indicating a potential AVT polymorphism that occurs in both species. It has been found that different alleles can be expressed in white blood cells of human heterozygous individuals, suggesting different cis-regulatory mechanisms in different cell or tissue types (Krishna

Pant et al., 2006). Preferential allelic expression would be another piece of evidence supporting a functional purpose for these neuropeptide hormone genes in immune cells.

As discussed in Chapter 1, functions such as chemotaxis (Chen et al., 2002), wound healing (Poutahidis et al., 2013), and cytokine regulation (Koike and Seki, 2013) have been demonstrated in mammalian lymphocytes. My findings provide preliminary support for the hypothesis that expression of neuropeptide hormone systems in immune cells is conserved in the earliest extant species to diverge from the vertebrate ancestor, however more stringent controls are needed. With fine-tuning of the culture medium and mitogens, *in vitro* primary culture of lamprey immune cells could become an important neuroendocrine and immune system research tool.

Chapter 5

Effects of vasotocin and oxytocin hormone exposure on sex pheromone release from male sea lamprey gills

5.1 Introduction

Sex pheromones are hormones secreted by an individual that are meant to move through an environment to be received by and attract a prospective mate of the same species. Pheromones can be produced by either males or females. In the case of the sea lamprey, males release a mating pheromone from cells in their gills when they begin spermiating (at reproductive maturation) that induces search and preference behaviors in ovulating females (Siefkes et al., 2003). These reproductive behaviors appear to be mediated over long ranges only by olfactory sensing (Johnson et al. 2012). According to Siefkes et al. (2003) the mechanism male sea lampreys employ to release the pheromone is by actively pumping it from the specialized cells in the gill epithelium in order to broadly distribute the signal. Finding out how the pheromone is released could lead to a way to block the mechanism, thereby providing a potential target to control this invasive species in the Great Lakes.

Although few species have been tested for a connection between pheromones and neuropeptides, one study in male red-bellied newts (amphibian; *Cynops pyrrhogaster*) demonstrated that arginine vasotocin (AVT) induces pheromone discharge from a contractile abdominal gland (Toyoda et al., 2003). This led to my hypothesis that AVT may mediate active pheromone release from the male lamprey gill cells.

Additionally, a study was conducted on a cichlid fish species in which excised testicular tissue fragments were incubated in Krebs-Ringer's medium containing various concentrations of AVT (Ramallo et al., 2012). They found that male steroid hormones (androgens) released from the tissue increased with increasing AVT concentration. Their

methodology served as a model for my first experiment on pheromone release from excised gills. Lampreys possess seven gill slits on each side, each open to an individual gill pouch. I found that the gill pouches can be easily removed and act as homogeneous individual treatment units.

To determine whether blocking AVT receptors would diminish pheromone release, I used a Vaptan receptor antagonist. Vaptans are highly specific molecules designed to fit into the receptor active sites of the various mammalian arginine vasopressin (AVP) receptor types, where the receptor's native hormone AVP would normally bind, blocking the conformational change and disrupting function (Decaux et al., 2008). A Vaptan called OPC-21268 (Sigma-Aldrich Chemical) was specifically designed to block the AVP V1a receptor. In a second experiment, intact animals were used, and the AVT or an alternate ligand, oxytocin (OXT) was injected directly into the heart. The first experiment was conducted late in the reproductive season while the second experiment was conducted early in the season. The different methods and timing were intended to be exploratory, and the results are not definitive, but provide preliminary evidence that warrants further study.

5.2 Materials and Methods

5.2.1 Experiment 1: Ex vivo *gill pouches*

Adult male reproductively-mature lampreys were obtained from USGS Hammond Bay Biological Station (HBBS) (Millersburg, MI) in August 2013 and were held without feeding in a 100-gallon stock tank at 15 ^{o}C for a maximum of two weeks after receipt. Spermiated condition was tested by pressing the abdomen and confirming the release of semen. Animals were euthanized in tricaine methanesulfonate (MS-222, Sigma, St. Louis, MO), pH 7.4, at 1 g/L. Animal treatment conformed to University of Minnesota animal care standards (IACUC protocol number 1305-30612A).

Krebs-Ringer's Solution (KR soln) components and concentrations (mM) are as follows: NaCl, 119.0; KCl, 2.5; NaH_2PO_4 (monobasic), 1.0; $NaHCO_3$, 25.0; $CaCl_2 \cdot 2H_2O$, 2.5; $MgSO_4 \cdot 7H_2O$, 1.2; HEPES buffer, 20.0; D-glucose (dextrose) $C_6H_{12}O_6$, 11.0, adjusted to pH 7.4 at 25°C with 1M NaOH. A 10^{-4} M AVT stock and 10^{-5} M OPC-21268 (Sigma-Aldrich Chemical) stock solution and subsequent dilutions were prepared in KR soln.

Lampreys were exsanguinated via heart puncture to remove most of the blood from the gills. Gill pouches were removed from seven animals by making an incision on the ventral side from the mouth to the heart and snipping connective tissue and vasculature to free each gill pouch. The cylindrical pouches were opened to a flat sheet by snipping on one side. Ten of the 14 gill pouches (from gill slits 2 through 6 on each side) from each animal were collected. The pouches were similar in size and presumed similar in function. Gills from one side of each animal were pre-incubated in medium only and the

other side in medium with a 10µM concentration of the Vaptan V1a AVP receptor antagonist OPC-21268 to potentially block the receptors prior to exposure to AVT. Figure 5-1 depicts the experimental design.

After 30 minutes the gills in medium only were moved randomly to individual wells of a 12-well plate containing 1 ml of medium with 1, 10, 50 or 100 nM AVT, and the gills in the medium + OPC were moved randomly to wells containing the same AVT concentrations + the 10 µM OPC (Figure 1). After a three-hour incubation with agitation the medium was collected into 2-ml cryovials and frozen at -20 ^{o}C overnight, then moved to a -70 ^{o}C freezer until shipment. Samples were shipped on dry ice to the laboratory of Dr. Weiming Li at Michigan State University for analysis. Dr. Li's lab was the first to identify and determine the nature and production of the bile acid-based male sex pheromone 3-keto petromyzonol sulfate (3kPZS), determine the female response, and identify the gills as the source of excretion to the environment (Li et al., 2002; Siefkes et al., 2003; Li, 2005; Johnson et al., 2012; Walaszczyk et al., 2013). Compounds analyzed included 3kPZS, 3-keto petromyzonol (3kPZ), petromyzonol sulfate (PZS), petromyzonol (PZ), allocholic acid (ACA), 3-keto allocholic acid (3kACA), petromyzonamine disulfate (PADS), petromyzosterol disulfate (PSDS), and 3,12-diketo-4,6-petromyzonene-24-sulfate (DKPES). Results were expressed as ng pheromone/1 ml incubation medium. Statistical analyses of the assay data was conducted using analysis of variance (ANOVA).

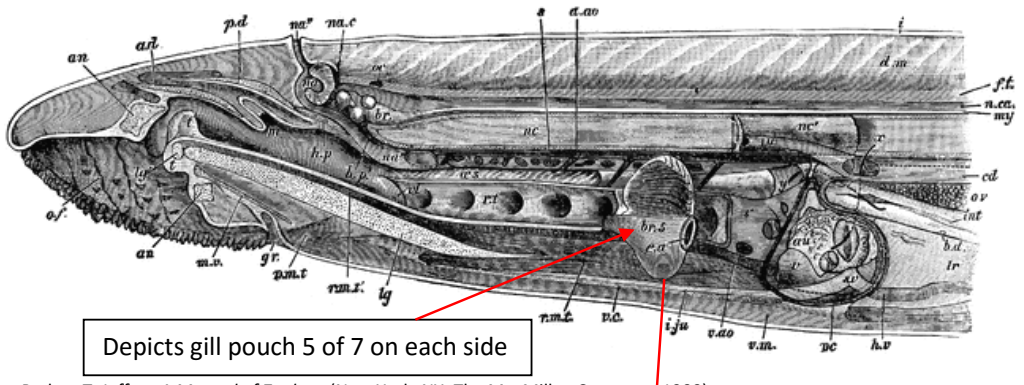


Illustration: Parker, T. Jeffrey *A Manual of Zoology* (New York, NY: The MacMillan Company, 1900)

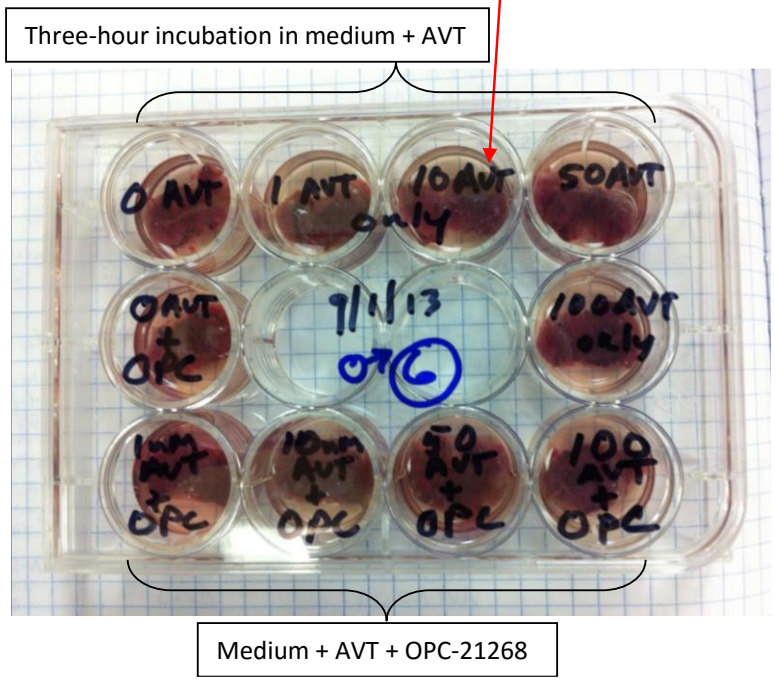


Figure 5-1. Experimental design schematic depicting lamprey gill pouches and the method of incubation with various concentrations of arginine vasotocin (AVT) or AVT + the mammalian vasopressin V1a receptor antagonist OPC-21268. Gill pouches two through six on each side (10 total) were used in experiments, pre-incubated with either buffer medium only or medium + OPC prior to treatment in 12-well plates as shown.

5.2.2 Experiment 2: Intact animal heart injection

Eighteen males were received from HBBS in mid-June during the lamprey run, and were processed within two weeks of receipt. Spermiated condition was tested and determined that the animals in this phase were not releasing sperm with pressure on the abdomen. In the spermiated condition as in the previous experiment, endogenous AVT, if involved in pheromone release upon gonadal maturation, could already have triggered the releasing process. Testing of mature upstream migrants not yet at spermiation was performed to determine whether exogenous AVT could induce the release of pheromones.

Dosage was determined based on studies in lampreys as well as in other species. In a study of AVT effect on kidneys in *Lampetra japonica* (Uchiyama and Murakami, 1994), injected doses of 1-100 ng/kg caused an antidiuretic response. Blood pressure responses in rainbow trout (*Oncorhynchus mykiss*) were elicited with similar doses (Conklin et al., 1997). Higher doses of 2-4 international units AVP (50 I.U. = 1 mg) per kg body weight were used to elicit cardiovascular responses in rabbits (Black, 1960). Weight range of lampreys in this experiment was 111-319 g, with a mean of 219 g ($\pm$ SD of 53 g). A total dose per animal of either 4 I.U. of AVT or OXT was chosen, delivered in a fish physiological saline solution (6.42 g/L NaCl, 0.15 g/L KCl, 0.22 g/L $CaCl_2$, 0.12 g/L $MgSO_4$, 0.084 g/L $NaHCO_3$, 0.06 g/L NaH_2PO_4); controls received saline carrier only.

Before injection, each lamprey was first anesthetized for 20 min under 0.1 g/L MS-222 in ~400 ml aerated de-chlorinated tap water with 200 µl (0.2 mg) fluorescein dye,

and resulting water containing gill exudates was sampled. Fluorescein dye was added to normalize the volume for calculation of analyzed pheromone concentrations (Suppl. Fig. 5-1, Appendix B). The lamprey was then moved to another container with a euthanasia dosage of MS-222 (1 g/L) for 10 min after which a small ventral incision was made to expose the heart. Saline carrier for controls or 4 I.U. (0.08 mg/100 µl) of AVT or OXT was directly injected into the heart ventricle. Seven males were injected with AVT, six males with OXT and five males with saline. The body was held by an adjustable, rubber-coated test tube clamp so that the head and gills were suspended in ~150 ml fresh aerated water with 0.05 mg fluorescein dye, and the water sampled after 20 min. Samples in 2-ml cryovials were collected from each pre- and post-injection incubation and frozen at -20 ^{o}C overnight, then moved to a -70 ^{o}C freezer until shipment on dry ice to the Li lab at Michigan State University for analysis. Analyses included PZ, 3kPZ, PZS, 3kPZS, ACA, 3kACA, PADS, and PSDS.

To determine the total volume of water in each sample, fluorescein was measured at a wavelength of 488 nm (Arcoumanis et al., 1990) and compared to a standard curve (Suppl. Fig. 5-1A). Absorbance of samples was read within 48 hours of sampling with samples held at 4 ^{o}C. A time course experiment shows the dye was stable over at least six days under these conditions (Suppl. Fig. 5-1B). Concentrations of pheromones supplied from the duplicate analytical results as ng/ml were averaged then multiplied by the total volume of incubation water (gill washings) and expressed as total ng released in the 20-min incubation period.

5.3 Results

5.3.1 Experiment 1: Ex vivo *gill pouches*

Gill pouches from seven spermiated male lampreys were incubated in media containing various concentrations of AVT, or AVT and an AVP receptor antagonist Vaptan to determine whether stimulation of AVT receptors in the gill cells mediate release of pheromones. Analyses of the pre-incubation media showed that the gills were exuding pheromones prior to exposure to AVT (Suppl. Table 5-1). Bile acids PSDS, DKPES and 3kPZ were not detected in any samples and PZ was detected at low levels in only three samples (data not shown). The allocholic acids ACA and 3kACA were detected in all gill samples from three lampreys (1, 2 and 4) at levels lower than any of the other compounds, and only intermittently among the other four lampreys (Suppl. Table 5-1). PADS was the only compound detected in all samples (Suppl. Table 5-1; Fig. 5-2A). Pheromone 3kPZS was detected at similar levels to those of PADS in samples from lampreys 1, 2 and 4, but at very low levels or not detected in lampreys 3, 5, 6, and 7 (Fig. 5-2B). PZS concentrations were the highest of all compounds, also in lampreys 1, 2 and 4. The 3kPZS concentrations from the 1-ml media/gill pouch incubations (lampreys 1, 2, and 4) were higher by an average factor of about 80,000 than 3kPZS concentrations in environmental samples from streams sampled in late June (Xi et al., 2011). Data is presented for individual lampreys due to the high variation between animals (Fig. 5-2), although statistical analyses were also conducted showing no correlation between pheromone levels and AVT concentration or interaction with Vaptan treatment ($p>0.05$).

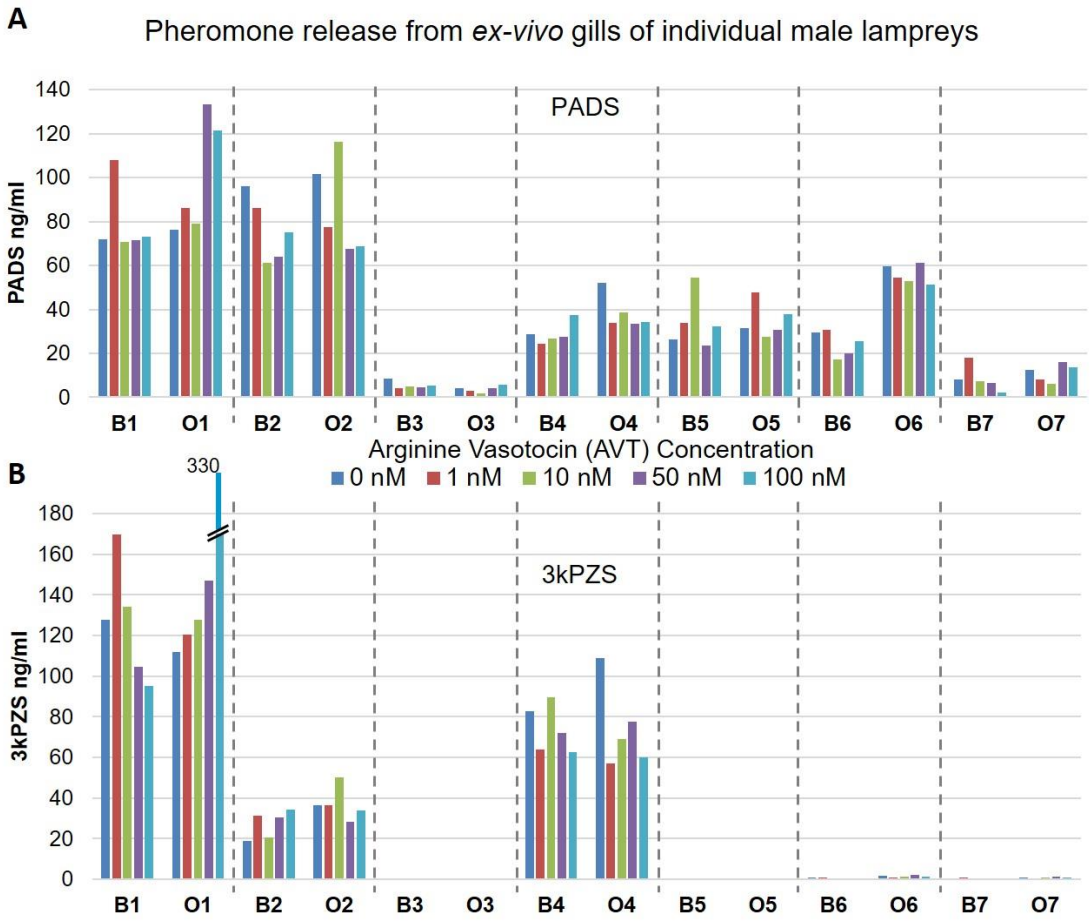


Figure 5-2. Results of Experiment 1 for pheromone collection from individual gill pouches. **(A)** Petromyzonamine disulfate (PADS) and **(B)** 3-keto-petromyzonol sulfate (3kPZS) pheromone release from excised gills of individual male lampreys. Five gills from each of seven lampreys were incubated in Krebs-Ringers buffer with arginine vasotocin (AVT) only (designated B), and the other five gills (designated O) from the same animal were incubated with AVT and 10 µM OPC-21268 (a mammalian V1a AVP receptor Vaptan antagonist). There were no significant differences among AVT concentrations and no significant interaction with OPC-21268 ($p>0.05$).

5.3.2 Experiment 2: Intact animal heart injection

The injection of AVT and OXT directly into the heart was an effort to deliver the hormones to the gill glandular cells to reach the AVT receptors in the same manner as endogenous AVT. Although the lampreys were exposed to euthanasia doses of MS-222, the heart often continued to beat. Test injections of bromophenol blue and fluorescein dye under the same conditions showed that the dye reached the gills even when the heart beat only a few times within the 20 min period (I have observed that the heart may beat for more than 30 min even when removed from the animal and placed in culture medium).

One animal in each treatment was disqualified from the analysis due to leakage of blood into post-injection samples. Due to dilute concentrations in initial analyses, all samples were concentrated by freeze drying prior to completing analyses. Nonetheless, in most samples, PZ, 3kPZ, and PSDS were not detected or were detected below the quantitation limit (Suppl. Table 5-2). Although inter-animal variation was again high as in Experiment 1, a potential correlation between treatment and pheromone release was observed in the PADS results. Three of the remaining six AVT-injected animals had post-injection total PADS collected that were 12- to more than 30-fold higher than the pre-injection gill washings. Three of the five OXT-injected males also had substantially higher post- than pre-injection totals, while this was not observed for any of the control males (Fig. 5-3A). However, results for 3kPZS indicated that among all lampreys tested only two of the OXT-injected animals had post-injection totals marginally higher than before injection (Fig. 5-3B). The ACA, 3kACA and PZS results were similar to the 3kPZS results (Suppl. Table 5-2).

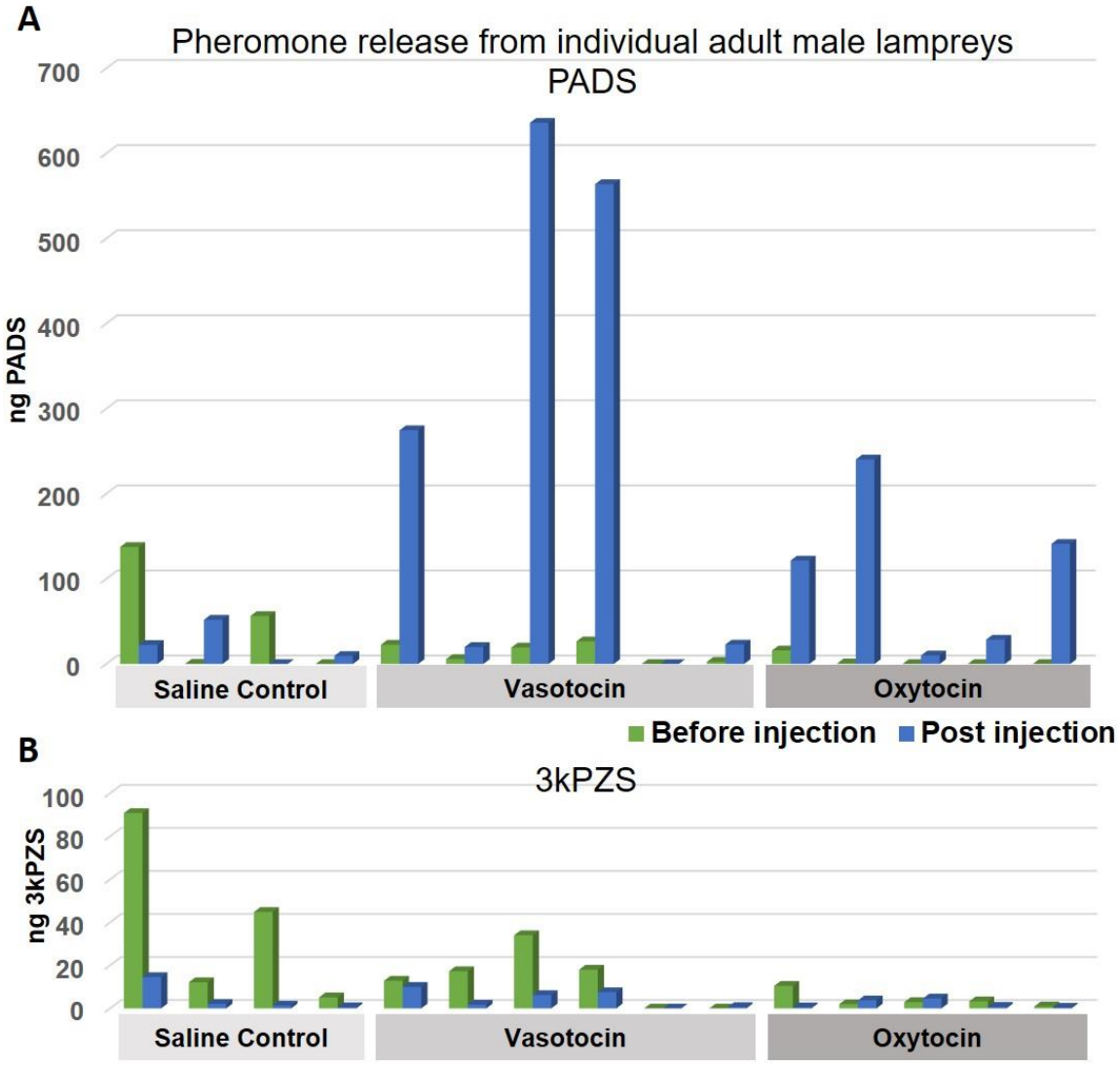


Figure 5-3. Results of Experiment 2: Pheromone release from intact individual male lampreys dosed with 4 I.U. of oxytocin or vasotocin directly injected in the heart. The total amount of petromyzonamine sulfate (PADS) **(A)** or 3-keto-petromyzonol sulfate (3kPZS) **(B)** released in 20-minute incubations (normalized to fluorescein dye) before and after injection are shown. Controls received the saline carrier only.

5.4 Discussion

No conclusive evidence was found in this study that AVT mediates the release of sex pheromones from adult male gill tissues. High variability was observed between lampreys in Experiment 1 but not between individual gill pouches within each lamprey. With the PADS results in Experiment 2 for some AVT- and OXT-injected male lampreys being higher in the post-injection samples than before injection, the possibility that these hormones are involved cannot be discounted.

Important outcomes of this study overall included the collection of pheromone release data from a total of 25 adult male lampreys. Methods of collecting pheromones from individual gills showed that similar levels of pheromone compounds seemed to be released from all gills in a particular animal. Also, the relative levels of each compound were consistent from one animal to another, but varied between the gill pouch and intact animal methods (Suppl. Tables 5-1 and 5-2). PADS was detected at substantially higher levels in some of the injected animals relative to 3kPZS in Experiment 2 (Fig. 5-3) than in the gill samples from Experiment 1, where the two compounds were at similar levels (Fig. 5-2). Generally, PZS was detected at approximately two to more than five times the concentrations of either 3kPZS or PADS in the gill samples from Experiment 1, while PZS and 3kPZS were at similar levels in Experiment 2.

One explanation for the relative concentration variation may be that the intact animal samples were more dilute, required concentration by freeze drying before analysis, and were closer to the limits of quantitation. Another explanation is that the differences in method reveal different sources for some compounds. It is possible that

PADS is more prevalent in the blood stream in pre-spermiating males, which is compatible with the current understanding that 3kPZS from the gills is the most potent female-attracting sex pheromone component (Walaszczyk et al., 2013; Johnson et al., 2009) while PADS is considered an active component of migratory pheromones released by ammocoete larvae. Olfaction tests have shown that ACA and PZS act as primary components in migratory pheromone mixtures (Li and Sorensen, 1997; Polkinghorne et al., 2001) and are released through larval feces along with PADS and PSDS (Hoye et al., 2007). 3kACA and DKPES may act as enhancers in sex pheromone mixtures (Yun, Scott and Li, 2003; Li, Brant, et al., 2013). Recent studies utilized implanted time-release steroid hormone pellets to show that production and plasma levels of 3kPZS were increased by the progesterone (~1000-4000 ng/ml) over controls (~10-40 ng/ml) and androstenedione (~50-75 ng/ml) in spermiating males (Bryan et al., 2015). Though all compounds seem to be made in the liver and travel through the bloodstream, perhaps the keto-compounds are more likely to be passed into the male gill glandular cells.

Results of this pilot study may provide clues as to the physiological source of the compounds at different maturation stages of the lampreys tested. Potential causes of variability and methodological adjustments for future efforts are discussed as follows.

Experiment 1

-This experiment was conducted very late in the season, with all seven lampreys tested in fully-mature, spermiated condition. If sex steroid hormones influence the release of endogenous AVT, and AVT mediates release of pheromones, some of the males may already have released their pheromones. Also, the animals may have been experiencing

system failures, nearing the end of their life cycle; 3kPZS levels were not detected in some animals in this study, but when detected, levels were similar to plasma levels reported by Bryan et al. (2015) in non-steroid treated males.

-The normalization methods were limited to the initial 1-ml volume of media dispensed and wet weight of the gills. Results were the same statistically whether or not the pheromone concentration was normalized to gill wet weight. Spiking compounds can be added to normalize the data and control for degradation.

-Although the lampreys were exsanguinated prior to removal of the gills, and gills were rinsed and pre-incubated prior to treatment, some blood may still have been present in the collected media (based on slight pinkish color). Therefore, because bile acid pheromones are thought to be transported to the gills from the liver via the blood (Siefkes et al., 2003; Li, 2005), some of the pheromone molecules may have entered the sample through the blood rather than being "pumped" out of the gill epithelial cells.

Experiment 2

-The incision made to expose the heart generally closed on itself after injection, however small amounts of blood were present in all samples and in three cases larger amounts of blood leaking into the sample caused the elimination of those results. Ideally, live lampreys could be held in a chamber with a gasket between the gills and posterior part of the animal (Siefkes et al., 2003), and a cannula could deliver intravenous AVT. This would allow longer term collection of pheromones from each animal.

-In the current study, some Experiment 1 males may have been tested too late in the life cycle while in Experiment 2 some of the animals may have been tested too early. The

above chamber method could be used to control for timing issues, by following individual males through their pre-spermiating and mature spermiated phases to determine how AVT affects pheromone release.

Chapter 6

Differential DNA methylation in the putative promoter of a sea lamprey (*Petromyzon marinus*) vasotocin receptor gene is correlated with tissue-specific and life stage-specific mRNA expression

To be submitted to *Frontiers in Endocrinology*, S. A. Mayasich, L .T. Bemis and B. L. Clarke, authors.

The sea lamprey genome has a different structure from both invertebrates and jawed vertebrates featuring high guanine-cytosine (GC) content. This raises the question of whether DNA methylation of cytosine-guanine (CpG) dinucleotides could function to regulate lamprey gene transcription. We previously characterized a lamprey vasotocin receptor gene (Pm807) that diverged into the separate vasotocin/vasopressin 1a (V1a) and oxytocin receptor (OXTR) genes in jawed vertebrates. Promoter region CpG methylation regulates mammalian expression and tissue localization of OXTR mRNA. We found CpG islands in the lamprey Pm807 V1a/OXT receptor gene promoter region. Lamprey Pm807 mRNA is highly expressed in adult heart but not expressed in adult liver, though it is expressed in larval liver. We hypothesize that differential DNA methylation is responsible for the difference in mRNA expression. Using High Resolution Melt (HRM) PCR on bisulfite-converted DNA, we pinpointed a region with tissue-specific differences in DNA melt characteristics, indicating differences in methylation level. Sequencing revealed a pattern of methylation at specific CpGs at consistently higher levels in adult heart and larval liver than adult liver. These CpGs are associated with putative transcription factor binding site sequences organized similarly to functional OXTR promoters in mammals, suggesting that CpG methylation functions in lamprey gene transcription regulation.

6.1 Introduction

Epigenetic modifications of genomic DNA are well known in mammals as regulators of tissue-specific gene expression. Cytosine-guanine dinucleotide (CpG) methylation is a highly studied epigenetic modification that is generally associated with transcriptional repression when observed in the promoter region of a gene, often near the transcription start site (TSS) (Hu et al., 2013). Whole genome bisulfite sequencing is frequently used to characterize methylation patterning in mammalian gene promoters. Recently, several genomes of non-mammalian animals have also been studied with this methodology. A wide variety of methylation patterning has been observed, however little is known about how methylation patterning in non-mammals relates to transcription of individual genes. The genome-wide pattern in general changes from mosaic in invertebrates to global methylation in vertebrate genomes (Hendrich and Tweedie, 2003). In the case of the sea lamprey which occupies a pivotal position in the evolutionary transition from invertebrate to vertebrate physiology, the methylation pattern is global (occurring throughout the genome) as in vertebrates (Tweedie et al., 1997). Very little is known about specific epigenetic events in lamprey gene regulation. The guanine-cytosine (GC) content and structure of the lamprey genome differs from both vertebrates and invertebrates (Smith et al., 2013), raising the question of whether CpG island methylation can be a functional regulator of lamprey gene transcription.

The arginine vasopressin (AVP)/oxytocin (OXT) hormone receptor system in mammals mediates multiple functions including social behavior, reproductive physiology and metabolism. The promoter regions of these genes in humans have been well studied

(Dhakar et al., 2013), since single-nucleotide polymorphisms, (SNPs), microsatellites and epigenetic modifications have been implicated in behavioral variation and disease. Receptor function in specific tissues during various life stages requires suppression in some tissues and expression in others. Locations of CpG islands, where CpG dinucleotides occur at a higher density than the expected probability, are found in both the human AVP V1a and OXT receptor promoter regions and regulate gene transcription (Kusui et al., 2001). Hypermethylation at CpGs in a region just upstream and downstream of the translation initiation site cause transcription suppression of the OXT receptor in the liver while the hypomethylated promoter allows expression in the uterus. Furthermore, the V1a receptor is highly expressed in mammalian liver (Kusui et al., 2001).

We previously characterized the vasotocin homologs of the vasopressin/OXT hormone receptors in the sea lamprey (*Petromyzon marinus*) (Chapter 2). We determined through phylogenetic and syntenic analyses that one ortholog found on scaffold 807 comprises characteristics of both the V1a and the OXT receptors. Messenger RNA was found to be expressed in several adult and parasitic-phase lamprey tissues including brain, gills, gonads, heart, and kidneys but not liver (Chapter 2, Fig. 2-4). Our hypothesis is that DNA methylation in the promoter region of the Pm807 V1a/OXT receptor regulates tissue- and life stage-specific transcription in lampreys. We used High Resolution Melt (HRM) polymerase chain reaction (PCR) and sequencing of bisulfite-modified genomic DNA to investigate the relationship of methylation level to mRNA expression at the individual gene level in the putative promoter region of the Pm807 V1a/OXT receptor.

6.2 Materials and Methods

6.2.1 Animal treatment and tissue collection

Sea lamprey larvae and adult upstream migrants were obtained from USGS Hammond Bay Biological Station (HBBS) (Millersburg, MI). Adult animals (which no longer feed) were held in a 100-gallon stock tank at 15^{o}C for a maximum of two weeks after receipt. Adults were euthanized in 2 g/L tricaine methanesulfonate (MS-222, Western Chemical, Inc., Ferndale, WA), pH 7.4 and stored at -20 ^{o}C prior to removing the heart and liver from each animal. A section of approximately 100 mg of each tissue was placed in a sterile petri dish and minced with a scalpel; 25 mg of the minced tissue was weighed into a 1.5 ml microcentrifuge tube for DNA isolation. The larvae (approximately 100-110 mm long) were maintained at the University of Minnesota-Duluth in a sand-lined 2-gallon aquarium in dechlorinated tap water at a temperature of 10^{o}C within a circulating water bath. A 24-hour feeding of 5 g of moist cake yeast slurry was provided prior to euthanization in 1 g/L MS-222 in dechlorinated tap water. Whole livers (approximately 15-20 mg each) were immediately harvested, placed in 1.5 ml microcentrifuge tubes and stored at -70 ^{o}C for DNA and RNA isolation. Animal treatment conformed to University of Minnesota animal care standards (IACUC protocol number 1305-30612A).

6.2.2 Nucleic acid isolation and treatment

Heart and liver tissues from four adult lampreys and livers from four larval lampreys were used in the HRM PCR screening of the Pm807 vasotocin receptor gene to determine if differences in CpG methylation status exist within the putative promoter

region. The tissues were crushed and homogenized in individual microcentrifuge tubes on dry ice using a sterile Teflon pestle. The DNeasy kit (Qiagen) was used to isolate genomic DNA following the manufacturer's protocol, with a three-hour incubation at 56 ^{o}C to lyse the tissue. DNA was eluted with AE buffer. Nucleic acid concentration was quantified by Nanodrop spectrophotometry and a Qubit® dsDNA BR assay kit (Life Technologies) was used with a Qubit® fluorometer to determine the specific DNA concentration. These results were subsequently used to normalize DNA to 1 µg for methylation and 2 µg for bisulfite conversion reactions. DNA quality was checked on a 1% agarose gel showing strong bands at 23 kb (Suppl. Fig. 6-1). Fully methylated reference DNA was created through *in vitro* methylation using 1 µg DNA and incubating for two hours at 37 ^{o}C with SssI methylase and S-adenosyl methionine as substrate according to the manufacturer's instructions (New England Biosciences). Each 20 µl reaction was stopped by heating at 65 ^{o}C for 20 min, and three reactions were pooled, precipitated in a final volume of 75% ethanol and resuspended in 25 µl nuclease-free water.

Bisulfite conversion of unmethylated cytosines in the methylated and untreated DNA samples was conducted with 2 µg of DNA per reaction. The EpiTect® Fast DNA Bisulfite Kit (Qiagen) was used according to the manufacturer's instructions for a 140 µl reaction. Thermal cycler conditions included two cycles of denaturation for 5 min at 95 ^{o}C and incubation for 15 min at 60 ^{o}C. The converted DNA was stored at 4 ^{o}C if used within 24 hours, thereafter at -20 ^{o}C.

Large quantities of RNA were present in the DNeasy eluate of the larval liver

tissue. RNA was then further isolated by simply treating with DNase I (DNA-*free*, Ambion) to remove genomic DNA according to the manufacturer's standard protocol in a 50 µl reaction. The larval liver RNA quality was checked on a 1% agarose gel, showing removal of the 23 kb genomic DNA band (Supplemental Fig. 6-1, Appendix B). RNA concentration was again measured by Nanodrop and 1 µg of RNA was reverse transcribed to cDNA using ThermoScript^{TM} RT (Invitrogen) in a 20 µl reaction. A mock reaction with the same RNA concentration and other conditions but without the RT enzyme served as a negative control to demonstrate the lack of DNA contamination. Note that the RNA is destroyed during bisulfite treatment with the EpiTect® kit according to the manufacturer (Qiagen), and therefore does not interfere with downstream assays.

6.2.3 Expression of Pm807 V1a/OXTR mRNA

Polymerase chain reaction (PCR) assays were conducted on larval liver cDNA using primers for a 285-bp segment of the coding region on the cDNA and the non-reverse-transcribed mock cDNA. Primers for a 280-bp segment of the lamprey β-actin gene were used to amplify the cDNA as a positive reference. Primer sequences are as follows: Pm807 forward AAGGCGTACATCACCTGGATGACA, reverse GACTGACCACATCTGCACGAAGAA; β-actin forward ATCATGTTCGAGACCTTCAACACGC, reverse TCTCCTTGATGTCACGCACGATCT. GoTaq Master-Mix (Promega) was used in a 25 µl reaction with 1 µl cDNA with a thermocycler profile of 2 min at 94 ^{o}C followed by 34 cycles of 30 s at 94 ^{o}C, 30 s at 60 ^{o}C, and 60 s at 72 ^{o}C, with a 5 min elongation step at 72 ^{o}C.

6.2.4 High Resolution Melt PCR screening of putative promoter region DNA

The location of the TSS and 5'-untranslated region (UTR) sequence of the lamprey Pm807 V1a/OXTR gene was previously determined using rapid amplification of cDNA ends (RACE) methodology (NCBI Accession number KC731437.1; see Chapter 2). The sequence is located on the minus strand of scaffold GL477135, GenBank Accession number AEFG01068454. Locations of CpG islands were determined in the region from the TSS to approximately 2300 base pairs (bp) upstream, and primers were designed using the MethPrimer program (Li and Dahiya, 2002) to screen the putative promoter region by discrete, generally overlapping segments (Fig. 6-1). Primer sequences and segment locations, sizes, CpG number and reaction characteristics for adult heart and liver HRM PCR are shown in Table 6-1. Full genomic sequences of the segments are shown in Supplemental Table 6-1 (Appendix B).

The 25 µl HRM reactions were performed using the EpiTect® HRM™ PCR kit (Qiagen), with duplicate reactions (except where indicated) of livers (n=8) and hearts (n=8) for each of the four animals and a 100% methylated reference sample. Based on DNA input to the bisulfite reactions as well as post-conversion Nanodrop quantification checks of 10% of the samples, DNA template concentrations used in the HRM reactions were 25-30 ng for amplicons under 200 bp and 50-60 ng for longer amplicons. Reaction conditions followed the manufacturer's instructions of an initial activation step of 5 min at 95 ºC, and cycles of 10 s denaturation at 95 ºC, 30 s annealing at 55 ºC and extension at 72 ºC with duration determined by amplicon length (generally, 10 s for products up to 150 bp, 8 s per 100 bp for larger products). Some longer segments were pre-amplified

using TaKaRa EpiTaq^{TM} HS polymerase kit (Clontech; see conditions for first-round PCR in Section 6.2.5) for bisulfite-treated DNA followed by semi-nested HRM reactions as noted in Table 6-1. High-resolution melting was conducted in 2 s, 0.1 ^{o}C increments from 65 ^{o}C to 95 ^{o}C, or 5 ^{o}C below the temperature at which half the amplicons are melted (Tm) to 5 ^{o}C above the Tm, when the Tm was known from preliminary trials.

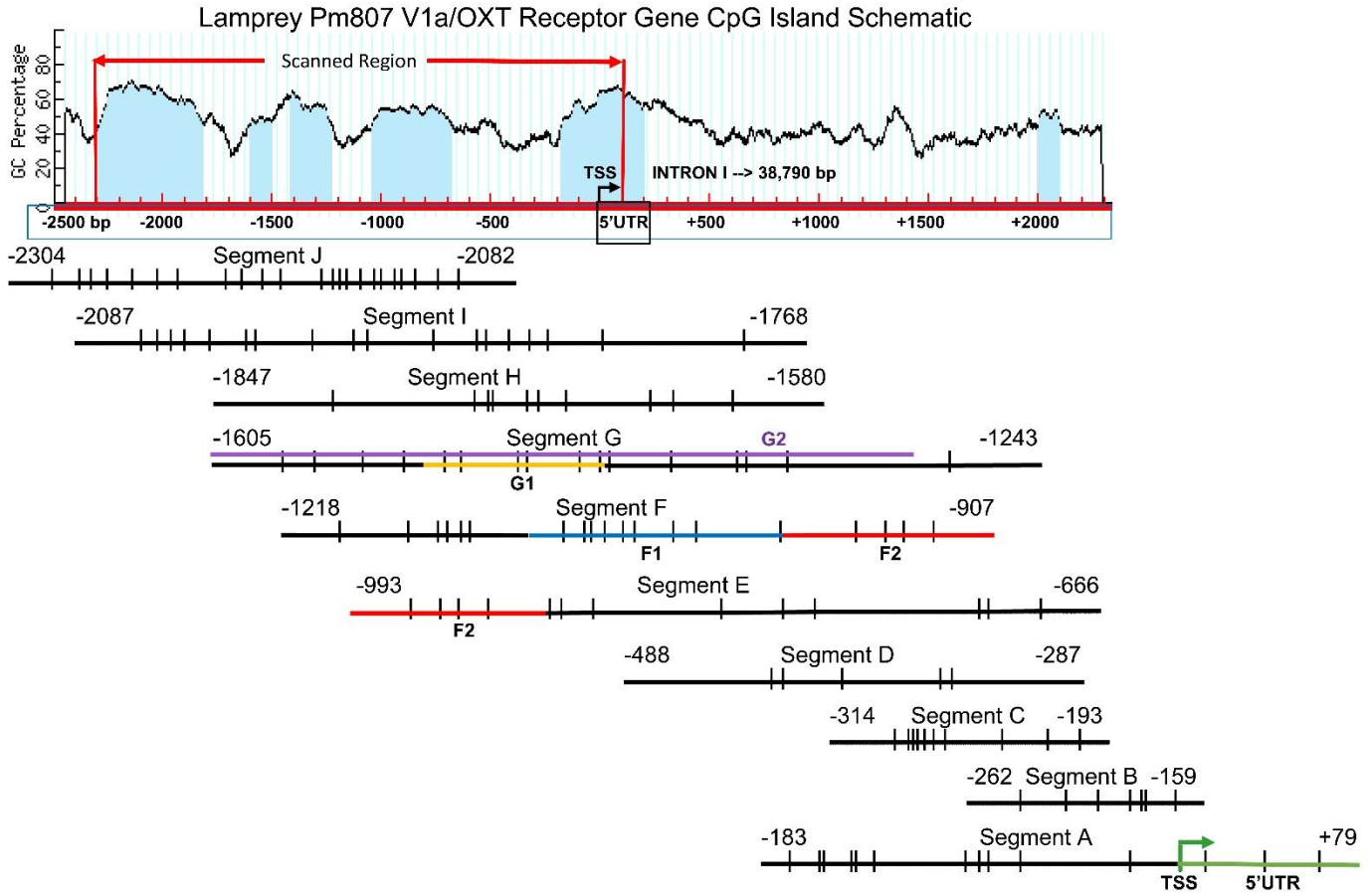


Figure 6-1. Schematic depicting the CpG island locations in the putative promoter region of the sea lamprey Pm807 vasotocin V1/OXT receptor gene constructed using MethPrimer online software (Li and Dahiya, 2002). A CpG island is a DNA stretch at least 100 base pairs (bp) long, >50% guanine/cytosine (GC) content, and observed/expected ratio of CpG dinucleotides >0.6. The region was divided into segments for which primers were designed to screen bisulfite-modified genomic DNA using High Resolution Melt PCR. Segments show location relative to the transcription start site (TSS) extending approximately 2300 bp upstream to 79 bp downstream into the 5'-untranslated region (UTR; black box). Genomic locations and characteristics of the segments are shown in Tables 1 and 2. Vertical lines on the segments indicate CpG locations.

Table 6-1. Characteristics of segments in the promoter region of the lamprey Pm807 vasotocin receptor screened for tissue-specific methylation in adult heart and liver DNA by High Resolution Melt (HRM) PCR. Primer sequences are designed to amplify bisulfite-modified DNA.

Segment	Scaffold GL477135	Relative to TSS		Primer sequences 5'-3'	Size (bp)	CpGs[1]	Total Cycles[2]
A*	241,714-241,975	-183 to +79	For Rev	GAAGTTGTAGAGYGTAATTTTTAGA CCTACAAATAAAAAACCRTACTATA	262	14	36
B	241,951-242,054	-262 to -159	For Rev	TGAGTGGTTGAAATTAGTAGTTT TCTAAAAATTACRCTCTACAACTTC	104	6	45
C	241,985-242,106	-314 to -193	For Rev	AGTTGTAAAGTTAGTTGTAAAGTTTGTT CTCATTATAAACRTCCCTTACCTCT	123	9	45
D	242,079-242,279	-488 to -287	For Rev	GGTTTGATAAAATAAATTTAGGAAATATTT AACAAACTTTACAACTAACTTTACAACTAA	201	5	45
F	242,698-243,009	-1218 to -907	For Rev	TTATTGTTTTGTGTGTATTTATTAA AATATCATTCACTACTAATTTTAACC	312	19	60
F1	242,766-242,918	-1123 to -971	For Rev	TGTGTATGTGTATATAGTTTGGTATAG TAATTCACTCATTCCTATTACCACC	153	9	45
F2	242,698-242,784	-993 to -907	For Rev	AATAGGAATGAGTGAATTATATATTG AATATCATTCACTACTAATTTTAACC	87	4	45
G2*	243,107-243,396	-1605 to -1316	For Rev	GGATGAAATGGGATTATTTTGTTTA AAAATAAAATAATTTACCAAAATTACACAT	290	15	36
H*	243,371-243,638	-1847 to -1580	For Rev	GTAATTTTAAGGAAATATAAAAAGTAGTTT TTAAACAAAATAATCCCATTTCATC	268	10	45
I	243,559-243,878	-2087 to -1768	For Rev	TAATGTAGTTAGGGTTGTGTTTAGGTT ATTACACACACAATATAATTAAAATTC	320	18	45
J	243,873-244,096	-2304 to -2082	For Rev	GTGGGGAAGTTTTTGGAATT ACATTACCAACAACTATCTAACCAC	224	23	45

*Segments A and H were pre-amplified using the same primers. Segment G2 was pre-amplified using semi-nested primers (primers for Segment G [Table 2] were used in the first round). See Section 2.4 for details. TSS – transcription start site. Y = C/T; R = A/G. [1]Number of C-G dinucleotides between the forward and reverse primers. [2]Total cycles using EpiTect HRM PCR kit on RotorGene Q thermocycler.

DNA isolated from livers of larval-stage animals were tested along with adults in the target region (Segments F and G), and downstream segment (E). Unmethylated reference DNA was also produced from adult heart and liver samples by first conducting standard PCR to cause the loss of methylation in the amplified product. Primers were designed for an 840 bp length of the native genomic DNA spanning just upstream and downstream of the target segments (Tables 6-1 and 6-2) as follows: Segment G forward TGGTTGGAATCGGCTCTAATG, Segment F reverse GCTGAGTTTGTCACGGAATTATG. GoTaq Master-Mix (Promega) was used in 50 µl reactions with 2 µl larval liver genomic DNA under a thermocycler profile of 2 min at 95 ^{o}C followed by 30 cycles of 30 s at 95 ^{o}C, 30 s at 55 ^{o}C, and 60 s at 72 ^{o}C with a 5 min elongation step at 72 ^{o}C. The product size was verified as a single band on a 1% agarose gel (data not shown) and purified using the QIAquick® PCR Purification Kit (Qiagen), eluted with nuclease-free water. The purified product was subsequently bisulfite converted for use in HRM assays.

Table 6-2. Characteristics of target segments in the promoter region of the lamprey Pm807 vasotocin receptor in adult heart, adult liver, and larval liver DNA analyzed by High Resolution Melt (HRM) PCR. Primer sequences are designed to amplify bisulfite-modified DNA, except for methylated-specific and unmethylated-specific primers designed for Segment G1; CpG positions 5 and 6 (For) and 9 and 10 (Rev) are underlined.

Segment	Scaffold GL477135	Relative to TSS		Primer sequences 5'-3'	Size (bp)	CpGs [1]	TC [2]
E	242,457-242,784	-993 to -666	For Rev	AATAGGAATGAGTGAATTATATATTG AAAATACAACTCTATTAAAAAACCC	328	13	50
F	242,698-243,009	-1218 to -907	For Rev	TTATTGTTTTGTGTGTATTTATTAA AATATCATTCACTACTAATTTTAACC	312	19	52
G	243,034-243,396	-1605 to -1243	For Rev	GGATGAAATGGGATTATTTTGTTTA CTCATTATAAACRTCCCTTACCTCT	363	16	45
G1 meth	243,224-243,312	-1521 to -1433	For Rev	GATATTTTATAGTTGGA<u>C</u>GTTTTG<u>C</u>GGA TAC<u>G</u>TAATAACC<u>G</u>AACTCCACACAATAA	89	2	40
G1 unmeth	243,224-243,312	-1521 to -1433	For Rev	GATATTTTATAGTTGGA<u>T</u>GTTTTG<u>T</u>GGA TAC<u>A</u>TAATAACC<u>A</u>AACTCCACACAATAA	89	2	40

TSS – transcription start site. R = A/G. [1]Number of C-G dinucleotides between the forward and reverse primers.
[2]TC - Total cycles using EpiTect HRM PCR kit on RotorGene Q thermocycler.

6.2.5 PCR for direct sequencing

Upon identifying segments (F and G2) with significant differences in melt curve shape between the adult heart and adult liver, the TaKaRa EpiTaq^{TM} HS polymerase kit (Clontech) for bisulfite-treated DNA was used to amplify Segments F and G for standard dideoxynucleotide (Sanger) DNA sequencing. The PCR master mix component final concentrations were 1.25 U Epitaq HS, 1X buffer, 2.5 mM $MgCl_2$, 0.3 mM dNTP mixture, 0.4 µM each primer, 50-60 ng bisulfite-converted DNA template, and nuclease-free water to 25 µl. PCR conditions were denaturation at 98 ^{o}C followed by 35 cycles of 10 s at 98 ^{o}C, 30 s at 55 ^{o}C, and 30 s at 72 ^{o}C with a 5 min elongation step at 72 ^{o}C. To produce enough product for sequencing, the process was repeated with 3 µl of the first-round product as template in 50 µl reactions. Product sizes were verified as single bands on a 1% agarose gel (data not shown) and purified using the QIAquick® PCR Purification Kit (Qiagen), eluted with nuclease-free water. The purified products were sequenced by GE SeqWright (Houston, TX) for adult samples and 100% methylated reference samples using an Applied Biosystems (ABI) 3730xl sequencer. The larval DNA and unmethylated reference samples were prepared and sequenced in the same manner as the adult DNA, but were sequenced at the University of Minnesota Genomics Center (St. Paul, MN).

To corroborate sequence results, we focused on a sub-segment (G1) in which two sets of primer pairs (Table 6-2) were designed to imitate the fully methylated (G1 meth) and unmethylated (G1 unmeth) states at CpGs 5-6 and 9-10. This assay was designed to reveal the differences in methylation level at CpGs 7 and 8 among the different tissues, as

well as the relative melt characteristics of the unmethylated to fully methylated states of all six CpGs. Except where noted, all assays for the target region included duplicate reactions for each of the four larval samples (n=8), two unmethylated (n=4) and two 100%-methylated (n=4) reference samples, and heart (n=4) and liver (n=4) DNA from two adult animals (animals 3 and 4). HRM assays were conducted as described for adults (Section 6.2.4) with DNA input adjusted as described in Section 6.2.2.

6.2.6 Data and statistical analysis

Fluorescence was normalized using the RotorGene Q HRM software. For greater clarity of the graphed curves, composite curves for each tissue type were generated by calculating means and population standard deviations at half-degree increments. This method provided adequate information for screening purposes.

6.3 Results

6.3.1 HRM screening analyses of putative promoter region DNA shows expression-correlated differences in melt curves

We performed HRM of bisulfite-modified lamprey DNA on overlapping sequence segments ranging from 87 to 363 bp in length (Fig. 6-1), from a tissue where the Pm807 V1a/OXT receptor gene is highly expressed (adult heart) and a tissue where the gene is not expressed (adult liver). The resulting normalized melt curve profiles and standard deviations among the relative fluorescence values at 0.5 ^{o}C temperature increments indicated negligible differences in the methylation levels between tissues for Segments A, B, C, D, or H (Fig. 6-2). High concentrations of CpGs in Segment I and J made primer design difficult, and may also have caused multiple melt peaks (data not shown) hindering interpretation. We also attempted to amplify the sequence between Segments D and E (Suppl. Table 6-1, Appendix B) containing 9 CpGs, but obtained no reaction.

Normalized fluorescence melt curves produced for Segments F and G2 showed differences between the adult heart and liver tissues that were consistent among all four animals tested. The curves for both segments indicate a higher temperature at melt start (take-off), more precipitous decline and lower temperature at melt completion (touchdown) for the liver than for heart samples. Segment F was further subdivided by Segment F1 containing the middle 9 CpGs and Segment F2 containing the last 4 CpGs. Similar composite curves were obtained but with smaller differences and less consistency within individual animals, as reflected in the observed standard error (Fig. 6-3). The Segment G2 melt curves showed a high degree of alignment among all individuals for each tissue type.

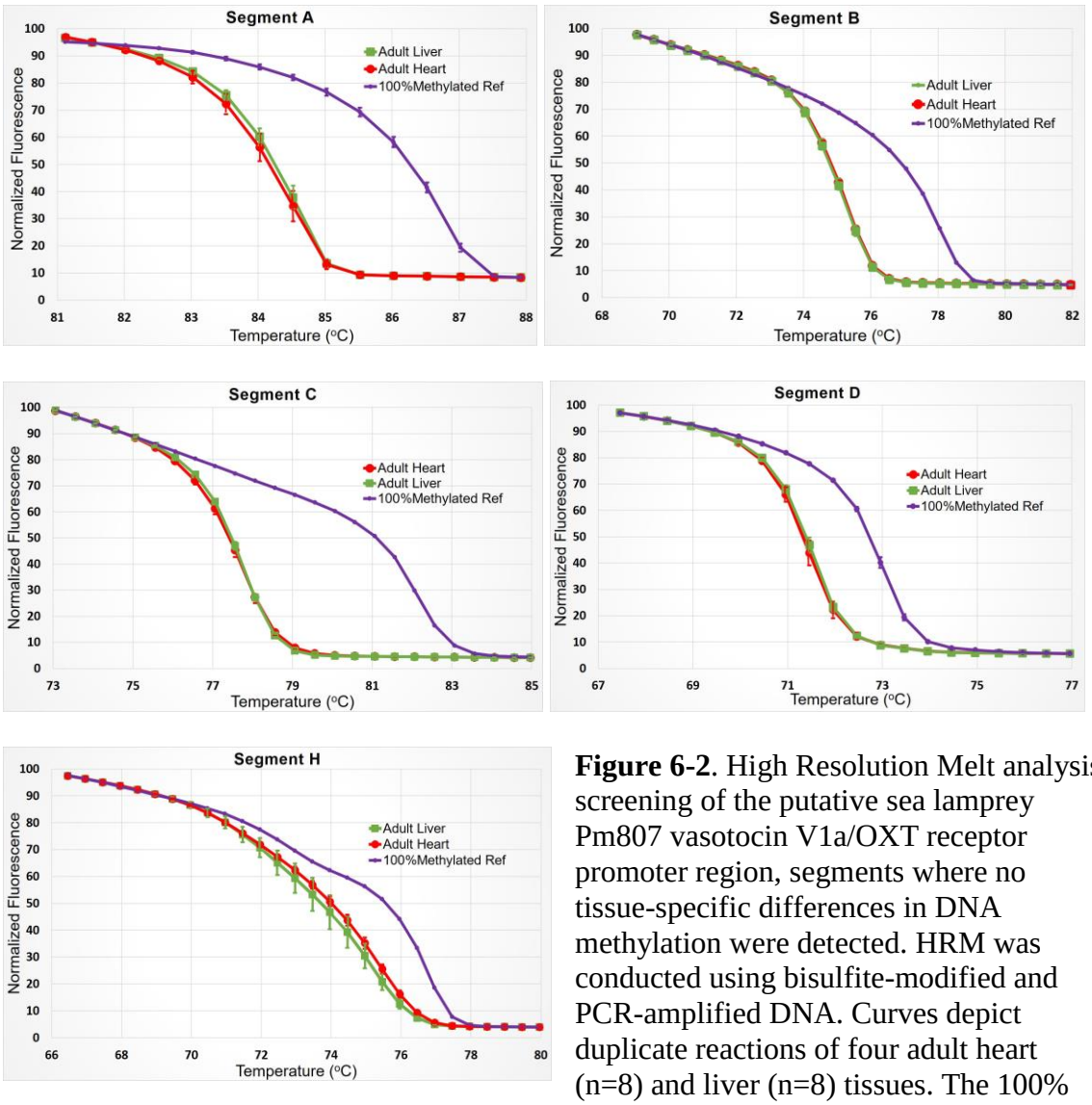


Figure 6-2. High Resolution Melt analysis screening of the putative sea lamprey Pm807 vasotocin V1a/OXT receptor promoter region, segments where no tissue-specific differences in DNA methylation were detected. HRM was conducted using bisulfite-modified and PCR-amplified DNA. Curves depict duplicate reactions of four adult heart (n=8) and liver (n=8) tissues. The 100% methylated reference was created by treating the DNA with SsssI methylase and S-adenosyl methionine prior to bisulfite conversion (Segments A and D n=4; Segments B, C and H n=2). See Table 1 for segment primers, characteristics and locations.

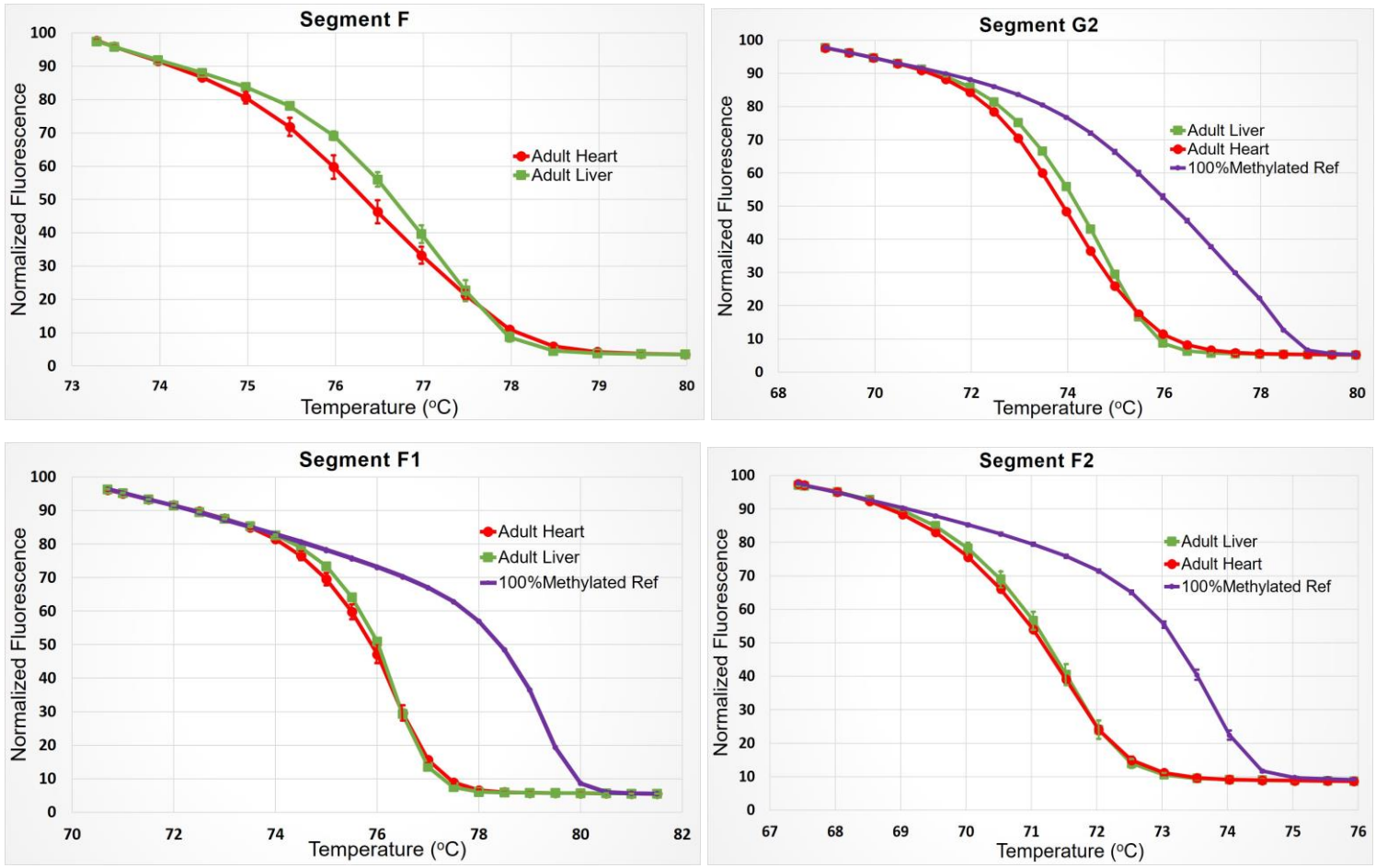
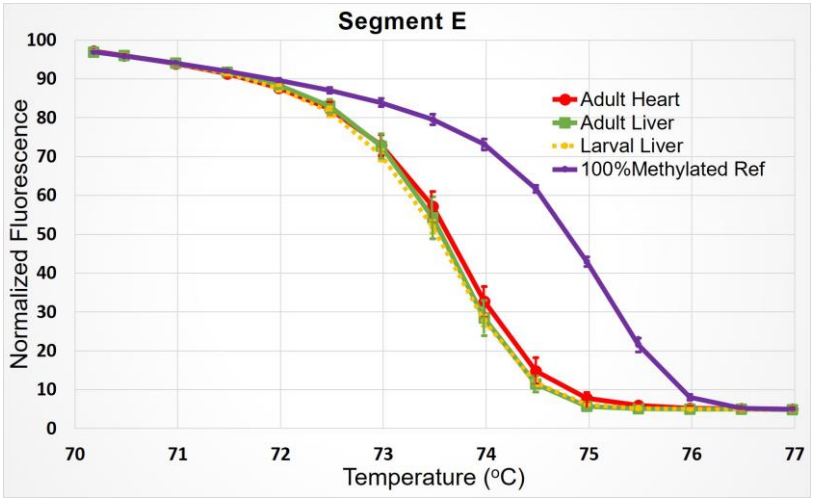


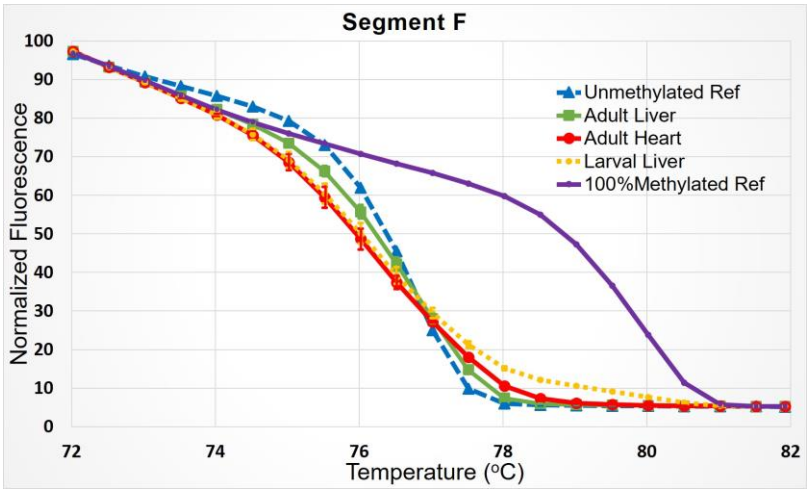
Figure 6-3. High Resolution Melt analysis screening of the putative sea lamprey Pm807 vasotocin V1a/OXT receptor promoter region, segments where tissue-specific differences in DNA methylation were detected. HRM analysis was conducted using bisulfite-modified and PCR-amplified DNA. Curves depict duplicate reactions of four adult heart (n=8) and liver (n=8) tissues (except Segments F1 and F2, run in triplicate). The 100% methylated reference was created by treating the DNA with SsssI methylase and S-adenosyl methionine prior to bisulfite conversion (Segments G2 n=4; F1 and F2 n=3). See Table 1 for segment primers, characteristics and locations.

To determine whether the Pm807 gene is expressed as mRNA in the livers of the larval-stage animals, polymerase chain reaction (PCR) assays were conducted on larval liver cDNA. Pm807 V1a/OXTR gene expression as mRNA in larval livers was demonstrated in samples from all four animals (Fig. 6-4). No bands were visible in the mock RT negative control reactions. Therefore, with expression differing in liver between life stages, we next tested whether DNA isolated from livers of larval-stage animals differed in melt characteristics from adult DNA in the target region (Segments F and G), and the neighboring downstream segment (E). HRM analysis for these segments also included an unmethylated reference in addition to the 100% methylated reference (Fig. 6-5). The unmethylated reference material was PCR generated only after segments with differential melt curves were discovered in the adult curves for Segments F and G2. The resulting curve for the unmethylated reference for Segment F features a higher takeoff temperature and precipitous drop similar to the adult liver curves. The amount of unmethylated template used to normalize the PCR reactions was estimated due to the requirement of cycling prior to bisulfite conversion. The adult liver samples, however, more closely followed the unmethylated curve, supporting the lower level of methylation observed in the adult tissue bisulfite sequencing results. The larval liver curve was more closely aligned with that of the adult heart but with a higher touchdown temperature. In addition to having a high CpG density, Segment F also had a high guanine concentration due to several tandem GT repeats. These characteristics may have resulted in less efficient PCR, requiring more cycles to reach a plateau for Segment F than for the longer Segment G.

The Segment G curves showed less difference than for Segment F in takeoff temperatures between the unmethylated reference and adult liver vs. the adult heart and larval liver (Fig. 6-5). The larval liver samples again reached the highest touchdown temperature. All tissue sample curves were closer to the unmethylated than the methylated reference, indicating that overall methylation is low throughout these sequences.

Figure 6-4. Pm807 V1a/OXT receptor mRNA expression in the larval liver. PCR products are 285bp Pm807 primers on cDNA (Lane 1) and mock cDNA (no reverse-transcriptase) (Lane2); and 280 bp actin primers on cDNA (Lane 3) for larval livers 1-4.





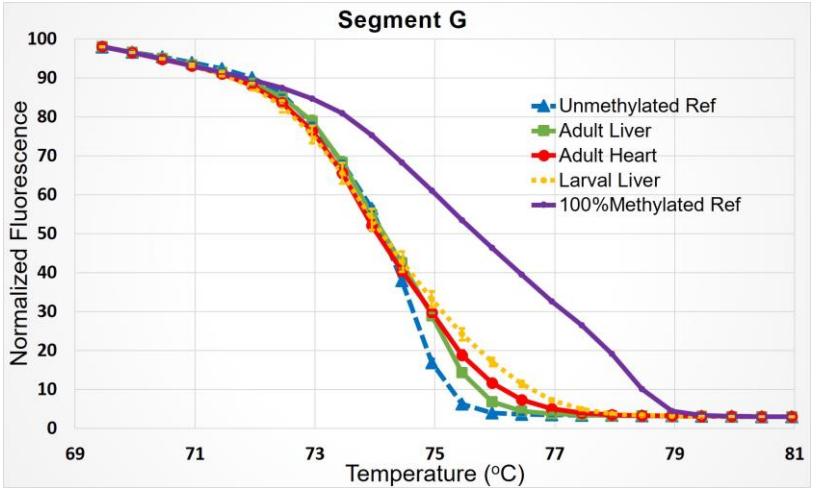


Figure 6-5. Target segments of the sea lamprey Pm807 vasotocin V1a/OXT receptor putative promoter region investigated for tissue-specific and life-stage-specific differences in DNA methylation. High Resolution Melt analysis was conducted using bisulfite-modified and PCR-amplified DNA. Curves depict duplicate reactions of two adult heart (n=4) and liver (n=4), and four larval liver (n=8) tissues. The 100% methylated reference was created by treating the DNA with SsssI methylase and S-adenosyl methionine prior to bisulfite conversion (Segment E n=4; F and G n=2). The unmethylated reference was created by PCR of the target region to abolish methylation marks prior to bisulfite conversion (Segment F n=2; G n=4). See Fig. 1 and Table 2 for segment primers, characteristics and locations.

6.3.2 Sequencing of target segments shows a pattern of CpG methylation

HRM screening results identify segments F and G as potentially harboring CpGs with tissue-specific methylation levels. In order to determine specifically which CpGs were responsible for differential HRM curves, we sequenced the full 312 bp Segment F and 363 bp Segment G bisulfite-converted DNA from all adult and larval samples, as well as 100% methylated and unmethylated reference samples from one adult heart and one liver sample. The sequences were consistent with the genomic sequence accounting for bisulfite conversion. The methylated reference sample sequences showed only cytosine at the expected CpG dinucleotides, indicating successful *in vitro* methylation with SssI methylase. Unmethylated reference samples had no cytosine peaks, indicating complete bisulfite conversion. Normalized sequence traces aligned using Geneious software (version 8.1.2) for both Segments F (Fig. 6-6) and G (Fig. 6-7) highlighting only the cytosine peaks also showed complete bisulfite conversion with no residual non-CpG cytosines.

In both segments the level of methylation was greater in the adult heart and larval liver than in the adult liver, and a pattern of partially methylated vs. unmethylated CpGs was clearly consistent among all animals. The results for Segment F showed a sequencing artifact between CpGs 14 and 15 in the adult sequencing runs. Also, possible slippage of the PCR polymerase in all samples starting at CpG 14 (3'-5'), likely originating in a GT repeat sequence, appears to have created second and in some samples third sequences, offset by two and four nucleotides (Fig. 6-6).

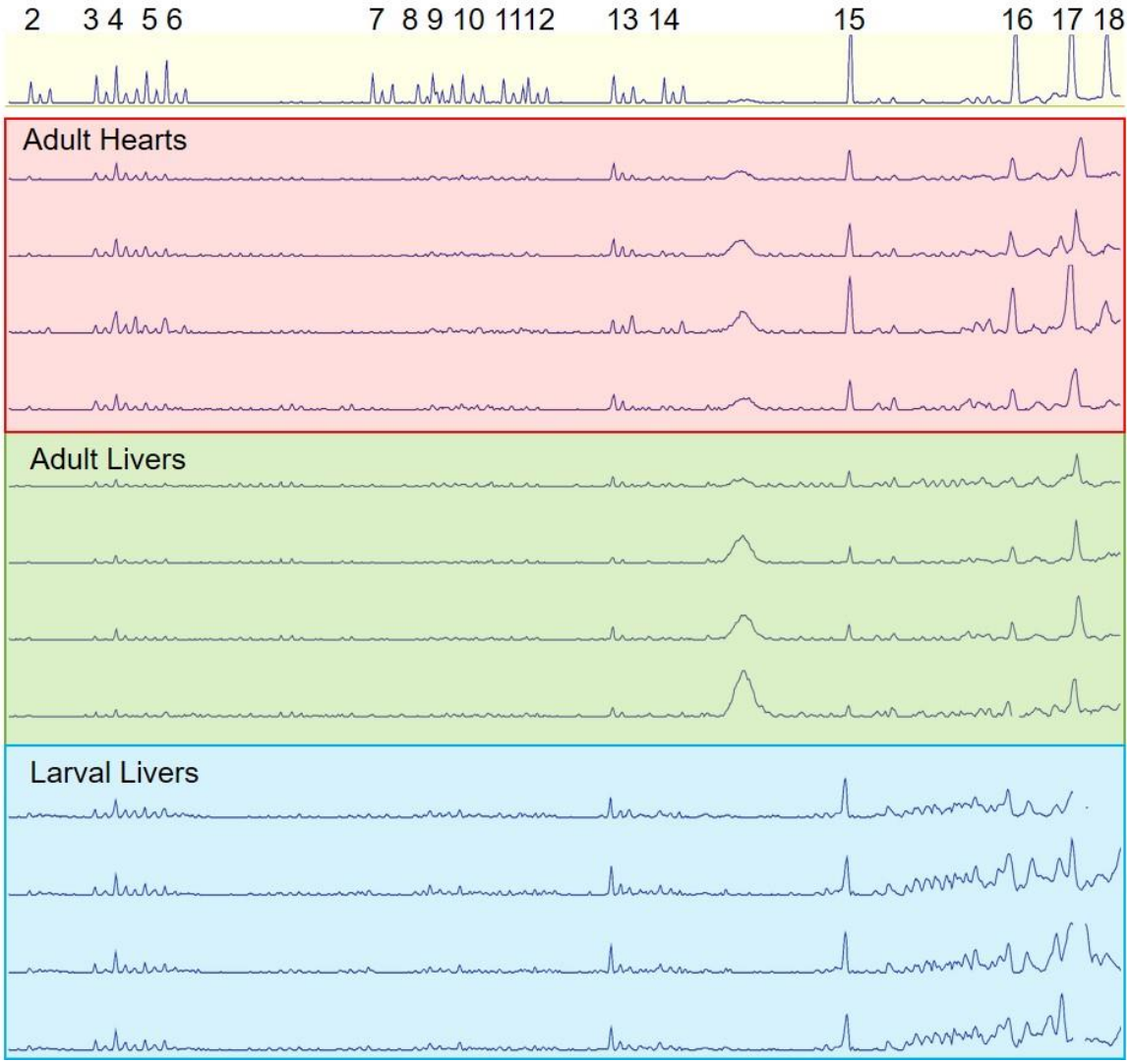


Figure 6-6. Segment F direct dideoxynucleotide (Sanger) DNA sequencing results of bisulfite-modified PCR products from four adult and four larval lampreys (adult heart and liver samples are sequentially from the same individuals). Genomic location and characteristics are shown in Fig. 6-1 and Table 6-1. The normalized trace was modified to show only the cytosine peaks (representing methylated cytosines), numbered in the 5'-3' direction. The top trace depicts 100% methylated reference DNA created by treating the genomic DNA with SsssI methylase and S-adenosyl methionine prior to bisulfite conversion. A sequencing artifact appears between CpGs 14 and 15 in the adult samples, CpG 1 (not shown) was not methylated in any sample, and CpG 19 (not shown) gave ambiguous results.

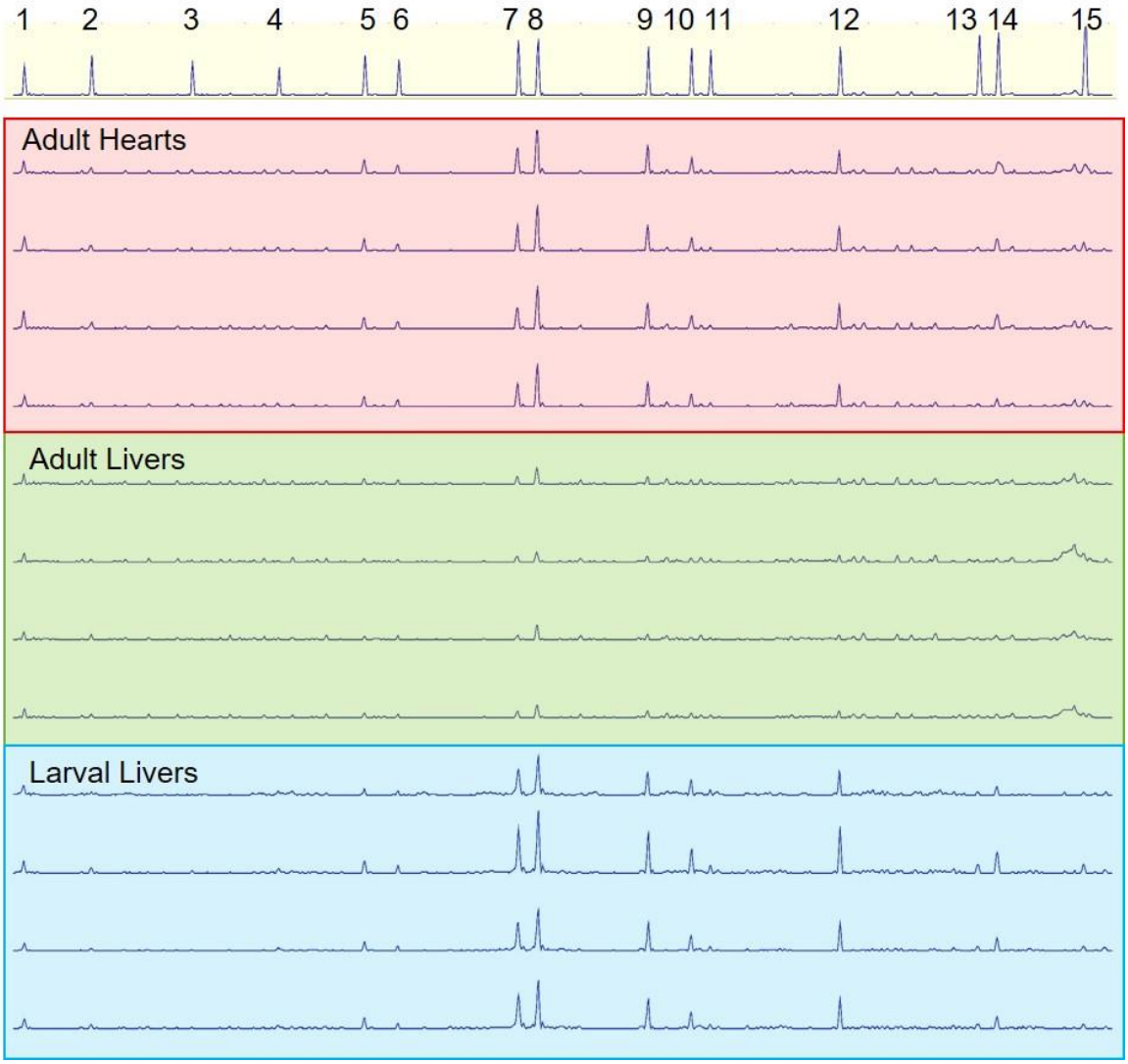


Figure 6-7. Segment G direct dideoxynucleotide (Sanger) DNA sequencing results of bisulfite-modified PCR products from four adult and four larval lampreys (adult heart and liver samples are sequentially from the same individuals). Genomic location and characteristics are shown in Fig. 6-1 and Table 6-1. The normalized trace was modified to show only the cytosine peaks representing methylated cytosines, numbered in the 5'-3' direction. The top trace depicts 100% methylated reference DNA created by treating the genomic DNA with SsssI methylase and S-adenosyl methionine prior to bisulfite conversion. CpG 16 (not shown) was not methylated in any sample.

In Segment F heart and larval liver samples, CpGs 3, 4, 5, 6, 13, 15, 16 and 17 were partially methylated; in adult liver samples CpGs 13, 15, 16 and 17 were also partially methylated, but methylation was very low or not discernable above background at 3, 4, 5 and 6. No methylation was detected at CpGs 1, 2, 7-12, 14 or 18 in any sample, and 19 was ambiguous due to its position at the end of the sequence. The more pronounced differences between tissues in CpGs 3 through 6 may account for the melt curve difference for the whole segment, considering that negligible differences were seen in Segment F1 covering CpGs 7 through 15 and F2 covering CpGs 16 through 19 (Fig. 6-3). This segment gave the first indication that methylation was higher in tissues with expressed mRNA than in the adult liver with no expression.

Segment G provided a clean sequence trace with minimal background, and showed more pronounced differences between tissues in methylation level than were seen in Segment F. Heart DNA showed moderate methylation at CpGs 5 and 6 and high methylation at CpGs 7, 8, 9 and 12, whereas adult liver samples showed low methylation at CpGs 7 and 8, and little to no methylation at the rest of the sites (Fig. 6-7).

Results for the larval liver samples support the similarity in CpG methylation levels and patterns seen in the adult heart samples, differing from the adult liver in both Segment F (Fig. 6-8) and Segment G (Fig. 6-9). Sequencing software tends to over-compensate for the loss of cytosines by exaggerating the remaining cytosine peaks when normalizing the peak heights with the other three bases (Boyd et al., 2007). Therefore, Figures 6-8 and 6-9 are composed of representative raw traces as a potentially more accurate picture of the true methylation levels. In Figures 6-8 and 6-9, peaks in the

cytosine trace proportionately show the level to which a CpG position was methylated in the tissue in relation to the thymine peaks at the same position, which indicate converted, unmethylated cytosines. These raw traces clearly show the positive correlation between mRNA expression and methylation level at specific CpGs, in tissue-specific patterns.

To further corroborate the sequencing results, Segment G1 primers representing methylated and unmethylated states were designed around the differentially methylated CpGs 5 through 10, a sequence of 89 bp. With the annealing temperature held relatively low (55^{o}C), all primers appeared to bind regardless of the methylation state of the sample (i.e., "methylated-specific" primers that contained cytosines at CpGs 5, 6, 9, and 10 bound to unmethylated reference DNA that contained only thymines, and "unmethylated-specific" primers that contained thymines rather than cytosines at those CpG positions bound to the 100% methylated reference DNA). The assay detected natural variability in methylation (not dictated by the primers) at CpG positions 7 and 8. The curves produced (Fig. 6-10) show more pronounced differences among the untreated samples and references than in longer segments, and follow the expected rule of higher melt temperature for higher methylation levels.

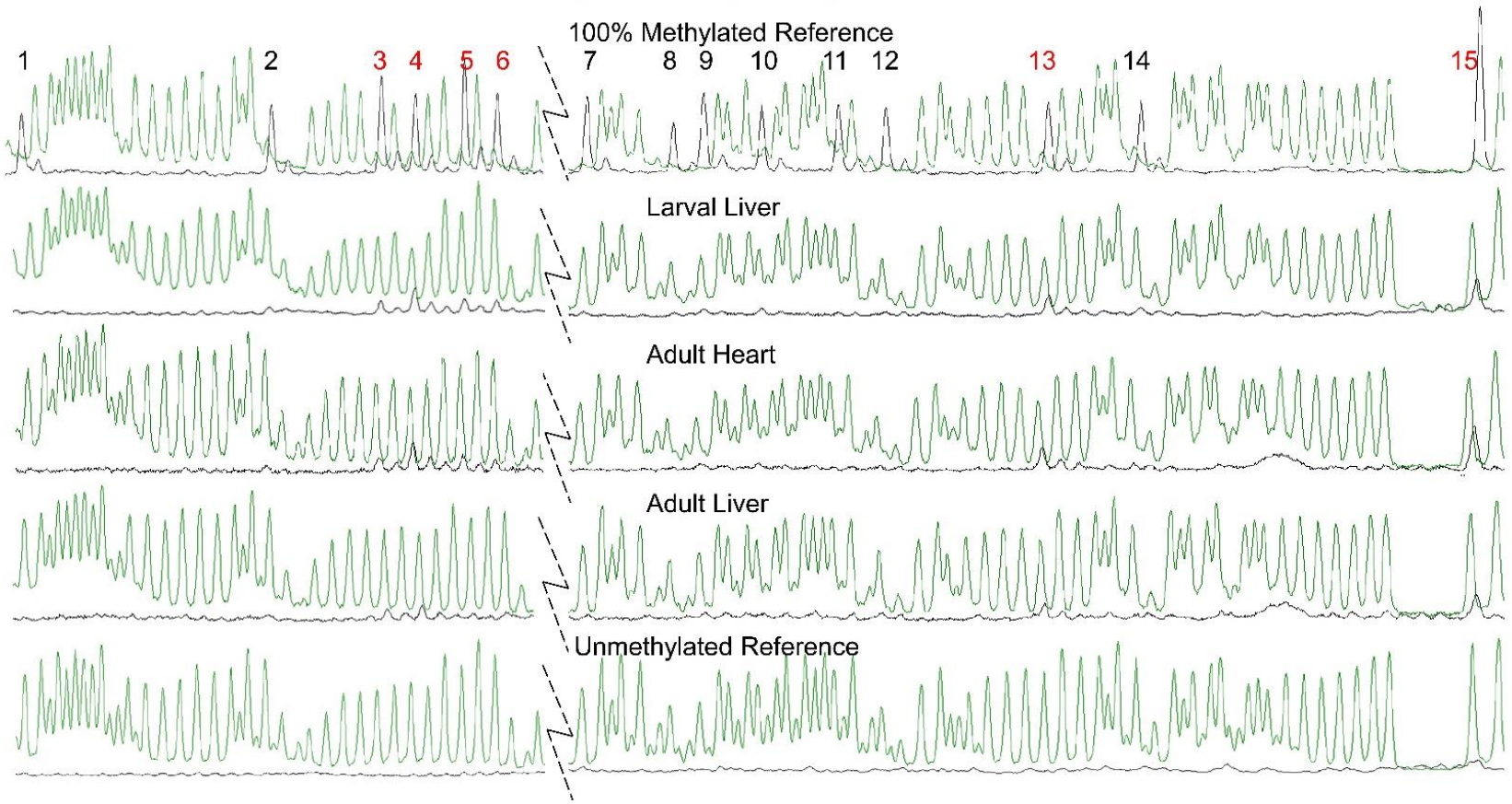


Figure 6-8. Representative Segment F raw traces from dideoxynucleotide (Sanger) DNA sequencing of bisulfite-modified PCR products from individual adult and larval lampreys. Genomic location and characteristics are shown in Fig. 6-1 and Tables 6-1 and 6-2. The traces are modified to show only the cytosine peaks (black) representing methylated cytosines, relative to the thymine peaks (green) representing the unmethylated state at the CpG locations numbered in the 5'-3' direction. The top trace depicts 100% methylated reference DNA. The unmethylated reference (bottom) was created by PCR of the target region to abolish methylation marks prior to bisulfite conversion. Red numbers indicate CpGs that show some degree of methylation in the tissue DNA.

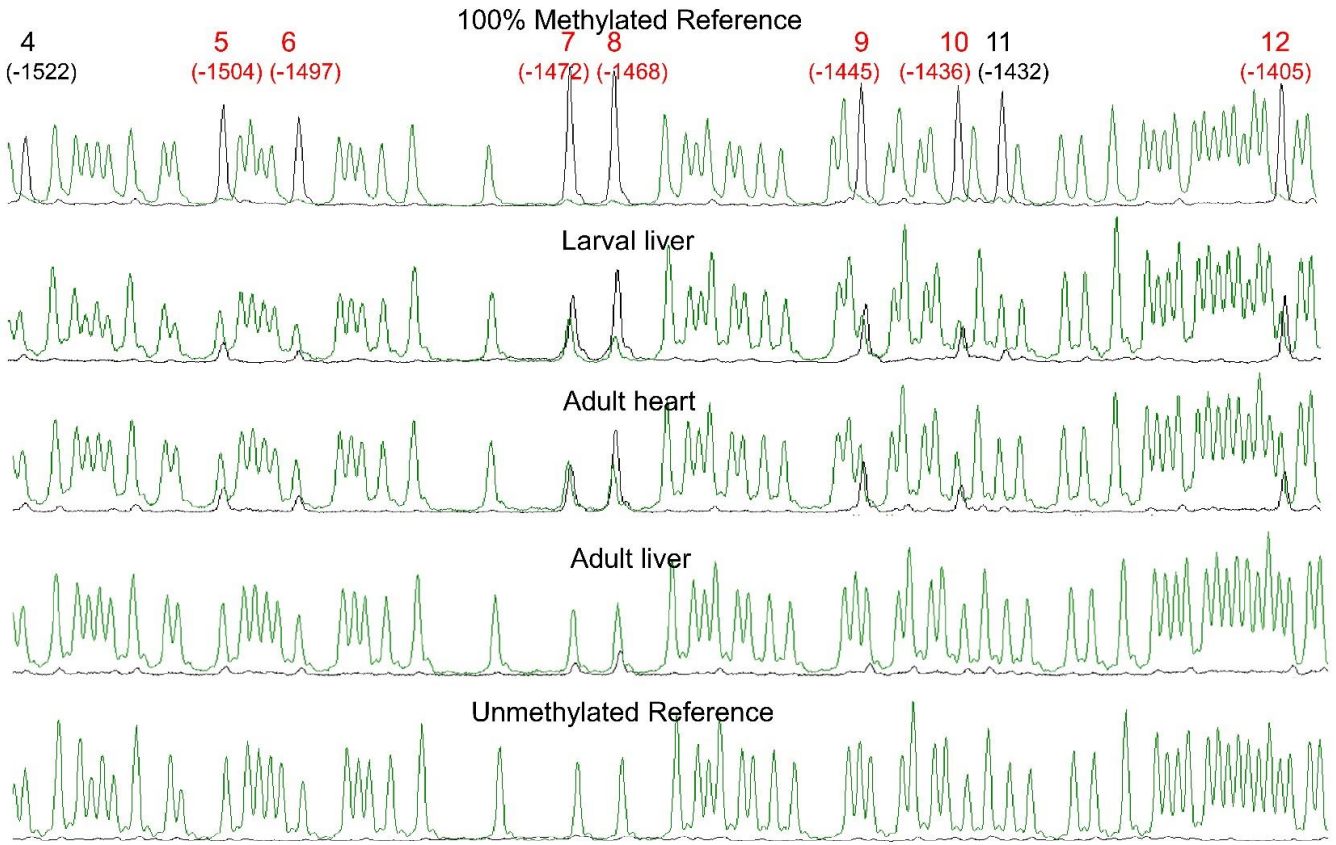


Figure 6-9. Representative Segment G raw traces from dideoxynucleotide (Sanger) DNA sequencing of bisulfite-modified PCR products from individual adult and larval lampreys, focusing on the area of differential methylation. The traces are modified to show only the cytosine peaks (black) representing methylated cytosines, relative to the thymine peaks (green) representing the unmethylated state at the CpG locations numbered in the 5'-3' direction. The top trace depicts 100% methylated reference DNA. The unmethylated reference (bottom) was created by PCR of the target region to abolish methylation marks prior to bisulfite conversion. Red numbers indicate CpGs that are differentially methylated in the heart, larval liver and adult liver tissue DNA, with numbers in parentheses indicating the upstream location relative to the transcription start site.

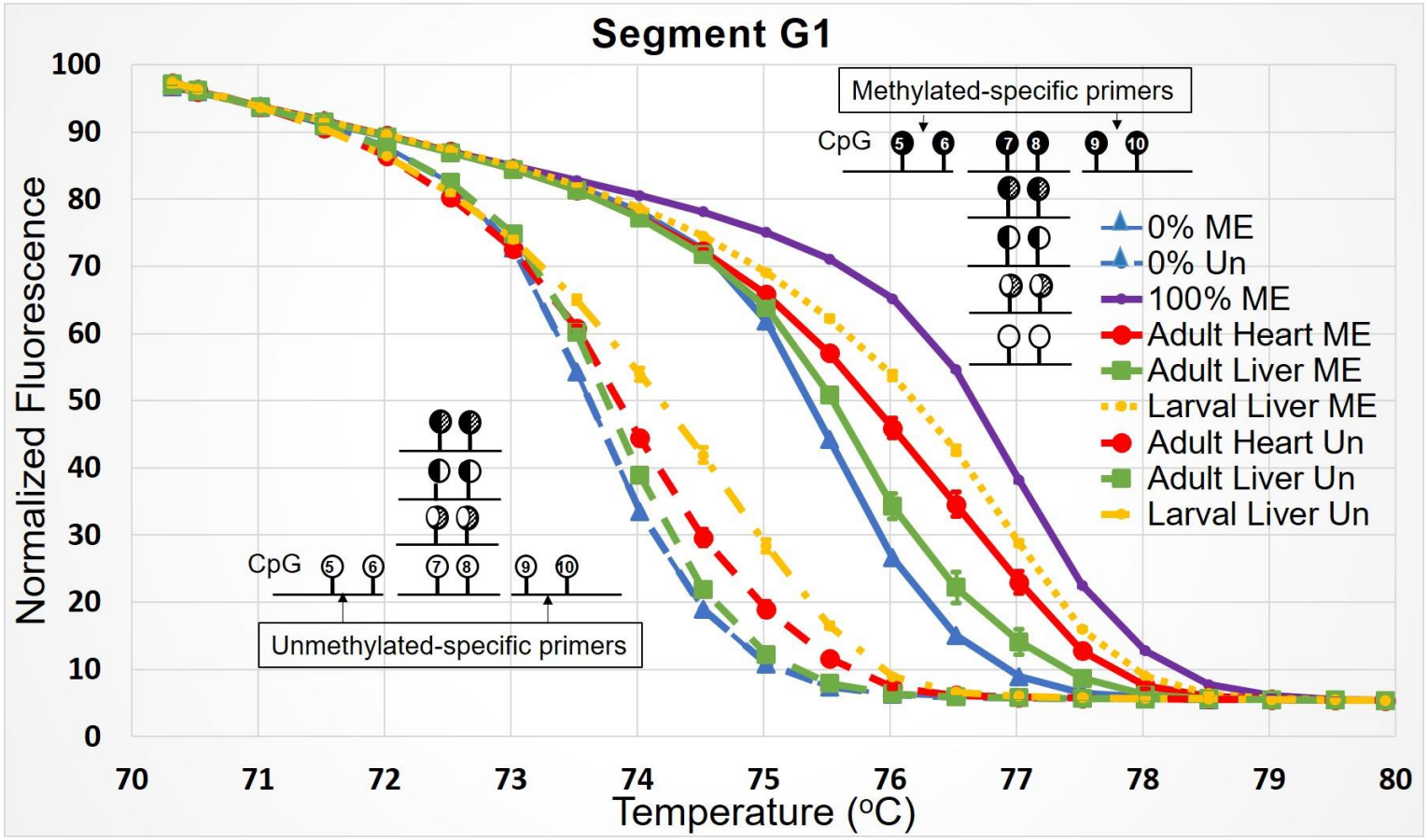


Figure 6-10. Segment G1 High Resolution Melt analysis in the sea lamprey Pm807 vasotocin V1a/OXT receptor putative promoter region investigating tissue-specific and life-stage-specific differences in DNA methylation. HRM was conducted using bisulfite-modified and PCR-amplified DNA. Curves depict duplicate reactions of two adult heart (n=4) and liver (n=4), and two larval liver (n=4) tissues. Circles indicate relative levels of methylation at CpG locations 5-10 (Fig. 6-9), from open for unmethylated to black for fully methylated. Primers for curves at left (dotted lines) represented unmethylated CpGs, primers for curves at right (solid lines) represented methylated CpGs. See text and Table 6-2 for primer design details.

Our results showed, for most of the screened segments of the Pm807 gene promoter region, that the melt curves did not differ between the tissue types. Curve shape similarity predicts that any methylation is occurring at the same CpGs, since melting rate and therefore melt curve shape is influenced by the locations in the sequence of C-G triple bonds between strands (Dwight et al. 2011). However, the HRM results for the sequence from approximately -1605 to -907 bp upstream of the TSS (Segments G and F) indicated repeatable melt curve differences between tissues corresponding to whether the gene is or is not transcribed. The high takeoff temperature and precipitous drop of the unmethylated reference curve, similar to the adult liver curves, may be due to greater efficiency generally seen in PCR of sequences that are less methylated (i.e., with lower GC content), resulting in more product (Warnecke et al., 1997). Sequencing also revealed consistent patterns of methylated vs. unmethylated CpGs in Segments G and F that clearly correlate between the two tissues with mRNA expression (adult heart and larval liver) and differ from the adult liver where no transcription of DNA has been detected.

6.4 Discussion

The general assumption for CpG island DNA methylation in mammals is that higher methylation correlates with lower transcription or gene silencing. However, variations exist in mammals and at other phylogenetic levels that become evident with deeper investigations of individual genes. Specifically, in the human OXTR promoter higher methylation at known CpGs inactivates transcription in liver (Kusui et al. 2001). However, positive correlations between methylation and expression have been shown for OXTR in mice in specific brain areas (Harony-Nicolas et al., 2014), and reversals occur in the same tissues (e.g., uterus and mammary) under different physiological conditions (Mamrut et al., 2013).

The putative V1a/OXT-like receptor mRNA is expressed in the sea lamprey tissues homologous to those expressing V1a and/or OXT receptors in jawed vertebrates (Chapter 2). Because this gene is pleiotropic it must have a mechanism of turning on and off in specific tissues for functions needed through the entire life span, such as in the heart; for reproductive behaviors and functions only required in adulthood; or for metabolic needs of the growing and developing larva. In this study our HRM analyses of the region upstream the lamprey Pm807 V1a/OXTR gene TSS showed overall low methylation levels that are expected to be associated with vertebrate CpG islands. However, higher methylation was found at specific CpGs in that region in the adult heart and larval liver where the gene is expressed than in the silenced adult liver. We discovered a DNA methylation pattern that may address both tissue-specific expression regulation at the same life stage, and life-stage-specific expression in the same tissue type. The pattern

was very robust in manifesting in the same manner in four individual adults and four larval animals, but did not follow the inverse correlation between higher methylation and lower transcription levels that is usually seen in mammals.

Comparing to the well-studied mouse and human OXTR promoters, the pattern in the lamprey Pm807 putative promoter does not occur in the same location as in the mouse and human OXTR genes relative to the translation initiation site, but appears well upstream of the TSS. This may be due to the comparatively long first intron (Fig. 6-1) in the lamprey gene, and may be related to its unclear orthological connection with the V1a receptor (Chapter 2), or its unique genomic characteristics. Thus, assuming this is truly the regulatory region of the Pm807 gene, there are fundamentally two ways to explain the high methylation/high transcription association we observed. One explanation is based on a whole-genome perspective, where the inherently divergent characteristics of lamprey DNA based on its evolutionary history and phylogenetic position dictate the prevalent regulatory role of DNA methylation for the species as a whole. Alternatively, the conserved functions of neuropeptide hormone family genes dictate a gene-specific role of promoter methylation.

A basis for the whole-genome perspective on the GC content and density in sea lamprey DNA was provided with the sequencing of the lamprey genome (Smith et al., 2013). Starkly different structures and patterns were found in the lamprey compared with both invertebrate and vertebrate genomes including the unique programmed loss of ~20% of germline DNA from somatic cells during early embryogenesis (Smith et al., 2012). Furthermore, high GC content in coding regions (highest in the third position of codons),

a high degree of repetitive elements, and no correlation between codon GC bias and adjacent non-coding regions indicate that GC-content heterogeneity in the lamprey genome differs from species possessing isochore structures (Smith et al., 2013). Isochore structures occur in warm-blooded animals; the ratio of methylated CpGs (5mC) to CpG density decrease with increasing GC content, supporting the findings of low methylation at CpG islands and hence the association with promoters and gene transcription (Varriale et al., 2010). Lamprey genomic GC heterogeneity may mean that CpG islands as defined for other vertebrate species, although they exist in lamprey DNA, may not follow the same biological rules.

Alternatively, the neuropeptide gene-specific explanation finds support in two recent studies on the manner of transcription regulation by specific CpG site methylation in the mouse Oxtr promoter region (Mamrut et al., 2013; Harony-Nicolas et al., 2014). Mamrut et al. (2013) found Oxtr levels peak in the uterus and mammary glands before and during labor; uterine levels drop post-partum while mammary levels remain high. The levels of Oxtr in the uterus were responsive to estrogen, and transcription was found to increase by introducing C-to-A mutations abolishing methylation at just two CpGs – one within a specificity protein (Sp)-1 binding site and the other adjacent to a nearby half-palindromic estrogen response element (ERE) motif. In contrast, the mammary gland levels of Oxtr transcription were not responsive to estrogen, and high levels of Oxtr expression directly correlated with high levels of methylation. The authors suggest that differential Oxtr promoter regulation in the mouse is methylation dependent in a tissue- and physiological condition-specific manner (Mamrut et al., 2013). Furthermore,

methylation at these same CpG positions in mouse brain tissue were found to have a positive correlation with mRNA levels, controlling Oxtr distribution in various brain regions in an estrogen-dependent manner (Harony-Nicolas et al., 2014).

Several half-palindromic ERE sequences are present in the 2kb of the lamprey Pm807 promoter upstream of the TSS that match sequences found in human promoters, including TGACT adjacent upstream of the unmethylated CpG at -1376 relative to the TSS (Fig. 6-11A), and two within 150 bp upstream of the TSS (Fig. 6-11B). A full ERE motif sequence at -101 to -88 (GGTGGtgcgTGACC; Fig. 6-11B) matches that found in the human C3 complement gene promoter (Gruber et al., 2004), except with four rather than three spacer nucleotides (shown in lower case). Estradiol appears to activate a jawed-vertebrate-like nuclear estrogen receptor in lampreys and bind to consensus DNA ERE sequences (Paris et al., 2008). The lamprey Pm807 proximal promoter area also has a significant length of GT-rich bases consistent with Sp-1 or Sp-2 binding (Kaczynski et al., 2003) which includes several CpGs within or immediately adjacent to the consensus binding motifs. No difference in HRM curves was observed in this segment (Segment A) in these tissues, however the organization of these elements in the lamprey promoter resembles that in the mouse Oxtr promoter, indicating possible regulatory significance in estrogen-responsive lamprey tissues.

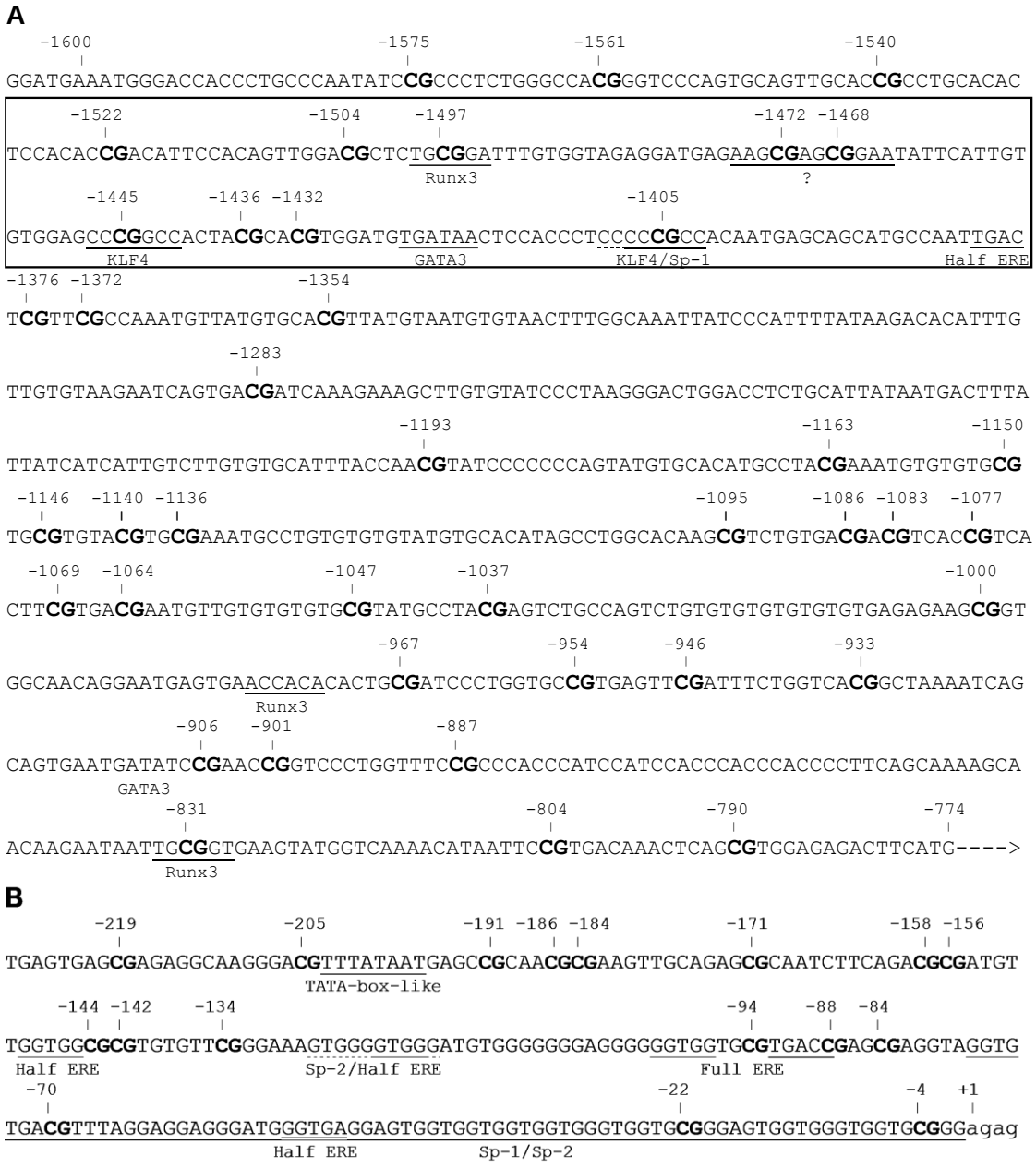


Figure 6-11. Locations of CpG dinucleotides (bold) and putative transcription factor/enhancer/response element binding motif consensus sequences (underlined and labeled) in **(A)** the target region upstream of the sea lamprey vasotocin V1a/OXT receptor transcription start site (TSS) and **(B)** the region proximal to the TSS. The sequence is shown in the 5'-3' direction and CpGs are numbered in relation to the TSS as +1. The boxed region covers Segment G1 where marked differences in methylation levels were observed among tissue types. Runx3 – Runt-related transcription factor 3; KLF4 – Krüppel-like factor 4; Sp-1/Sp-2 – Specificity proteins 1 and 2; GATA3 – GATA-binding protein 3; ERE – estrogen response element. The (?) at the highly methylated CpGs -1472 and -1468 is an unknown near-palindromic sequence of interest.

Chromatin-immunoprecipitation (ChIP) studies would also be needed to demonstrate that CpGs within the differentially methylated region indeed interact with transcription-associated protein complexes. However, in addition to the ERE and Sp-1 sites, several known transcription factor binding consensus sequences can be found in the lamprey Pm807 putative promoter. RUNX3-like motifs (Kopf and Miskin, 2005) are present at CpG 5 (-1504; TGmCpGGA) in Segment G, as well as in Segment F (ACCACA) and at CpG 8 (-831) in Segment E (TGCGGT) (Fig. 6-11A). The RUNX genes have been studied in the Japanese lamprey, exhibiting highly conserved RUNT DNA binding subunit homologies through humans (Nah et al., 2014). A GATA3 enhancer protein binding site (TGATAA) exists beginning 1422 bp upstream of the TSS, and at the end of Segment F (TGATAT) upstream of CpG -906. A GATA3 binding site is also found in the human V1a receptor promoter (UCSC Accession: wgEncodeEH001639), and a RUNX3 binding site is found in the human OXT receptor promoter (UCSC Accession: wgEncodeEH002330). An interesting though unknown motif surrounding CpGs 7 (-1472) and 8 (-1468) of Segment G is a near-palindromic sequence which reads (5'-3', methylated) as AAGmCpGAGmCpGGAA.

Although most transcription factors in mammals are hindered from binding methylated DNA, Hu et al. (2013) showed that of the ~1300 human transcription factors, 47 could bind methylated CpGs and convey transcriptional activity. One of these was Krüppel-like factor 4 (KLF4), an evolutionarily conserved zinc-finger (ZnF) DNA-binding protein associated with stem cell regulation as well as gut and cardiovascular transcription regulation. KLF4 acts as either a repressor or activator, depending on

sequence motif and context (Hu et al., 2013). In the lamprey the CpG locations with highest methylation in the adult heart/larval liver compared to almost no methylation in the adult liver were Segment G CpGs 5, 6, 7, 8, 9, 10 and 12 (-1504, -1497, -1472, -1468, -1445, -1436, and -1405 respectively relative to the TSS). CpG 12 (-1405) matches exactly the consensus sequence that Hu et al. (2013) found to be strongly associated with KLF4 binding when methylated, as CCmCpGCC (Fig. 6-11A). The sequence at CpG 9 (-1445) is very similar but has an extra guanine (CCmCpGGCC). Other methylated motif sequences have also been identified as interacting with KLF4 through its three ZnF cysteine2-histidine2 (C2H2) domains to activate transcription (Liu et al., 2014). Similar GC-rich methylated motifs can also be bound by the closely-related Sp-1 transcription factor (Höller et al., 1988; Kaczinski et al., 2003).

In Figure 6-12 we compare the KLF4 carboxy-terminal binding domains in human, frog, and fish amino acid sequences from GenBank with predicted sequences from the sea lamprey (Ensembl) and Japanese lamprey (*Lethenteron japonicum*) genome databases. While the amino-terminal regions of KLF4 are divergent among vertebrates and even within mammals (Kaczynski et al., 2003), the ZnF-C2H2 binding domains and specific interacting residues are highly conserved in all species examined (Fig. 6-12). Therefore, a KLF4-like transcription factor is present in lampreys and likely binds methylated motifs in a manner similar to that demonstrated in mammals (Hu et al., 2013; Liu et al., 2014).

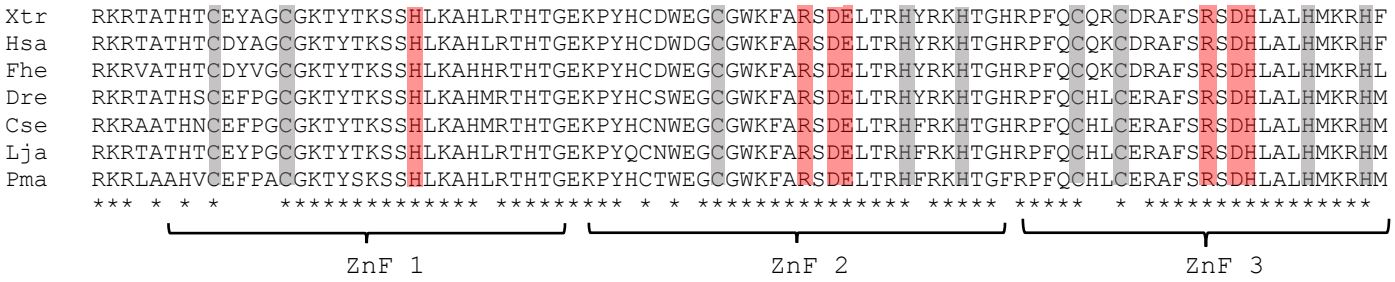


Figure 6-12. Conserved carboxy-terminal zinc-finger (ZnF) cysteine2-hystidine2 domains of Krüppel-like factor 4 transcription factor amino acid sequences. Gray highlighted C2H2 residues are involved in zinc binding, red highlighted residues (RDH/E) interact with specific DNA bases (Liu et al., 2014). Species and accession numbers: frog (Xtr - *Xenopus tropicalis*; AAI70893.1), human (Hsa – *Homo sapiens*; AAH30811.1), killifish (Fhe - *Fundulus heteroclitus*; XP_012724460.1), zebrafish (Dre – *Danio rerio*; AAI65781.1), flatfish (Cse - *Cynoglossus semilaevis*; XP_008332950.1), Japanese lamprey (Lja – *Lethenteron japonicum*; predicted from Scaffold KE994719.1), and sea lamprey (Pma – *Petromyzon marinus*; predicted from Ensembl Scaffolds GL490429.1 and GL476342.1).

Covelo-Soto et al. (2015) used methylation-sensitive amplified polymorphism (MSAP) genotyping to detect DNA methylation pattern differences between adult and larval lampreys, finding these differences occurred in genes related to metamorphosis including the vasotocin hormone precursor, lymphocyte receptor and Hox genes. But they compared only muscle tissue in adult and larval animals without demonstrating expression levels for these genes in muscle tissues, for the two life stages. We found the vasotocin hormone precursor to be expressed in muscle at all life stages (Chapter 2 and unpublished). Therefore Covelo-Soto et al. (2015) cannot correlate expression level with methylation level, only methylation level with life stage in muscle for those genes.

For future studies, we offer two related hypotheses on the function this receptor might have in the larval liver that is abolished upon metamorphosis to the parasitic phase and no longer needed by the adult. In reviewing the cellular responses that define functionality at the organ level, note that calcium-mediated cellular contraction often associated with secretion is the result of the ligand-V1a or OXTR signaling cascade. In the liver, vasopressin is known to mediate glycogenolysis and gluconeogenesis in mammals (Hems and Whitton, 1973). Therefore, one hypothesis is that the feeding larvae need vasotocin-mediated metabolism but the reproducing adults no longer feed and therefore have reduced metabolic needs. However, this does not account for the lack of the vasotocin receptor in parasitic-phase animals that feed on blood.

The fact that vasopressin regulates glycogenolysis and gluconeogenesis metabolic activities through the V1a receptor by enhancing bile acid secretion (Ghazaee et al., 2010) leads to our second hypothesis. V1a receptor gradients from the peripheral toward

the central cells of the liver create an intercellular calcium wave that facilitates bile acid secretion (Tordjmann et al. 1998). Larval lampreys feed on detritus in stream sediments for several years, utilizing these metabolic processes to grow and prepare for transformation to the parasitic phase. During this time, they also release bile acids in their feces that attract spawning adults to these streams, acting as a signal of suitable spawning habitat (Polkinghorne et al., 2001; Sorensen et al., 2003). These bile acids were present in the livers and gall bladders of larvae but not the livers of parasitic or maturing adult animals. The gall bladder and bile ducts are lost during metamorphosis (Polkinghorne et al., 2001). Therefore, we hypothesize that the vasotocin V1a/OXT receptor may mediate the release of the bile acid migratory pheromone, a liver-related function that shuts down upon metamorphosis.

In conclusion, we report a novel mechanistic landscape in which higher methylation in the DNA from the adult heart and larval liver leads to higher transcription in those tissues. This is plausible considering that the KLF4 and Sp-1 transcription factors in humans bind to methylated DNA to activate transcription. Considering the location of the CpG island near the TSS, with differential methylation in a pattern relative to known regulatory consensus sequences, the lamprey Pm807 gene appears to have a typical vertebrate promoter region. More in-depth studies on individual genes will reveal whether lamprey gene regulation follows a vertebrate-specific or species-specific DNA methylation rule, or whether neuropeptide hormone receptor genes share conserved promoter regulatory organization across phyla.

Chapter 7

General Discussion, Future Studies and Conclusions

7.1 General discussion

To restate the central hypothesis of my research from Chapter 1, ***the molecular structures of genes encoding the AVT neuropeptide hormone and its receptors in the sea lamprey are evolutionarily conserved across vertebrate phyla, and are under the control of conserved cis-regulatory elements that allow selection of pleiotropically-linked traits*.** I tested the elements of this hypothesis through sequencing; determination of receptor expression among fifteen different tissues in the central nervous system and peripheral organs, and immune system cells; functional experiments on pheromone release; and gene promoter analysis.

Aim 1 (Chapter 2) was to test the hypothesis that the lamprey possesses AVT receptor orthologs of predecessors to the AVP/OXT family of G protein-coupled receptors found in mammals, providing insights into the early branching into the mammalian V1a, V1b, V2a and OXT receptors. Based especially on molecular phylogeny analyses, the hypothesis was partially supported in that the lamprey putative AVT receptor genes did seem to form precursor branches to the larger V1/OXTR and V2 clades. The synteny analysis however, partially rejected the hypothesis indicating divergence to the point of uniqueness of the receptors beyond assigning specific orthological relationships with the jawed vertebrate types. The Pm807 and Pm2017 are paralogs of each other that can at best be said to be V1a/OXT/V1b receptor-like. The lamprey sports two V2b/c-like receptors and no V2a was found. The Pm3133 appears to be a duplicate of Pm807, and not a separate receptor type.

Tissue expression distribution for mRNA showed that all of the receptors except

Pm2017 were transcribed in a wide variety of tissues. All of the putative receptors were expressed in the brain and reproductive tissues, especially in the cloaca and penis tissues which could indicate functions in oviposition and sperm release reported for jawed vertebrate species (Goodson and Bass, 2001). The fact that the Pm807 was the most highly expressed gene and most easily amplified through PCR made it a logical choice amenable to focused study in the subsequent chapters.

Another important conclusion was that this data supports the newly-proposed likelihood that only one round (1R) of whole-genome duplication (WGD) occurred in the ancestral vertebrate lineage (Smith and Keinath, 2015). A scenario in which segmental duplication in the jawed vertebrates created the V1b and V2a genes did not seem to occur in the lamprey lineage, supporting 1R, which is consistent with patterns seen in other gene families (Mehta et al., 2013; and Nah et al., 2014). My sequencing work also resulted in one near full-length (Pm4769) and four full-length putative AVT receptor coding sequences and a partial actin sequence that have been deposited in GenBank.

Aim 2 (Chapter 3) represents the initial attempt to show that the full-length coding sequences found in Chapter 2 are translated to functional proteins by cloning for the purpose of expressing the putative receptor genes into mammalian cells. Due to the high GC content of Pm3133, Pm2017 and Pm644, I was unable to generate clones with the correct native sequences (according to the genome database and sequences determined in Chapter 2). However, I was able to clone the Pm807 gene in preparation for future expression and functional assay studies, which are continuing in our lab. This paves the way for studies of ligand affinity and the discovery of specific antagonists.

Ligand affinity and receptor antagonist studies would not only help resolve orthology but would also become useful research tools to show whether blocking essential functions could help control the invasive lamprey in the Great Lakes. Molecular exploitation is a concept that could apply to the lamprey's Pm807 and Pm2017 receptors, which show OXTR-like structures and synteny even though the lamprey appears to possess no OXT peptide-producing gene. Molecular exploitation has been demonstrated (Bridgham et al., 2006) in which the existing mineralocorticoid hormone receptor in the vertebrate ancestor may have been co-opted to interact with a new ligand through a small number of amino acid changes in the receptor over a long period of time. It can therefore be speculated that when the OXT hormone appeared in jawed vertebrates, it was able to exploit at least one of the existing AVT receptors. The divergence of the more OXT-compatible receptor from the other AVT receptors then resulted over time from the selection pressure provided by the difference in the ligands of one amino acid (leucine in OXT instead of arginine in AVT in the eighth position).

Cloning the putative Pm644 V2b/c receptor is a high future priority, considering the changes in the V2 receptor allowing utilization of cyclic AMP as a second messenger. I speculate that the switch in second messengers may be attributed to sequence changes in a specific amino acid motif within intracellular loop (ICL) 3 from the V2b and V2c receptors to the V2a receptors. The amino acid sequence QVLIFREIHASL was shown to be important for interaction with the G_S protein in the V2a receptor (Liu and Wess, 1996; Erlenbach and Wess, 1998). The greatest decline in cAMP generation occurred when the Q (glutamine) and E (glutamic acid) were substituted. I compared amino acid alignments

among various species' AVP and AVT receptors to identify the ICL3 motif differences among the V2a, V2b and V2c receptor types potentially affecting G_s protein interaction. The QVLIFREIHASL motif is generally conserved in the V2a type (Fig. 7-1). The Q is conserved in all V2b/c-type receptors and the first I (isoleucine) is conserved among all receptor types (V1, OXT and V2). Therefore, the critical change in the motif to allow interaction with the G_s protein seems to have been in the appearance of the FREIH or FKEIH sequence. In addition to mammals, activation of adenylyl cyclase has been demonstrated for V2a-type receptors in Japanese tree frog (FREIH; Kohno et al., 2003), African lungfish (FKEIH; Konno et al., 2009), and medaka (FREIH) and gray bichir fish (FKEIH; Konno et al., 2010). The duplicated V2a receptors in teleost fish can take both the FREIH and FKEIH forms. Neither chicken V2c receptors (CKIIK) (Tan et al., 2000) nor, as noted, the shark V2bR1 (CRAIQ) (Yamaguchi et al., 2012), when transfected into mammalian cells, have shown cAMP accumulation upon stimulation with AVT. Based on analysis of the ICL3 motif amino acid sequences of the lamprey putative V2b/c-like receptors Pm644 (CRIIR) and Pm4769 (CRVIR) (Fig. 7-1), I would predict no AVT-inducible interaction with the G_s protein, and therefore no adenylyl cyclase activity and no cAMP accumulation.

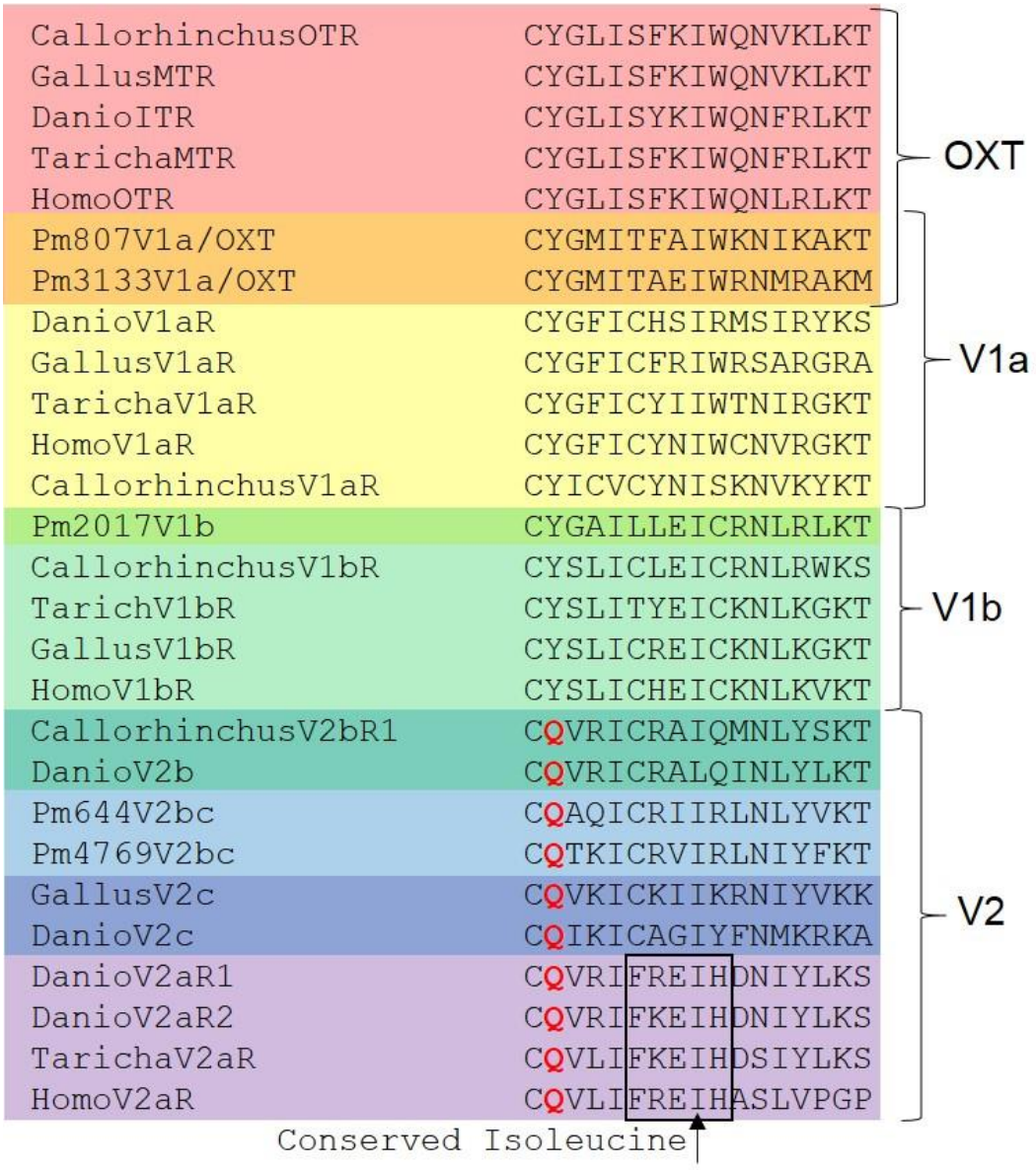


Figure 7-1. Comparison of intracellular loop (ICL) 3 amino acid motifs among receptor types and species. The boxed sequence is proposed to confer the ability to interact with the G_s protein resulting in cAMP accumulation.

Cloning and expression of genes in mammalian cell lines or *Xenopus* oocytes for functional studies is standard practice, as it seems to be assumed that the G-proteins and other components of the intracellular signaling cascade are conserved in all animals and operate in the same manner. As we have seen in Chapter 6 for the conservation of transcription factor and steroid hormone receptor binding domains, perhaps this is true. However, demonstration of receptor function *in vivo* would provide more definitive verification that the translated native protein can not only spark the accumulation of a second messenger, but can result in a measurable physiological end effect in the animal. Chapters 4 and 5 were dedicated to laying the groundwork for future studies to demonstrate physiological effects of exposure to exogenous AVT in lamprey cellular and organ systems.

Establishing a primary culture method and determining whether lamprey immune cells express neuropeptide hormones was the focus of *Aim 3* (Chapter 4). Live lamprey immune cells were successfully maintained in an *in vitro* culture medium for periods of at least six days, which has never before been demonstrated. Also, mRNA expression was preliminarily demonstrated for neuropeptide hormones GnRH and AVT, and the Pm807 putative AVT receptor. Neuropeptides and receptors are similarly expressed in jawed vertebrates, and functional roles have been demonstrated (Akbulut et al., 2001; Chen et al., 2002; Baurenhofer et al., 2003; Hansenne et al., 2004; Yada et al., 2004; Shaik Mohamed and Khan, 2006; Ndiaye et al., 2008; Loser et al., 2010; Koike and Seki, 2013; Poutahidis et al., 2013). My results support the growing body of evidence that neuroendocrine-immune interaction is an important fitness characteristic conserved

across vertebrate phyla. In humans and other mammals the CNS-like actions of these hormones in lymphocytes had long been dismissed as a novelty or minor function, and the importance of neuroendocrine-immune regulation has only more recently been recognized (Demas et al., 2011). Additional assays are needed to verify the transcription, and show translation of the proteins in lamprey immune cells. However, with the neuropeptide hormone functionality in immune cells already demonstrated in the literature not only in vertebrate but also invertebrate animals (Zhang et al., 2000; Demas et al., 2011), the neuroendocrine role in immune cell survival, regulation, and proliferation must be taken seriously as not only evolutionarily conserved, but as an evolutionary driver.

To be an evolutionary driver, a trait must be encoded by varying DNA sequences, presenting different functional options in either the coding or non-coding regions upon which natural selection can act. One or all of the traits attributed to a single gene must be under selection pressure. Immune systems are always under pressure because pathogen genomes are constantly mutating. Therefore, sequence diversity in a single pleiotropic gene that enhances immunity can ensure the survival and evolution of populations and species. Also, with not only epigenetic factors acting differently on the same promoter sequence in different tissues, but also heterozygosity offering phenotypic plasticity, traits can be fine-tuned within different tissues of individual organisms. One way in which results from the immune cell experiments demonstrates these concepts is the expression of an AVT hormone allele from the freshly isolated larval immune cells that differed from the sequence for *Petromyzon marinus* found in GenBank but matched the Japanese

lamprey AVT hormone precursor sequence. This indicates possible heterozygosity (the *P. marinus* species appears to have at least two alleles), and potentially the preferential expression of a specific AVT gene allele in immune cells. The mechanism responsible could be found in DNA methylation of the AVT gene promoter, which will be discussed below in conjunction with the results of Chapter 6.

Aim 4 (Chapter 5) describes the effort to demonstrate AVT hormone system function *in vivo* at the organ and system level. My hypothesis was that sex pheromone release from male sea lamprey gills is mediated by AVT and possibly OXT hormone exposure, and can be decreased or blocked by a mammalian V1a receptor antagonist. My hypothesis was based on studies showing that AVT mediated pheromone release from an abdominal gland in the amphibian red-bellied newt (Toyoda et al., 2003). Additionally, excretory roles are well known for AVP such as antidiuresis and insulin release (Decaux et al., 2008), and for OXT such as milk excretion and seminal fluid ejaculation (Gimpl and Fahrenholtz, 2001). A related example of an olfactory signal mediated by the AVT/AVP/OXT hormone family through the V1a receptor is scent or flank marking, used by some mammals to mark territory and attract mates (Caldwell and Young, 2006). If my hypothesis were supported, AVT receptors would provide a target for invasive sea lamprey control in the Great Lakes, and would also add lamprey pheromones to the list of AVT-mediated secondary sexual traits.

In the first of two experiments, in which gill pouches were removed and individually treated with AVT and V1a receptor antagonist OPC-21268, the results did not support the hypothesis, showing no significant treatment effects or interactions

($p>0.05$). In the second experiment, samples from some animals contained higher levels of the pheromone PADS after injection with either AVT or OXT than before injection, while in saline-injected control lampreys post-injection levels did not increase, indicating a possible effect of the hormones. Pheromone levels from AVT-injected males were higher than from OXT-injected males. However, levels of the main sex pheromone component 3kPZS was not higher after injection, and maturation timing and difficulty preventing blood from entering the samples may have affected results.

Important outcomes of the pheromone study include the novel method of collecting pheromones from individual gills which appears to confirm that similar levels of pheromone compounds are released from all gills in a particular animal. High variability from animal to animal in both experiments potentially shows that some animals may have already released a finite amount of pheromones, or were starting to experience physiological failure in the late season, or in the early season experiments were not yet mature enough to release pheromones. Utilizing a chamber (Siefkes et al., 2003) to collect samples from live, cannulated animals over time, or implanting a time-release hormone pellet (Bryan et al., 2015) may solve several experimental issues. Pheromone analysis data from this pilot study were somewhat surprising in the detected levels of the non-keto-bile acids ACA, PZS, PADS and PSDS known to be active in migratory pheromone mixtures released by larvae (Li and Sorensen, 1997; Polkinghorne et al., 2001; Hoye et al., 2007). However, higher relative levels in excised gill samples of the main sex pheromone attractant component 3kPZS were consistent with its known activity, along with other keto-compounds 3kACA and DKPES, in sex pheromone

mixtures released from adult male gills (Yun et al., 2003; Li et al., 2013).

Although the AVT effects on pheromone release were not definitive, as cited above other studies have shown that reproductive functions and secondary sexual characteristics are mediated by the AVT/AVP/OXT hormone-receptor systems. In the lamprey, tissue expression distribution of mRNA shows several systems, including reproductive tissues, in which AVT could function (Chapter 2), and my results of primary immune cell culture preliminarily showed transcription of AVT hormone and receptor genes (Chapter 4) representing potential involvement in the ability to respond to pathogenic threats. *Aim 5* (Chapter 6) was designed to tie differential mRNA expression and potential functional diversity to a stretch of DNA upstream of an AVT receptor transcription start site (TSS).

The high GC content in coding regions of lamprey genes raises the question of whether DNA methylation of cytosine-guanine (CpG) dinucleotides could function to regulate lamprey gene transcription. However, previous studies showed the lamprey genome to be globally methylated similarly to jawed vertebrates (Tweedie et al., 1997). In one recent paper DNA methylation pattern differences between adult and larval lamprey were found in muscle tissue for genes related to metamorphosis, including the vasotocin hormone precursor, lymphocyte receptor and Hox genes (Covelo-Soto et al., 2015). However, the authors did not determine levels of mRNA expression. In Chapter 6 our hypothesis was that differential DNA methylation of CpG dinucleotides in the promoter region of the lamprey Pm807 AVT receptor gene is correlated with life-stage and tissue-specific differences in mRNA expression. Genomic DNA from adult heart and larval liver where the Pm807 gene is transcribed, was compared to adult liver where no

mRNA expression has been observed. Bisulfite conversion and High Resolution Melt (HRM) PCR of the genomic DNA with subsequent sequencing showed for the first time in a specific lamprey gene that DNA methylation was in fact correlated with transcription. Methylation of CpGs occurred in a specific pattern in a CpG island area (as defined by Li and Dahiya, 2002) from approximately -900 to -1600 bp upstream of the TSS. However, because the high levels of methylation in the heart and larval liver appeared to result in high levels of transcription, contrary to the prevalent mode in mammals, it was important to ascertain how this could work.

The putative Pm807 receptor promoter region was searched and found to contain consensus sequences matching those of regulatory elements known to control gene transcription in other species. Estrogen response element (ERE) sites and half-sites are present in the Pm807 putative promoter. As noted in Chapter 1, the estrogen nuclear receptor in lampreys binds the classic estradiol molecule, and the activated receptor binds ERE consensus sequences (Paris et al., 2008). Moreover, recent research actually provides evidence of regulatory element sequences containing integrated CpGs that require methylation in order for certain transcription factors to bind and initiate transcription (Hu et al., 2013; Liu et al., 2014). One of these transcription factors is Krüppel-like factor 4 (KLF4). The KLF4 consensus DNA binding site sequence is found in the Pm807 promoter DNA; the CpGs within this sequence were highly methylated in adult heart and larval liver samples (high transcription), but unmethylated or at low methylation levels in the adult liver (no transcription). Also, we show that the KLF4 protein gene exists in both the *P. marinus* and *L. japonicum* genomes, and that the DNA

binding domain is highly conserved among jawed vertebrates including humans.

The significance of these results are that these regulatory elements integrated with epigenetic factors are the basis for gene pleiotropy. Potentially, dynamic methylation changes could be responsible for the Pm807 gene transcription in the lamprey immune cells when incubated with mitogens. The silencing of one AVT hormone precursor allele over another could also be controlled by promoter DNA methylation. And, as discussed in Chapter 6, the shutdown of Pm807 transcription in the liver upon metamorphosis is choreographed with life-stage-specific metabolic activities. The de-methylation at specific CpGs in the KLF4 binding sequence from the larval to adult DNA may occur during metamorphosis. KLF4 is also known to participate in chromatin remodeling by mediating histone protein 3 acetylation (Sprujit and Vermeulen, 2014; He et al., 2015); this would be consistent with the KLF4 binding sites at ~1400 bp upstream of the Pm807 TSS operating as an enhancer to expose the proximal promoter near the TSS for the binding of other transcription factors including the activated estrogen receptor.

The possible involvement of KLF4 is also important as a potential connecting factor between gut metamorphosis and the particular bile acid compounds that become pheromones. Bile ducts and the gall bladder are degraded upon metamorphosis so that bile compounds are no longer secreted to the digestive tract (Polkinghorne et al., 2001). In addition to bile duct loss, the gut in lampreys is repurposed toward reproductive function. KLF4 could be involved in this process as it is highly expressed in the gut epithelium of mammals and is involved in cell reprogramming (Sprujit and Vermeulen, 2014). Interestingly, conversion of the bile compounds to keto forms found in the

bloodstream and/or gills of metamorphosed adults, as shown by pheromone analysis (Chapter 5), may have been an adaptation to reduce toxicity (Cai et al., 2013). In terms of medical knowledge, the degradation of bile ducts as a human disease (biliary atresia) is currently being studied in lampreys to explore how the animal has adapted to the interruption of bile excretion through the gut (Cai et al., 2013; Chung-Davidson et al., 2015). The lamprey might also find future use as a model for the study of colorectal cancer related to DNA methylation (Curtin et al., 2011) and KLF4 (Hu et al., 2011).

7.2 Future Studies

Five putative AVT receptors and one AVT hormone precursor were found in the sea lamprey, corroborated by searches of the Japanese lamprey database. These most likely represent all of the genes in that family that exist in lampreys, however genome and chromosome construction is an ongoing process and new techniques may yield future surprises. The Pm4769 coding sequence can likely now be completed by designing primers with the aid of the sequence from the Japanese lamprey genome database. For the currently known genes, the 5'- and 3'- untranslated regions need to be confirmed so that gene regulation can be studied, including the translation-regulating effects of micro-RNAs acting on the 3'-UTRs. The TSS of the Pm2017 gene, of interest because it has the most limited tissue expression distribution, was determined in this study but the promoter region has not yet been explored. The full coding sequence of the AVT hormone precursor in *P. marinus* has not been determined, and the discovery in Chapter 4 of potential differential allele expression needs to be further investigated.

The fact that the AVT receptor genes in the lamprey still must be called "putative" needs to be remedied. The Pm807 gene was successfully cloned and is now ready for expression in a mammalian cell line. Function will be analyzed using a method such as flow cytometry which is currently being developed in our lab. The other receptors need to be successfully cloned, transfected and assayed to determine their intracellular signaling capabilities and second-messenger pathways. A fairly extensive effort not included in this dissertation was expended on immunohistochemistry in an attempt to show that the receptors are produced as proteins in the lamprey tissues. Custom antibodies were not yet

available so a mammalian antibody was purchased which was produced using an antigen peptide sequence with approximately half the amino acids conserved in the lamprey putative Pm807 protein. Specific staining was not achieved in any of the sectioned and stained tissues, including gills, heart, brains or liver. Custom antibodies were recently made for Pm807, Pm2017 and Pm644, and are currently being purified and validated by Western blot for many future applications.

Much can yet be learned about the lamprey immune cells through primary culture. My research has provided preliminary information on culture conditions. Proliferation and apoptosis assays are needed, as well as quantitative PCR to test whether cytokines, AVT receptors or other neuropeptide systems such as GnRH and melanocortin are active in the lamprey immune system. The cells could also be used to test how stimulation with pathogen-associated molecules affect DNA methylation in the promoters of the neuropeptide hormone and receptor genes. Behaviors correlated with neuropeptide receptor expression patterns in the brain have shown an association with expression in immune cells (Provencal et al., 2014). Immune cells have been collected using minimally invasive techniques to study behavior in humans (Unternaerer et al., 2012; Jack et al., 2012; Gregory et al., 2009) and paralogs within a receptor family may be differentially expressed in immune cell subsets (Provencal et al., 2014; Hansenne et al., 2004). Differential expression could be associated with dynamic methylation, and methylation could be compared between immune cells and specific brain areas at different life stages or when the animals are stressed or exposed to chemicals. The immune cells could act as a less-invasive surrogate monitoring tool for not just lampreys but other non-mammalian

wildlife and economically-important species.

As discussed in the previous section, AVT was not definitively linked to pheromone release but may still be a factor. Methods are available to potentially resolve this question such as constructing a chamber (Siefkes, et al., 2003) to collect samples from live, cannulated animals, or implanting a time-release hormone pellet (Bryan et al., 2015). An exciting line of research on AVT involvement in related reproductive behaviors could use similar methods, but these also require more elaborate resources than were available for this project, such as an experimental stream (see e.g., Johnson et al., 2009; Hume et al., 2013).

The tie that binds transcription in all of the pertinent cells and tissues and at the proper times is the gene promoter region landscape. Chromatin immunoprecipitation studies especially are needed to confirm actual DNA binding of transcription factors to putative consensus sequences.

7.3 Conclusions

The Good Genes hypothesis put forth by Hamilton and Zuk (1982) provides for the selection of enhanced reproductive and immune traits simultaneously. The Good Genes hypothesis states that secondary traits that are attractive to the opposite gender indicate genetic quality and resistance to disease, thus, choosing the attractive male will insure the female's direct reproductive success and fitness of her offspring. Prior to the work of Zuk (1996) and Siva-Jothy and Skarstein (1998), the "good genes" that link mating signals and immune function had been treated as a black box. Zuk (1996) put forth a "good-genes" candidate. Her work focused on testosterone, which is involved in the development of secondary sexual traits as well as regulating immune system function (Zuk, 1996). Siva-Jothy and Skarstein (1998) review the highly diverse major histocompatibility complex (MHC) genes that work together with T- and B-lymphocyte genes to produce specific immunoglobulin-based antibodies. However, although the testosterone receptor is a gene product, the testosterone hormone itself is not, and many upstream genes and events are involved in its production and release. MHC gene products are not involved in reproduction per se, except for the sensory processing of pheromones (Siva-Jothy and Skarstein, 1998).

Despite the existence of a plethora of data to support the role of neuropeptide hormone systems in multiple functions including reproduction and immunity, no one has formally nominated these good-genes candidates. In this dissertation I have laid additional groundwork from the perspective of a basal vertebrate to propose that neuropeptide hormones and their receptors are the central genes in a "good genes

network," in conjunction with steroid hormone systems, that spans from at least the jawless vertebrate species and links reproductive, homeostatic and immune system traits. A schematic summarizing the neuropeptide connections linking mating signals with reproduction, immune function and fitness is presented in Figure 7-2.

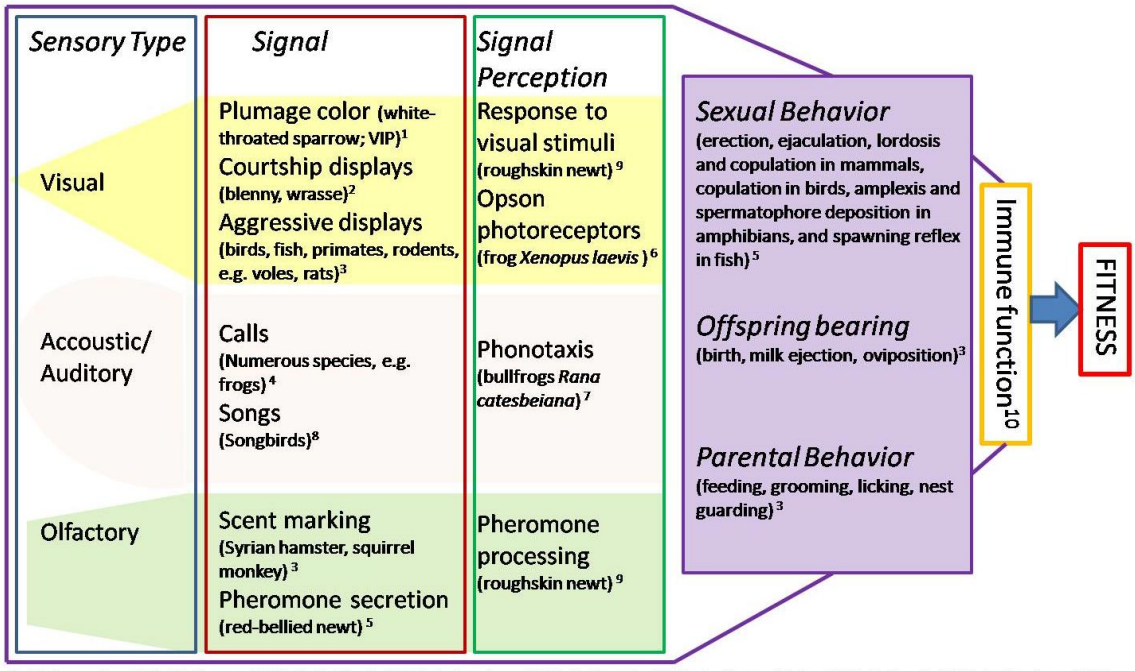


Figure 7-2. Known mating signals, perception and reproductive behaviors and functions mediated by neuropeptide hormones.

In summary, the AVP/OXT homologs and their receptors provide the conduit for activating the intracellular processes resulting in specific muscular movement, vocalization, excretion, sensory perception, or even higher brain functions like a feeling or the retrieval of a memory (Choleris et al., 2013). Additionally, gene promoter modifications on which natural selection may act, that affect the expression of these genes and associated behaviors in the central nervous system, potentially also affect their

expression and functionality in peripheral organs and the immune system. The ideas gained from this research may further our understanding of evolution, provide practical applications in the enhancement of non-mammalian immune cell primary culture, and a basis for future animal health and human medical applications concerning gene promoter regulation.

Bibliography

Acher, R., Chauvet, J., Chauvet, M.T. (1995). Man and the chimaera. Selective versus neutral oxytocin evolution. Adv. Exp. Med. Biol. 395, 615-627.

Akbulut, S., Byersdorfer, C. A., Larsen, C. P., Zimmer, S. L., Humphreys T. D., Clarke, B.L. (2001). Expression of melanocortin 5 receptor on rat lymphocytes. Biochem. Biophys. Res. Comm. 281, 1086-1092.

Albalat, R., Marti-Solans, J., Canestro, C. (2012). DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. Briefings in Functional Genomics pp1-14, doi:10.1093/bfgp/els009.

Alder, M. N., Herrin, B.R., Sadlonova, A., Stockard, C.R., Grizzle, W.E., et al. (2008) Antibody responses of variable lymphocyte receptors in the lamprey. Nature Immunol. 9, 319-327.

Alexander, C., Rietschel, E.T. (2001). Bacterial lipopolysaccharides and innate immunity. J. Endotoxin Res. 7, 167-202.

Alvarez-Viejo, M. et al. (2003). Co-localization of mesotocin and opsin immunoreactivity in the hypothalamic pre-optic nucleus of *Xenopus laevis*. Brain Research 969: 36-43.

Arcoumanis, C., McGuirk, J. J., Palma, J. M. L. M., (1990). On the use of fluorescent dyes for concentration measurements in water flows. Exper. Fluids 10, 177-180.

Baeyens, D.A., Cornett, L.E. (2006). The cloned avian neurohypophysial hormone receptors. Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 143, 12–19.

Balment, R.J., Lu, W., Weybourne, E., Warne, J.M. (2006). Arginine vasotocin a key hormone in fish physiology and behaviour: A review with insights from mammalian models. Gen. Comp. Endocrinol. 147, 9-16.

Beets, I., Janssen, T., Meelkop, E., Temmerman, L., Suetens, N., et al. (2012). Vasopressin/oxytocin-related signaling regulates gustatory associative learning in *C. elegans*. Science 338, 543-545. http://dx.doi.org/10.1126/science.1226860

Binder, T.R., McDonald, D.G. (2007). Is there a role for vision in the behaviour of sea lampreys (Petromyzon marinus) during their upstream spawning migration?" Can. J. Fish. Aquat. Sci. 64, 1403-1412.

Birnbaumer, M. (2000). Vasopressin receptors. Trends. Endocrinol. Metab. 11, 406-410.

Black, J.W. (1960). Electrocardiographic changes produced in rabbits by vasopressin (pitressin) and their alteration by prolonged treatment with a commercial heart extract. J. Pharm. Pharmacol. 12, 87-94.

Bodily, P.M., Fujimoto, M.S., Ortega, C., Okuda, N., Price, J.C., et al. (2015). Heterozygous genome assembly via binary classification of homologous sequence. BMC Bioinformatics 16 (Suppl 7):S5

Boyd, S. (1994). Arginine vasotocin facilitation of advertisement calling and call phonotaxis in bullfrogs. Horm. Behav. 28, 232-240.

Boyd, V.L., Schroeder, B.G., Lee, L.G. et al. (2007). "New and improved methods for analysis of methylated DNA using capillary electrophoresis." In: *Progress in DNA Methylation Research*, ed. H.P. Neumann (Nova Science Publishers, Inc.: New York). ISBN: 9781600217227

Breathnach, R., Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. 50, 349-383.

Bridgham, J.B., Carroll, S.M., Thornton, J.W. (2006). Evolution of hormone-receptor complexity by Molecular Exploitation. Science 312, 87-101.

Bryan, M.B., Scott, A.P., Li, W. (2008). Sex steroids and their receptors in lampreys. Steroids 73, 1-12.

Bryan, M.B., Chung-Davidson, Y.-W., Ren, J., Bowman, S., Scott, A.P., et al. (2015). Evidence that progestins play an important role in spermiation and pheromone production in male sea lamprey *(Petromyzon marinus*). Gen. Comp. Endocrinol. 212, 17-27.

Burbach, J.P.H. (2010). Neuropeptides from concept to online database www.neuropeptides.nl. Euro. J. Pharmacol. 626, 27–48.

Caldwell, H.K., Young III, W.S. (2006). "Oxytocin and Vasopressin: Genetics and Behavioral Implications." In: *Handbook of Neurochemistry and Molecular Neurobiology, 3rd Edition: Volume 9: Neuroactive Proteins and Peptides*, by Abel Lajtha (ed), 573-607. New York: Springer.

Cai, S.-Y., Lionarons, D.A., Hagey, L., Soroka, C.J., Mennone, A., Boyer, J.L. (2013). Adult sea lamprey tolerates biliary atresia by altering bile salt composition and renal excretion, Hepatology 57, 2418–2426.

Carroll, S.B. (2005a). *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom.* W.W. Norton and Sons, New York.

Carroll, S.B. (2005b). Evolution at Two Levels: On Genes and Form. PLoS Biol 3, e245: 1159-1166.

Chen, A., Ganor, Y., Rahimipour, S., Ben-Aroya, N., Koch, Y., Levite, M. (2002). The neuropeptides GnRH-II and GnRH-I are produced by human T cells and trigger laminin receptor gene expression, adhesion, chemotaxis and homing to specific organs. Nature Medicine 8, 1421-1426.

Choleris, E., Pfaff, D.W., Kavaliers, M., eds. (2013). *Oxytocin, Vasopressin, and Related Peptides in the Regulation of Behavior*. (Cambridge University Press) ISBN: 9780521190350.

Chung-Davidson, Y.-W., Bryan, M.B., Teeter, J., Bedore, C.N., Li, W. (2008). Neuroendocrine and behavioral responses to weak electric fields in adult sea lampreys (*Petromyzon marinus*). Horm. Behav. 54, 34–40.

Chung-Davidson, Y.-W., Priess, M.C., Yeh, C.-Y., Brant, C.O., Johnson, N.S., et al. (2013) A thermogenic secondary sexual character in male sea lamprey. J Exp Biol 216, 2702-2712.

Chung-Davidson, Y.-W., Yeh, C.-Y., Li, W. (2015). The sea lamprey as an etiological model for biliary atresia. BioMed Research International 2015, Article ID 832943, 1-17. http://dx.doi.org/10.1155/2015/832943

Close, D.A., Yun, S.-S., McCormick, S.D., Wildbill, A.J., Li, W. (2010) 11-Deoxycortisol is a corticosteroid hormone in the lamprey. Proc. Nat. Acad. Sci. 107, 13942-13947.

Conklin, D., Chavas, A., Duff, D., Weaver, L., Zhang, Y., Olson, K.R., (1997). Cardiovascular effects of arginine vasotocin in the rainbow trout *Oncorhynchus mykiss*. J. Exp. Biol. 200, 2821–2832.

Cornett, L.E., Kirby, J.D., Vizcarra, J.A., Ellison, J.C., Thrash, J., et al. (2003). Molecular cloning and functional characterization of a vasotocin receptor subtype expressed in the pituitary gland of the domestic chicken (*Gallus domesticus*): avian homolog of the mammalian V1b-vasopressin receptor. Regul. Pept. 110, 231–239.

Covelo-Soto, L., Saura, M., Morán, P. (2015). Does DNA methylation regulate metamorphosis? The case of the sea lamprey (*Petromyzon marinus*) as an example. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 185, 42-6. doi: 10.1016/j.cbpb.2015.03.007.

Curtin, K., Slattery, M.L., Samowitz, W.S. (2011). CpG island methylation in colorectal cancer: past, present and future. Patholog Res Int 2011, Article ID 902674, 1-8.

Decaux, G., Soupart, A., Vassart, G., (2008). Non-peptide arginine-vasopressin antagonists: the vaptans. Lancet 371, 1624–32. doi: 10.1016/S0140-6736(08)60695-9.

DeKoning, J., Kaattari, S. (1992). An improved salmonid lymphocyte culture medium incorporating plasma for in vitro antibody production and mitogenesis. Fish Shellfish Immunol. 2, 275-285.

Demas, G. E., Adamo, S., French S.S. (2011). Neuroendocrine-immune crosstalk in vertebrates and invertebrates: implications for host defense. Func. Ecol. 25, 29-39.

Dhakar, M.B., Stevenson, E.L., Caldwell, H.K. (2013). "Oxytocin, vasopressin, and their interplay with gonadal steroids," in Oxytocin, Vasopressin, and Related Peptides in the Regulation of Behavior, ed. E. Choleris, D.W. Pfaff and M. Kavaliers (Cambridge University Press) ISBN: 9780521190350.

Donaldson, Z.R., Bai, Y., Kondrashov, F.A., Stoinski, T.L., Hammock, E.A.D., Young, L.J. (2008). Evolution of a behavior-linked microsatellite-containing element of the 5' flanking region of the primate AVPR1A gene. BMC Evol Biol 8, 180–188.

Donaldson, Z. R. and Young, L. J. (2008). Oxytocin, Vasopressin, and the Neurogenetics of Sociality. Science 322, 900-904.

Donaldson, Z.R., Young, L.J. (2013). The relative contribution of proximal 5' flanking sequence and microsatellite variation on brain vasopressin 1a receptor (Avpr1a) gene expression and behavior. PLoS Genet 9(8), e1003729. doi:10.1371/journal.pgen.1003729

Dwight Z., Palais, R., Wittwer, C.T. (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics 27, 1019-1020. doi: 10.1093/bioinformatics/btr065.

Erlenbach. I., Wess, J., (1998). Molecular basis of V2 vasopressin receptor/Gs coupling selectivity. J. Biol. Chem. 273, 26549-58.

Ellens, E.R., Kittilson J.D., Hall J.A., Sower S.A., Sheridan M.A. (2013). Evolutionary origin and divergence of the growth hormone receptor family: Insight from studies on sea lamprey. Gen. Comp. Endocrinol. 192, 222-236.

Fleming, J.G., Spencer, T.E., Safe, S.H., Bazer, F.W. (2006). Estrogen regulates transcription of the ovine oxytocin receptor gene through GC-rich SP1 promoter elements. Endocrinol. 147, 899–911. doi: http://dx.doi.org/10.1210/en.2005-1120.

Fondon III, J. W. et al. (2008). Simple sequence repeats: genetic modulators of brain function and behavior. Trends Neurosci. 31, 328-334.

Ghazaee, S.P., Gorenko, Z.A., Karbovska, L.S., Veselsky, S.P., Yanchuk, P.I., Makarchuk M.Y. (2010). Desmopressin stimulates bile secretion in anesthetized rats. Gen. Physiol. Biophys. 29, 151–159. doi:10.4149/gpb_2010_02_151.

Gimpl, G., Fahrenholz, F. (2001). The oxytocin receptor system: structure, function, and regulation. Physiol. Rev. 81, 629-683.

Goodson, J. L. (2005). The vertebrate social behavior network: evolutionary themes and variations. Horm. Behav. 48, 11-22.

Goodson, J.L., Bass, A.H. (2001). Social behavior functions and related anatomical characteristics of vasotocin /vasopressin systems in vertebrates. Brain Res. Rev. 35, 246-265.

Goodson, J.L., Kabelik, D. (2009). Dynamic limbic networks and social diversity in vertebrates: from neural context to neuromodulatory patterning. Front. Neuroendocrinol. 30, 429-441.

Gregory, S.G., Connelly, J.J., Towers, A. J., Johnson, J.,Biscocho, D., Markunas, C.A., et al. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Medicine 7:62. doi:10.1186/1741-7015-7-62.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59, 307-321.

Guo, P., Hirano M., Herrin, B.R., Li, J., Yu, C., et al. (2009). Dual nature of the adaptive immune system in lampreys. Nature 459, 796-801.

Gruber, C.J., Gruber, D.M., Gruber, I.M.L., Wieser, F., Huber, J.C. (2004). Anatomy of the estrogen response element. Trends Endocrinol. Metab. 15, 73-78. doi:10.1016/j.tem.2004.01.008.

Gwee, P.C., Tay, B.H., Brenner, S., Venkatesh, B. (2009). Characterization of the neurohypophysial hormone gene loci in elephant shark and the Japanese lamprey: origin of the vertebrate neurohypophysial hormone genes. BMC Evol. Biol. 9, 47.

Haitina, T., Klovins, J., Takahashi, A., Löwgren, M., Ringholm, A., Enberg, J., et al. (2007). Functional characterization of two melanocortin (MC) receptors in lamprey showing orthology to the MC1 and MC4 receptor subtypes. BMC Evol. Biol. 7, 101.

Hall J.A., Decatur W.A., Daukss, D.M., Hayes, M.K., Marquis, T.J., et al. (2013). Expression of three GnRH receptors in specific tissues in male and female sea lampreys *Petromyzon marinus* at three distinct life stages. Front. Neurosci. 7 (88) 1-9.

Hamilton, W. D., Zuk, M. (1982). Heritable true fitness and bright birds: a role for parasites? Science 218, 384-387.

Hansen, L.H., King, Jr., E.L., Howell, J.H., Smith, A.J. (1974). A culture method for sea lamprey larvae. The Progressive Fish-Culturist 36, 122-128.

Hansenne, I., Rasier, G., Charlet-Renard, C., Defresne, M.-P., Greimers, R., et al. (2004). Neurohypophysial receptor gene expression by thymic T cell subsets and thymic T cell lymphoma cell lines. Clin. Devel. Immunol. 11, 45–51.

Harony-Nicolas,H., Mamrut, S., Brodsky, L., Shahar-Gold, H., Barki-Harrington, L., Wagner, S. (2014). Brain region-specific methylation in the promoter of the murine oxytocin receptor gene is involved in its expression regulation. Psychoneuroendo. 29, 121-131. http://dx.doi.org/10.1016/j.psyneuen.2013.10.004

Hasunuma, I., Toyoda, Kadono, Y., Yamamoto, K., F., Namiki, H., Kikuyama, S. (2010). Localization of three types of arginine vasotocin receptors in the brain and pituitary of the newt *Cynops pyrrhogaster*. Cell Tissue Res. 342, 437–457.

Hasunuma, I., Sakai, T., Nakada, T., Toyoda, F., Namiki, H., Kikuyama, S. (2007). Molecular cloning of three types of arginine vasotocin receptor in the newt, *Cynops pyrrhogaster*. Gen. Comp. Endocrinol. 151, 252–258.

He, M., Zheng, B., Zhang, Y., Zhang, X.-H., Wang, C., et al., (2015). KLF4 mediates the link between TGFβ1-induced gene transcription and H3 acetylation in vascular smooth muscle cells. FASEB J. 29, 1-12.

Hems, D.A., and Whitton, P.D. (1973). Stimulation by vasopressin of glycogen breakdown and gluconeogenesis in the perfused rat liver. Biochem. J. 136, 705-709.

Hendrich, B., Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Gen. 19, 269-277. doi:10.1016/S0168-9525(03)00080-5.

Henricson, A., Forslund, K., Sonnhammer, E.L.L. (2010). Orthology confers intron position conservation. BMC Genomics 11, 412.

Heusingveld, H.J., Westerink, R.H.S. (2011). Caveats and limitations of plate reader-based high-throughput kinetic measurements of intracellular calcium levels. Toxicol. Appl. Pharmacol. 255, 1–8

Höller M., Westin, G., Jiricny, J., Schaffner, W. (1988). Sp-l transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev. 2, 1127-1135. doi: 10.1101/gad.2.9.1127.

Hu, S.-B., Zhao, Z.-S., Yhap, C., Grinberg, A., Huang, S.-P., et al. (2003). Vasopressin receptor 1a-mediated negative regulation of B cell receptor signaling. Journal of Neuroimmunology 135, 72-81.

Hu, R., Zuo, Y., Zuo, L., Liu, C., Zhang, S., et al. 2011. KLF4 expression correlates with the degree of differentiation in colorectal cancer. Gut and Liver, 5, 154–159. http://doi.org/10.5009/gnl.2011.5.2.154.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife, 2, e00726. doi: 10.7554/eLife.00726.

Hume, J.B., Adams, C.E., Mable, B., Bean, C.W. (2013). Sneak male mating tactics between lampreys (Petromyzontiformes) exhibiting alternative life-history strategies. J. Fish. Biol. 82, 1093-100. doi: 10.1111/jfb.12047.

Jack, A., Connelly, J.J., Morris, J.P. (2012). DNA methylation of the oxytocin receptor gene predicts neural response to ambiguous social stimuli. Front. Hum. Neurosci. 6:280. doi:10.3389/fnhum.2012.00280.

Janeway, C. (2008). *Janeway's Immunobiology – Seventh Edition.* K. Murphy, P. Travers, and M. Walport, eds. Taylor and Francis, New York.

Johnson, N.S., Yun, S.-S., Buchinger, T.J., Li, W. 2012. Multiple functions of a multi-component mating pheromone in sea lamprey *Petromyzon marinus*. J. Fish Biol. 80, 538-554.

Johnson, N.S., Yun, S.-S., Thompson, H.T., Brant, C.O., Li, W. (2009). A synthesized pheromone induces upstream movement in female sea lamprey and summons them into traps. Proc. Nat. Acad. Sci. 106, 1021-1026.

Jones, M. L. (2007). Toward improved assessment of sea lamprey population dynamics in support of cost-effective sea lamprey management. J.Great Lakes Res. 33:35-47.

Joseph, N.T., Aquilina-Beck, A., MacDonald, C., Decatur, W.A., Hall, J. A., Kavanaugh, S. I., et al. (2012). Molecular cloning and pharmacological characterization of two novel GnRH receptors in the lamprey (*Petromyzon marinus*). Endocrinology 153, 3345–3356.doi:10.1210/en.2012-1217

Kaczynski, J., Cook, T., Urrutia, R. (2003). Sp1- and Krüppel-like transcription factors. Genome Biol. 4:206, 1-8. doi:10.1186/gb-2003-4-2-206.

Kanda, A., Satake, H., Kawada, T., Minakata, H. (2005). Novel evolutionary lineages of the invertebrate oxytocin/vasopressin superfamily peptides and their receptors in the common octopus (*Octopus vulgaris*). Biochem. J. 387, 85–91.

Kavanaugh, S.I., Nozaki, M., Sower, S.A. (2008). Origins of gonadotropin-releasing hormone (GnRH) in vertebrates: identification of a novel GnRH in a basal vertebrate, the sea lamprey. Endocrinology 149, 3860–3869.doi:10.1210/en.2008-0184.

Knapp, R. (2004). Endocrine mediation of vertebrate male alternative reproductive tactics: The next generation of studies. Integrated Comp. Biol. 43: 658-668.

Kohno, S., Kamishima, Y., Iguchi, T. (2003). Molecular cloning of an anuran V(2) type [Arg(8)] vasotocin receptor and mesotocin receptor: functional characterization and tissue expression in the Japanese tree frog (*Hyla japonica*). Gen. Comp. Endocrinol. 132, 485-98.

Koike, Y., Seki, S. (2013). Children with growth hormone (GH) deficiency treated with recombinant human GH have an increased frequency of CD57+ T and CD56+ NK cells and an augmented IFN-γ production in lymphocytes. HOAJ Biology 2, 1-6. http://dx.doi.org/10.7243/2050-0874-2-1

Konno, N., Kurosawa, M., Kaiya, H., Miyazato, M., Matsuda, K., Uchiyama, M., (2010). Molecular cloning and characterization of V2-type receptor in two ray-finned fish, gray bichir, *Polypterus senegalus* and medaka, *Oryzias latipes*. Peptides 31, 1273–1279.

Konno, N., Hyodo, S., Yamaguchi, Y., Kaiya, H., Miyazato, M., et al. (2009). African lungfish, *Protopterus annectens*, possess an arginine vasotocin receptor homologous to the tetrapod V2-type receptor. J. Exp. Biol. 212, 2183-2193.

Kopf, E., Miskin, R. (2005). A RUNX/AML-binding motif residing in a novel 13-bp DNA palindrome may determine the expression of the proximal promoter of the human uPA gene. J. Thromb. Haemost. 3, 2057–64. doi: 10.1111/j.1538-7836.2005.01510.x.

Krishna Pant, P.V., Tao, H., Beilharz, E.J., Ballinger, D.G., Cox, D.R., Frazer, K.A. (2006). Analysis of allelic differential expression in human white blood cells. Genome Res. 16, 331-339.

Kuenzel WJ, Kang SW, Jurkevich A. (2013). Neuroendocrine regulation of stress in birds with an emphasis on vasotocin receptors (VTRs). Gen Comp Endocrinol. 190, 18-23. doi: 10.1016/j.ygcen.2013.02.029.

Kuraku, S. (2013). Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. Semin. Cell Dev. Biol. 24, 119-127.

Kuraku, S. (2010). Paleophylogenomics of the vertebrate ancestor—Impact of hidden paralogy on hagfish and lamprey gene phylogeny. Integr. Comp. Biol. 50, 124-129.

Kuraku, S., Meyer, A., Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? Mol. Biol. Evol. 26, 47-59.

Kusui C., Kimura T., Ogita K., Nakamura H., Matsumura Y., et al. (2001). DNA methylation of the human oxytocin receptor gene promoter regulates tissues-pecific gene suppression. Biochem. Biophys. Res. Com. 289, 681–686. doi:10.1006/bbrc.2001.6024.

Lagman, D., Ocampo Daza, D., Widmark, J., Abalo, X.M., Sundström, G., Larhammar, D. (2013). The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. BMC Evol. Biol. 13, 238.

Lema, S.C. (2010). Identification of multiple vasotocin receptor cDNAs in teleost fish: Sequences, phylogenetic analysis, sites of expression, and regulation in the hypothalamus and gill in response to hyperosmotic challenge. Mol. Cell. Endocrinol. 321, 215–230.

Li, L.C., Dahiya R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics 18, 1427-31. doi: 10.1093/bioinformatics/18.11.1427.

Li, W., Sorensen, P.W. (1997). Highly independent olfactory receptor sites for naturally occurring bile acids in the sea lamprey, *Petromyzon marinus*. J. Comp. Physiol. A 180, 429-438.

Li, W., Scott, A.P., Siefkes, M.J., Yan, H., Liu, Q., Yun S.-S., Gage, D.A. (2002). Bile acid secreted by male sea lamprey that acts as a sex pheromone. Science 296, 138-141.

Li, W. (2005) Potential Multiple Functions of a Male Sea Lamprey Pheromone. Chem. Senses 30 suppl.1, i307-i308.

Li, K., Brant, C.O., Siefkes, M.J., Kruckman, H.G., Li, W. (2013) Characterization of a novel bile alcohol sulfate released by sexually mature male sea lamprey (*Petromyzon marinus*). PLoS ONE 8(7): e68157. doi:10.1371/journal.pone.0068157.

Liu J, Wess J. (1996). Different single receptor domains determine the distinct G protein coupling profiles of members of the vasopressin receptor family. J. Biol. Chem. 271, 8772-8.

Liu, Y., Olanrewaju, Y.O., Zheng, Y., Hashimoto, H., Blumenthal, R.M., et al. (2014). Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res. 42, 4859-67. doi: 10.1093/nar/gku134.

Loser, K., Brzoska, T., Oji, V., Auriemma, M., Voskort, M., et al. (2010). The neuropeptide alpha-melanocyte-stimulating hormone is critically involved in thedevelopment of cytotoxic CD8+ T cells in mice and humans. PLoS ONE 5(2): e8958. doi:10.1371/journal.pone.0008958.

Macciò A1, Madeddu C, Chessa P, Panzone F, Lissoni P, Mantovani G. (2010). Oxytocin both increases proliferative response of peripheral blood lymphomonocytes to phytohemagglutinin and reverses immunosuppressive estrogen activity. In Vivo. 24, 157-63.

Mamrut, S., Harony, H., Sood, R., Shahar-Gold, H., Gainer, H., et al. (2013). DNA methylation of specific CpG sites in the promoter region regulates the transcription of the mouse oxytocin receptor. PLoS One 8, e56869. doi:10.1371/journal.pone.0056869.

Mehta, T.K., Ravi, V., Yamasaki, S., Lee, A.P., Lian, M.M., Tay, B.-H., et al. (2013). Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). Proc. Natl. Acad. Sci. U. S. A. 110, 16044-16049.

Mellenius, H., Ehrenberg, M. (2015). DNA Template Dependent Accuracy Variation of Nucleotide Selection in Transcription. *PLoS ONE*, *10*(3), e0119588. http://doi.org/10.1371/journal.pone.0119588

Mellstedt H. (1975) In vitro activation of human T and B lymphocytes by pokeweed mitogen. Clin. Exp. Immunol. 19, 75-82.

Michopoulos, V., Maney, D.L., Morehouse, C.B., Thomas, J.W. (2007). A genotyping assay to determine plumage morph in the white-throated sparrow (*Zonitrichia albicollis*). The Auk 124: 1330-1335.

Miller, N.W., Clem, L.W. (1988). A culture system for mitogen-induced proliferation of channel catfish (*Ictalurus punctatus*) peripheral blood lymphocytes. J. Tissue Cult. Meth. 11, 69-73.

Nah, G.S.S., Tay, B-H., Brenner, S., Osato, M., Venkatesh, B. (2014). Characterization of the Runx gene family in a jawless vertebrate, the Japanese lamprey (*Lethenteron japonicum*). PLoS ONE 9(11): e113445. doi:10.1371/journal.pone.0113445.

Nair, H.P., Young, L.J., (2006). Vasopressin and pair-bond formation: genes to brain to behavior. Physiology (Bethesda). 21, 146-52.

Ndiaye, K., Poole, D.H., Pate, J.L. (2008). Expression and regulation of functional oxytocin receptors in bovine T lymphocytes. Biol. Reproduction 78, 786–793.

Ocampo Daza, D., Lewicka, M., Larhammar, D. (2012). The oxytocin/ vasopressin receptor family has at least five members in the gnathostome lineage, including two distinct V2 subtypes, Gen. Comp. Endocrinol. 175, 135–143.

Ohno, S. (1970). Evolution by Gene Duplication (Springer, New York).

Pancer, Z., Amemiya, C.T., Ehrhardt, G.R., Ceitlin, J., Gartland, G.L., Cooper, M.D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature. 430, 174-80.

Paris, M., Pettersson, K., Schubert, M., Bertrand, S., Pongratz, I., et al. (2008). An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol Biol 8:219, 1-20.

Polkinghorne, C.A., Olson, J.M., Gallaher, D.G., Sorensen, P.W. (2001). Larval sea lamprey release two unique bile acids to the water at a rate which is sufficient to produce a detectable pheromonal plume. Fish Physiol. Biochem. 24: 15–30.

Poutahidis, T., Kearney, S.M., Levkovich, T., Qi, P., et al. (2013). Microbial symbionts accelerate wound healing via the neuropeptide hormone oxytocin. PLoS ONE 8(10): e78898. doi:10.1371/journal.pone.0078898

Provençal, N., Suderman, M.J., Guillemin, C., Vitaro, F., Côté, SM, et al. (2014). Association of childhood chronic physical aggression with a dna methylation signature in adult human T cells. PLoS ONE 9(4): e89839. doi:10.1371/journal.pone.0089839

Prud'homme, B., Gompel, N., Carroll, S.B. (2007). Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. 104, 8605-8612.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1072.

Qiu, H., Hildebrand, F., Kuraku, S., Meyer, A. (2011). Unresolved orthology peculiar coding sequence properties of lamprey genes: the KCNA gene family as a test case. BMC Genomics 12, 325.

Ramallo, M. R., Grober, M., Cánepa, M. M., Morandini, L., Pandolfi, M. (2012). Arginine-vasotocin expression and participation in reproduction and social behavior in males of the cichlid fish *Cichlasoma dimerus*. Gen. Comp. Endocrinol. 179, 221-231.

Rantala, M., Kortet, R. (2004). Male dominance and immunocompetence in a field cricket. Behavioral Ecology 15, 187-191.

Reversi, A., Cassoni, P. Chini, B. (2006). Oxytocin receptor signaling in myoepithelial and cancer cells. J. Mamm. Gland Biol. Neoplasia 10, 221-229. DOI: 10.1007/s10911-005-9583-7

Rosenfeld, C.S. (2015). Bisphenol A and phthalate endocrine disruption of parental and social behaviors. Front. Neurosci. 9:57, 1-15.

Searcy, B.T., Walthers, E.A., Heppner, B.L., Thompson R.R., Moore, F. L. (2009). Identification of mesotocin and vasotocin nucleotide sequences in two species of urodele amphibian. Gen. Comp. Endocrinol. 160, 111–116.

Shaik Mohamed, J.S., Khan, I.A. (2006). Molecular cloning and differential expression of three GnRH mRNAs in discrete brain areas and lymphocytes in red drum. J Endocrinol. 188, 407-16.

Siefkes, M.J., Scott, A.P., Zielinski, B., Yun, S.-S., Li. W. (2003). Male sea lampreys, *Petromyzon marinus* L., excrete a sex pheromone from gill epithelia. Biology of Reproduction 69, 125-132.

Silver, M.R., Nucci, N.V., Root, A.R., Reed, K.L., Sower,S.A. (2005). Cloning and characterization of a functional type II gonadotropin-releasing hormone receptor with a lengthy carboxy-terminal tail from an ancestral vertebrate, the sea lamprey. Endocrinology 146, 3351–3361. doi:10.1210/en.2005-0305

Siva-Jothy, M.T., Skarstein, F. (1998). Towards a functional understanding of ``good genes". Ecol. Let. 1:178-185.

Sladek, C.D., Somponpun, S.J. (2008). Estrogen Receptors: Their Roles in Regulation of Vasopressin Release for Maintenance of Fluid and Electrolyte Homeostasis. Front. Neuroendocrinol. 29, 114–127.

Smith, E., Jones, M.E., Drew, P.A. (2009). Quantitation of DNA methylation by melt curve analysis. BMC Cancer 9:123 doi:10.1186/1471-2407-9-123.

Smith, J.J., Baker, C., Eichler, E.E., Amemiya, C.T., (2012). Genetic consequences of programmed genome rearrangement. Curr Biol 22, 1524–1529.

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat. Genet. 45, 415-21. doi: 10.1038/ng.2568.

Smith, J.J., Keinath, M.C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 25, 1-10.

Sorensen, P.W., Vrieze, L.A., Fine, J.M., (2003). A multi-component migratory pheromone in the sea lamprey. Fish Physiol. Biochem. 28, 253–257. doi: 10.1023/B:FISH.0000030545.39156.2b.

Sower, S.A., Freamat, M., Kavanaugh, S.I. (2009). The origins of the vertebrate hypothalamic–pituitary–gonadal (HPG) and hypothalamic–pituitary–thyroid (HPT) endocrine systems: New insights from lampreys. Gen. Comp. Endocrinol. 161:20–29.

Sower, S.A., Baron, M.P. (2011). The interrelationship of estrogen receptor and GnRH in a basal vertebrate, the sea lamprey. Front. Endocrinol. 2(58), 1-8.

Sower, S.A., Decatur, W.A., Joseph, N.T., Freamat, M. (2012). Evolution of vertebrate GnRH receptors from the perspective of a basal vertebrate. Front. Endocrinol. 3, 140.

Spruijt, C.G., Vermeulen, M. (2014). DNA methylation: old dog, new tricks? Nat. Struct. Mol. Biol. 21, 949-954. doi: 10.1038/nsmb.2910.

Stafflinger, E., Hansen, K.H., Hauser, F., Schneider, M., Cazzamali, G. et al. (2008). Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc. Nat. Acad. Sci. 105, 3262-3267.

Strober, W. (1997). “Trypan blue exclusion test of cell viability.” In: *Current Protocols in Immunology* A.3B.1-A.3B.2. John Wiley & Sons, Hoboken, NJ.

Tan, F.L., Lolait, S.J., Brownstein, M.J., Saito, N., MacLeod, V., et al. (2000). Molecular cloning and functional characterization of a vasotocin receptor subtype that is expressed in the shell gland and brain of the domestic chicken. Biol. Reprod. 62, 8–15.

Tanriverdi, F., Gonzalez-Martinez, D., Hu, Y., Kelestimur, F., Bouloux P.M.G. (2004). Gonadotropin-releasing hormone type-I (GnRH-I) expression in peripheral lymphocytes and possible immune action. Ericyes Med. J. 26, 105-112.

Thibonnier, M., Preston, J.A., Dulin, N., Wilkins, P.L., Berti-Mattera, L.N., Mattera, R. (1997). The human V3 pituitary vasopressin receptor: Ligand binding profile and density-dependent signaling pathways. 138, 4109-4122.

Thompson, R.R., Dickson, P.S., Rose, J.D., Dakin, K.A., Civello, G.M., et al. (2008). Pheromones enhance somatosensory processing in newt brains through a vasotocin-dependent mechanism. Proc. Royal Soc. B 275, 1685-1693.

Tordjmann, T., Berthon, B., Jacquemin, E., Clair, C., Stelly, N., et al. (1998). Receptor-oriented intercellular calcium waves evoked by vasopressin in rat hepatocytes. EMBO J. 17, 4695–4703. doi: 10.1093/emboj/17.16.4695

Toyoda, F., Yamamoto, K., Ito, Y., Tanaka, S., Yamashita, M., Kikuyamar S. (2003). Involvement of arginine vasotocin in reproductive events in the male newt *Cynops pyrrhogaster*. Horm. Behav 44, 346-353.

Tweedie, S., Charlton, J., Clark, V., Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol. Cell. Biol. 17, 1469–1475.

Uchiyama, M., Murakami, T. (1994). Effects of AVT and vascular antagonists on kidney function and smooth muscle contraction in the river lamprey, *Lampetra japonica*. Comp. Biochem. Physiol. 107A, 493-499.

University of Oklahoma. (2013). "Sea lamprey genome mapped." ScienceDaily. ScienceDaily, 28 February 2013, <www.sciencedaily.com/releases/2013/02/130228171510.htm>.

Unternaehrer, E., Luers, P., Mill, J., Dempster, E., Meyer, A.H., et al. (2012). Dynamic changes in DNA methylation of stress-associated genes (OXTR, BDNF) after acute psychosocial stress. Transl. Psychiatry 2, e150.

Vaccari, C., Lolait, S.J., Ostrowski, N.L. (1998). Comparative distribution of vasopressin V1b and oxytocin receptor messenger ribonucleic acids in brain. Endocrinol. 139, 5015-5033.

Varriale, A., Bernardi, G. (2010). Distribution of DNA methylation, CpGs, and CpG islands in human isochores. Genomics 95, 25-28. doi:10.1016/j.ygeno.2009.09.006.

Vogel, C., Marcotte, E.M. (2012). Insights into the regulation of protein abundance and proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227-232.

Walaszczyk, E. J., Johnson, N. S., Steibel, J. P., Li, W. (2013). Effects of sex pheromones and sexual maturation on locomotor activity in female sea lamprey Petromyzon marinus. J. Biol. Rhythms 28, 218-226.

Warne, J.M., Hyodo, S., Harding, K., Balment, R.J., (2000). Cloning of pro-vasotocin and pro-isotocin cDNAs from the flounder *Platichthys flesus*; levels of hypothalamic mRNA following acute osmotic challenge." Gen. Comp. Endocrinol. 119, 77–84.

Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 25, 4422-4226.

Wolstenholme, J.T., Edwards, M., Shetty, S.R., Gatewood, J.D., Taylor, J.A., et al. (2012). Gestational exposure to bisphenol a produces transgenerational changes in behaviors and gene expression. Endocrinology 153, 3828-3838. doi: 10.1210/en.2012-1195.

Xi, X., Johnson, N.S., Brant, C.O., Yun, S.-S., Chambers, K.L., Jones, A.D., Li, W. (2011). Quantification of a male sea lamprey pheromone in tributaries of Laurentian Great Lakes by liquid chromatography-tandem mass spectrometry. Environ. Sci. Tech. 45, 6437-6443.

Yamaguchi, Y., Kaiya, H., Konno, N., Iwata, E., Miyazato, M., et al. (2012). The fifth neurohypophysial hormone receptor is structurally related to the V2-type receptor but functionally similar to V1-type receptors. Gen. Comp. Endocrinol. 178, 519-528.

Yachnin S, Svenson RH. (1972). The immunological and physicochemical properties of mitogenic proteins derived from *Phaseolus vulgaris*. Immunology 22, 871-83.

Yada, T., Misumi, I., Muto, K., Azuma, T., Schreck, C.B. (2004). Effects of prolactin and growth hormone on proliferation and survival of cultured trout leucocytes. Gen. Comp. Endocrinol. 136, 298-306.

Yun, S.S., Scott, A.P., Li, W. (2003) Pheromones of the male sea lamprey, Petromyzon marinus L.: structural studies on a new compound, 3-keto allocholic acid, and 3-keto petromyzonol sulfate. Steroids 68, 297-304.

Zhang, L., Wayne, N.L., Sherwood, N.M., Postigo, H.R., Tsai, P.-S. (2000). Biological and immunological characterization of multiple GnRH in an opisthobranch mollusk, *Aplysia californica*. Gen.Comp. Endocrinol. 118, 77–89.

Zuk, M. (1996). Disease, endocrine-immune interactions and sexual selection. Ecology 77, 1037-1042.

Appendix A - Method for culture of sea lamprey larvae in a recirculating system

The sea lamprey (*Petromyzon marinus*) is an ancient vertebrate that is a destructive invasive species in the North American Great Lakes system. Lamprey larvae are sand-burrowing filter feeders that are cultivated for use in many types of basic and evolutionary research as well as to develop population control mechanisms. A steady, year-round supply of research animals is often needed for these types of research, however high maintenance and mortality can be problematic when keeping the larvae in aquaria with static or recirculating water systems. We have developed an inexpensive method of utilizing 2-liter (L) zebrafish breeder tanks placed within larger (38-L) aquaria. The individual sand bed compartments in the breeder tanks can be used with either flowing or recirculating systems, and also be used with other small, sand-burrowing organisms. The method allows for reduced maintenance time, ease of animal treatment and retrieval, and improved survival over systems where the sand media covers the aquarium bottom. We were able to maintain a stock of sea lamprey larvae in a recirculating system for use as needed over a 29-month period, with attrition comparable to a flow-through system.

Introduction

The sea lamprey (*Petromyzon marinus*) is an ancient vertebrate in a key position at the cusp of radical morphological and physiological changes from the jawless to jawed vertebrates. The sea lamprey is a species native to ocean and estuarine ecosystems, however it is a severe problem in the Great Lakes fisheries, having been accidentally introduced when canals were opened to bypass Niagara Falls. Lampreys hatch in freshwater in both ocean and lake systems, and the filter-feeding ammocoete larvae live in stream sediment for approximately five to seven years (Sower et al. 2009). Following metamorphosis, lampreys migrate from streams into the ocean or Great Lakes to spend 12 to 18 months in the parasitic phase, attaching to and extracting fluids from fish such as lake trout and salmon. Lamprey adults return to streams to spawn and then die (Jones 2007). In addition to research on pheromones (Johnson et al. 2009; Polkinghorne et al. 2001; Sorenson et al. 2003) to control invasive sea lamprey populations, recent studies have been aimed at probing the origins of vertebrate systems such as reproductive hormones (Bryan et al. 2007; Sower et al. 2009), thyroid hormones and developmental systems (Gross and Manzon 2011), the central nervous system (Shifman and Selzer 2007), the immune system (Amemiya et al. 2007; Guo et al. 2009) and the development of the collagenous skeleton (Zhang et al. 2006). For many of these types of studies, larval lampreys must be maintained in the laboratory.

If available, flow-through lake, river, or well water systems are utilized by labs studying lampreys. However, for research institutions without access to such systems, maintaining the animals without a flow-through system can be time intensive and result in high mortality rates. The animals are filter feeders that need to burrow in sand, which is more prone to bio-fouling problems and anoxic conditions than gravel used in fish-rearing tank systems. Additionally, based on established methods (Hanson et al. 1974; Mallatt 1983; and Rodriguez-Munoz et al. 2003) the larvae are fed yeast, introducing large amounts of nitrogen-rich biomass into the system.

Mallatt (1983) and Rodriguez-Munoz et al. (2003) studied larval lamprey growth using varying temperatures, animal densities, food types and feeding rates, and flowing

water systems; but did not report mortality. Also, the experimental periods of these studies were relatively short (six months and 10 days, respectively) compared with our study. Culture conditions in the Hanson et al. (1974) study were closest to those employed in the current study although they used a Lake Huron flow-through water system, but because the flow was turned off during each twice-weekly 24-hour feeding period the system was periodically static. They reared captive-spawned larvae, finding that the animals grew and thrived best when fed yeast at a rate of 105 grams (g) yeast/meter^2 (m^2), dependent on the area of the tank bottom rather than volume. After one year, 108 of the original 200 animals survived.

Our initial aquarium used under-gravel filters plus particulate/bio-filtration units, however these failed to control ammonia even though a smaller number of animals was contained in each aquarium. It was also difficult to remove animals from the sand without disturbing the whole system. We needed to develop an economical system to reduce the build-up of wastes and reduce maintenance time, as well as providing convenient treatment units and easy access to animals within the tanks. The compartment system we have devised has met these needs, and would also be useful in flow-through systems for isolating experimental treatment groups, as well as for use with other sand-borrowing species.

Methods

All animal housing and treatments were in accordance with the University of Minnesota Institutional Animal Care and Use Committee protocol (no. 1006A84140). We received 120 stream-caught lamprey larvae from Hammond Bay Biological Station (Millersburg, MI) ranging in total length from 42 to 116 millimeters (mm). They were initially distributed among one 57-L and three 38-L glass aquaria, with a 5-6 centimeter (cm) layer of Lake Superior beach sand that had been sieved through a No. 18 standard sieve. The 38-L aquariums have a bottom area of 25 cm by 50 cm (0.125 m^2). Tap water originating from Lake Superior, supplied by the City of Duluth was de-chlorinated using Tap Water Conditioner (API, Chalfont, PA). Aquaria were kept in a temperature-controlled room at 15^{o}C, under 27-watt full-spectrum fluorescent lamps on a 16-hr light:8-hr dark cycle. All aquaria were equipped with small-capacity particulate and bio-filtration units (Askoll model A-6..RN, Italy). In one 38-L aquarium, sand was placed directly on the tank bottom. In the other three, an under-gravel filter system with flow-through plastic plates was used (Penn-Plax, inc., Hauppage, NY). A layer of landscape fabric, which is non-toxic, durable in aquatic soils and allows flow of water and gases, was used to cover the plates and hold the sand above the free-water space beneath the plastic plates. Yeast suspensions were supplied on Mondays and Thursdays, generally corresponding to those of Hanson et al. (18-19 g for the 57-L aquarium and 13-14 g for the 38-L aquarium) during which filters were turned off for 24 hours.

The theory behind the under-gravel filter system is that the air stones force circulation to down-well through the sand and then up the standpipe through the activated-carbon filter head. However, within two weeks in all tanks ammonia levels tested greater than 8 milligrams (mg)/L (API ammonia test kit). The aquaria were also fouled with fungal and algal growth. While the dense sand layer likely prevented sufficient down-welling current, the results were no different in the aquarium without the under-gravel system. The use of zeolite ammonia-removal filter heads (Penn-Plax Ammonia-X under-gravel filter cartridges) and water treatment with AmQuell (Kordon LLC, Hayward, CA), as well as weekly water changes of one-half the aquarium volume

did not alleviate the problem. After five weeks, ammonia concentrations still exceeded 8 mg/L.

The aquaria were then dismantled and reconfigured with four individual 2-L zebrafish breeder tanks (Aquatic Habitats, Inc., Apopka, FL) within each of the three 38-L aquaria. Eleven of the original 120 animals died or were used in experiments leaving 109 animals to be moved to the breeder tanks. Each breeder tank has an upper container nested into a lower holding container. The upper containers are 190 mm long by 85 mm wide at bottom by 95 mm deep; the holding containers are 198 mm long by 90 mm wide by 118 mm deep. Each 2-L breeder tank set (with lid and divider) costs approximately $25. The bottom of the upper container is honey-combed with 3-mm hexagonal holes to allow the passage of fertilized zebrafish eggs when used for zebrafish breeding, but was covered with a layer of landscape fabric and 5 cm of sand into which the lamprey larvae burrow. This resulted in individual "sand bed" compartments in which eight to 10 larvae were housed. Only the upper container is used in the aquarium so that water can flow around and beneath each container (Figure A-1). Feet at the corners of the container provide a 3-mm space beneath the container. Two air stones are used in each aquarium, outside of the sand beds, and the same small-capacity filtration units, feeding regimen, temperature and light:dark cycle were used as in the under-gravel configuration.
Every three to four weeks for the first ten months, algal/fungal mats on the sand surface were removed by "vacuuming" with a siphon tube and the exposed surfaces of the sand bed containers were wiped. The lower containers were used when removing the sand beds from the aquaria, so that water remained full to the top of the 2-L tanks while the aquarium was being cleaned (i.e., the inhabited sand beds are not left drained of water). The aquarium surfaces were then thoroughly cleaned and fresh water added.

When animals were treated or needed for experiments, an individual sand bed could be removed from the aquarium without disturbing the other beds. The bed was placed in its lower holding container and the landscape fabric pulled out from under the sand; the sand was sifted with light agitation into the holding container leaving the larvae free of sand in the upper container. A new sand bed was prepared in another container into which the animals were quickly placed after treatment.

At random intervals from one week to two months when treatments occurred or to take animals for experiments, the animals were removed from the sand and the sand was rinsed or exchanged with used sand that was recycled (rinsed and allowed to dry). Beach sand previously unused in the aquaria was also occasionally incorporated to replace sand lost over time due to waste siphoning. Animal density was held relatively constant by consolidating the remaining animals after animals were removed for use in experiments. During one period (11-14 months) when no animals were removed for experiments, water was changed at seven to 21-day intervals but sand was not changed or flushed.

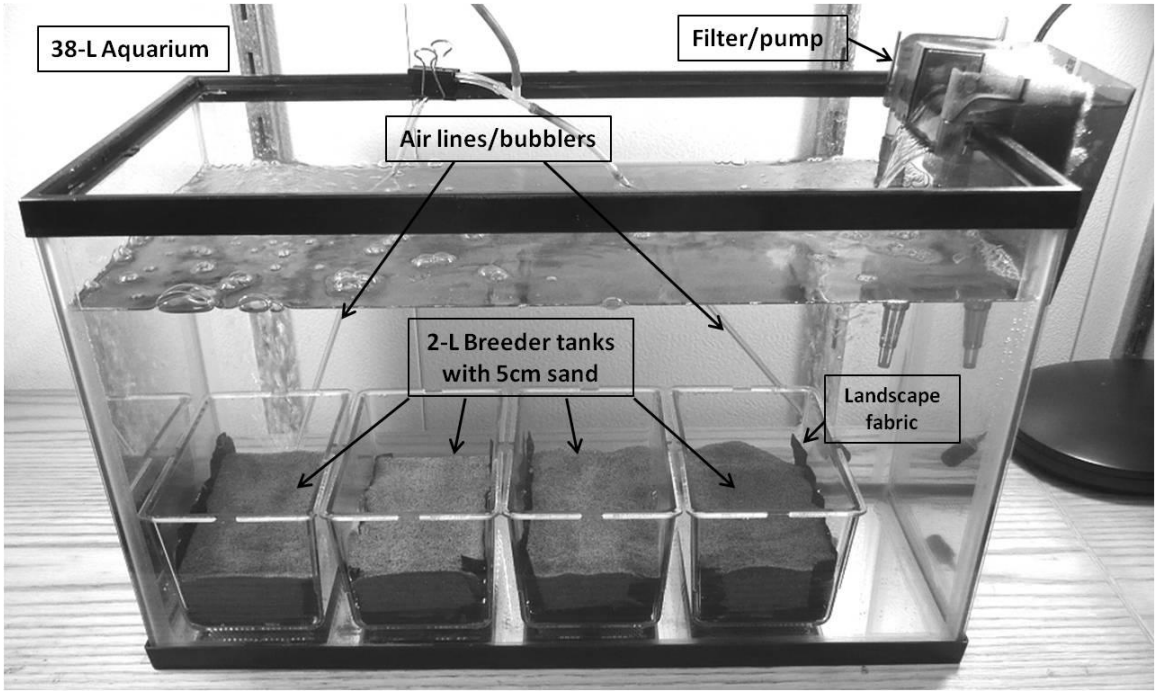


Figure A-1. The breeder tank sand bed compartments set up with 2-L zebrafish breeder tanks (Aquatic Habitats, Inc., Apopka, FL) in a 38-L aquarium.

Results and Discussion

With the initial (under-gravel filter) aquarium set-up, weekly water changes, in-place tank and filter cleaning and ammonia treatments took approximately eight hours per week and were ineffective at maintaining appropriate water quality. Ammonia concentrations remained in excess of 8 mg/L at five weeks, beyond the expected ammonia spike at approximately two weeks and decline by three weeks in a typical aquarium nitrogen cycling process. This was probably due to the loading of nitrogen from yeast at a level much higher than that of usual fish feeding levels, and the inability of the recirculating filtration system to remove the uneaten yeast and larval waste materials. Using the breeder tanks, the sand beds can be removed from an aquarium with minimal disturbance, the entire aquarium cleaned and water changed, and beds replaced in the aquarium in about one hour per aquarium. This cleaning schedule reduced maintenance time four fold.

When animals were placed in a sand bed with overlying water, they burrowed easily and quickly (usually within seconds). Once animals were burrowed and the beds placed in the tanks, an animal occasionally left the sand at night and was found the next day swimming at the bottom of the aquarium. These animals sometimes found their way back into a sand bed, or were caught with a dip net and placed back into a bed. Escapes could be controlled if necessary by covering the sand bed containers with screen having fairly large openings so as not to impede water flow and filter feeding of yeast. However, escapes did not cause any serious problems and during treatment studies we used small 10-L aquaria with one sand bed per aquarium to insure that animals did not move from one treatment group to another.

The size-weight distribution of the 109 animals when placed in the breeder tank compartment ranged from 42 mm (0.15 g) to 116 mm (2.29 g) (Figure A-2). The size of animals used in experiments was not random; larger (longer and heavier) animals were generally selected to assure sufficient quantity of the tissues and cells being studied. Animals were removed or lost at irregular intervals and therefore growth rate was not monitored. In the first half (15 months) using the breeder tank system, 28 healthy larvae were euthanized for use in experiments, and 34 animals remained, for a loss of 47

animals or 43%. This included 12 small larvae that died due to being pulled up and trapped in the filters, which would put deaths due to unknown causes at 27.5%. A screen could be placed over the intake to prevent larval entrapment.

Additionally, no animals were removed or lost in the first four months, and total mortality in the first ten months was only 22 animals (20%); more than half of the mortality (25 larvae) occurred in months 11-14, when the sand was not changed or flushed. In aquaculture systems, washing the sand removes accumulated waste material and prevents anaerobiosis while reusing sand maintains denitrifying bacterial flora (New and Valenti 2000). During the first month to six weeks after sand is flushed, ammonia did not exceeded 1 mg/L, but ammonia concentration when sand was unchanged for more than six weeks reached 2 mg/L. Thus, if the sand culture is left unflushed, its bacterial flora and the recirculating filter are not sufficient to remove all nitrogenous waste; occasional washing, similar to back-flushing a sand filter for wastewater treatment, is necessary to control nitrogenous waste and maintain survival. Sand can be more easily flushed in the individual containers than when lining the entire bottom of an aquarium. The 34 animals that remained after 15 months were weighed and measured, and ranged from 51 mm (0.16 g) to 93 mm (1.17 g) (Figure A-3). The cultivation period ended at 29 months because the final nine otherwise healthy animals died overnight due to an over-feeding error, and five had been trapped in the filter during the second half (14 months) of the cultivation period. Five animals were euthanized and 15 animals died due to unknown causes in those 14 months, or 44% of the 34 animals that had already been in captivity for 15 months. Therefore the 45 total deaths due to unknown causes over the full 29 months were actually 41% of the initial 109 animals.

The rate of attrition that Hanson et al. (1974) experienced (35% at six months and 46% at 12 months) may have been due to the young age of the animals in their study, with average lengths of 9 mm at the beginning of the study to 54 mm (35-65 mm range) after 12 months. They reported more favorable results in preliminary unpublished experiments feeding yeast to large, stream-caught lamprey, but did not provide quantitative details. Polkinghorne et al. (2001) reported a very low mortality rate (1% in a year) while maintaining larvae in a flow-through system during migratory pheromone

studies. They did not report the sizes of the larvae but they did report that some larvae metamorphosed during the study. The animals we cultivated were in a size category when received that was below the minimum size for probable metamorphosis within their natural stream habitats as statistically determined by Henson et al. (2003) of 120 mm and 3 g; the mean size before actual metamorphosis varied between 141 mm (4.54 g) and 150 mm (5.12 g), respectively, in the two rivers they studied. The largest animals we cultivated were the earliest removed for experiments and thus would not have been expected to metamorphose within the period of captivity in our system.

Employing 2-L zebrafish breeder tanks within larger aquaria has resulted in an effective and inexpensive method to maintain and experimentally treat larval lampreys, with or without flowing water systems, with survivorship comparable to that for similar-aged larvae in a flow-through system. The method can be utilized for many types of studies of this and related jawless vertebrate species or other sand-burrowing filter feeders.

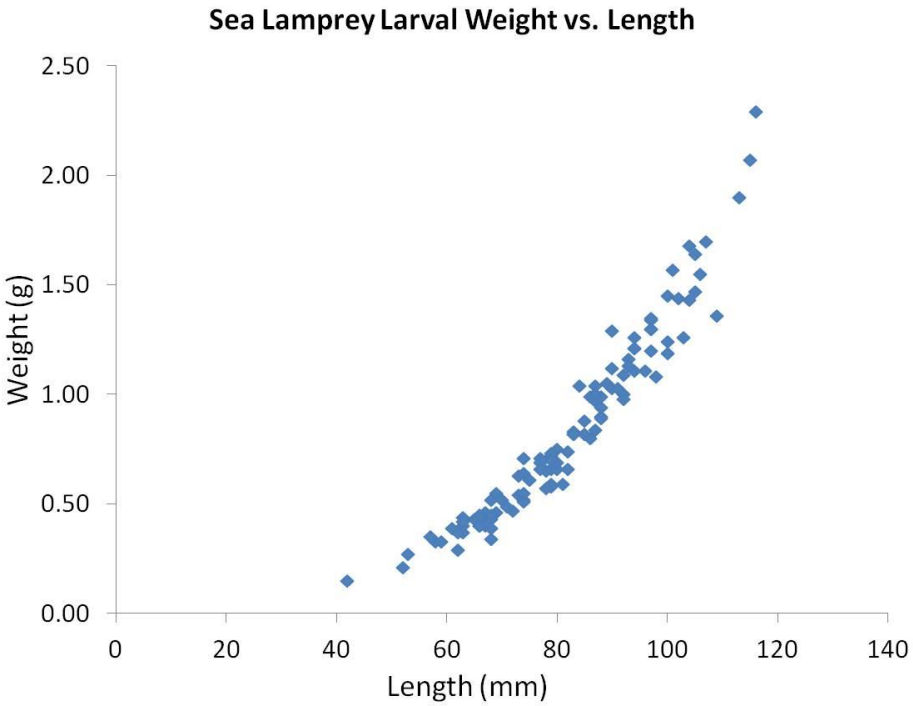


Figure A-2. Initial length-weight distribution of larval lampreys housed in sand beds (n=109).

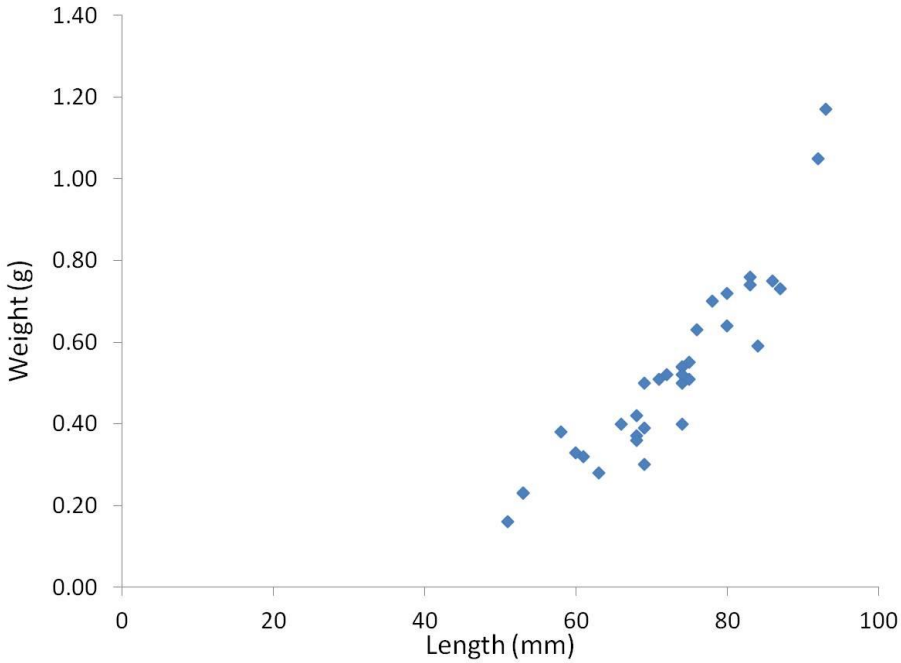


Figure A-3. Length-weight distribution of larval lampreys remaining after 15 months in recirculating system (n=34).

References

Amemiya, C. T., N. R. Saha, and A. Zapata. 2007. Evolution and development of immunological structures in the lamprey. Current Opinions in Immunology 19:535–541.

Bryan, M. B., A. P. Scott, and W. Li. 2007. The sea lamprey (*Petromyzon marinus*) has a receptor for androstenedione. Biology of Reproduction 77:688–696.

Gross, T.N., and R.G. Manzon. 2011. Sea lamprey (*Petromyzon marinus*) contain four developmentally regulated serum thyroid hormone distributor proteins. General and Comparative Endocrinology 170:640-649.

Guo, P., Hirano, M., Herrin, B. R., Li, J., Yu, C., Sadlonova, A., et al. 2009. Dual nature of the adaptive immune system in lampreys. Nature 459:796-801.

Hansen, L.H. E.L. King, Jr., J.H. Howell, and A. J. Smith. 1974. A culture method for sea lamprey larvae. The Progressive Fish-Culturist 36:122-128.

Henson, M. P., R. A. Bergstedt, and J. V. Adams. 2003. Comparison of spring measures of length, weight, and condition factor for predicting metamorphosis in two populations of sea lamprey (*Petromyzon marinus*) larvae. Journal of Great Lakes Research 29 (Supplement 1):204–213.

Johnson, N. S., S.-S.Yun, H. T. Thompson, C. O. Brant, and W. Li. 2009. A synthesized pheromone induces upstream movement in female sea lamprey and summons them into traps. Proceedings of the National Academy of Sciences 106:1021-1026.

Jones, M. L. 2007. Toward improved assessment of sea lamprey population dynamics in support of cost-effective sea lamprey management. Journal of Great Lakes Research 33:35-47.

Mallatt, J. 1983. Laboratory growth of larval lampreys (*Lampetra (Entosphenus) tridentata* Richardson) at different food concentrations and animal densities. Journal of Fish Biology 22:293-301.

New, M. B., and W. C. Valenti. 2000. Freshwater prawn culture: the farming of *Macrobrachium rosenbergii*. Blackwell Science, Malden, MA, 443pp.

Polkinghorne, C. N., J. M. Olson, D. G. Gallaher, and P. W. Sorensen. 2001. Larval sea lamprey release two unique bile acids to the water at a rate sufficient to produce detectable riverine pheromone plumes. Fish Physiology and Biochemistry 24:15–30.

Rodriguez-Muñoz, R., A.G. Nicieza, and F. Braña. 2003. Density-dependent growth of Sea Lamprey larvae: evidence for chemical interference. Functional Ecology 17:403-408.

Shifman, M.I., and M.E. Seltzer. 2007. Differential expression of class 3 and 4 semaphorins and netrin in the lamprey spinal cord during regeneration. The Journal of Comparative Neurology 501:631–646.

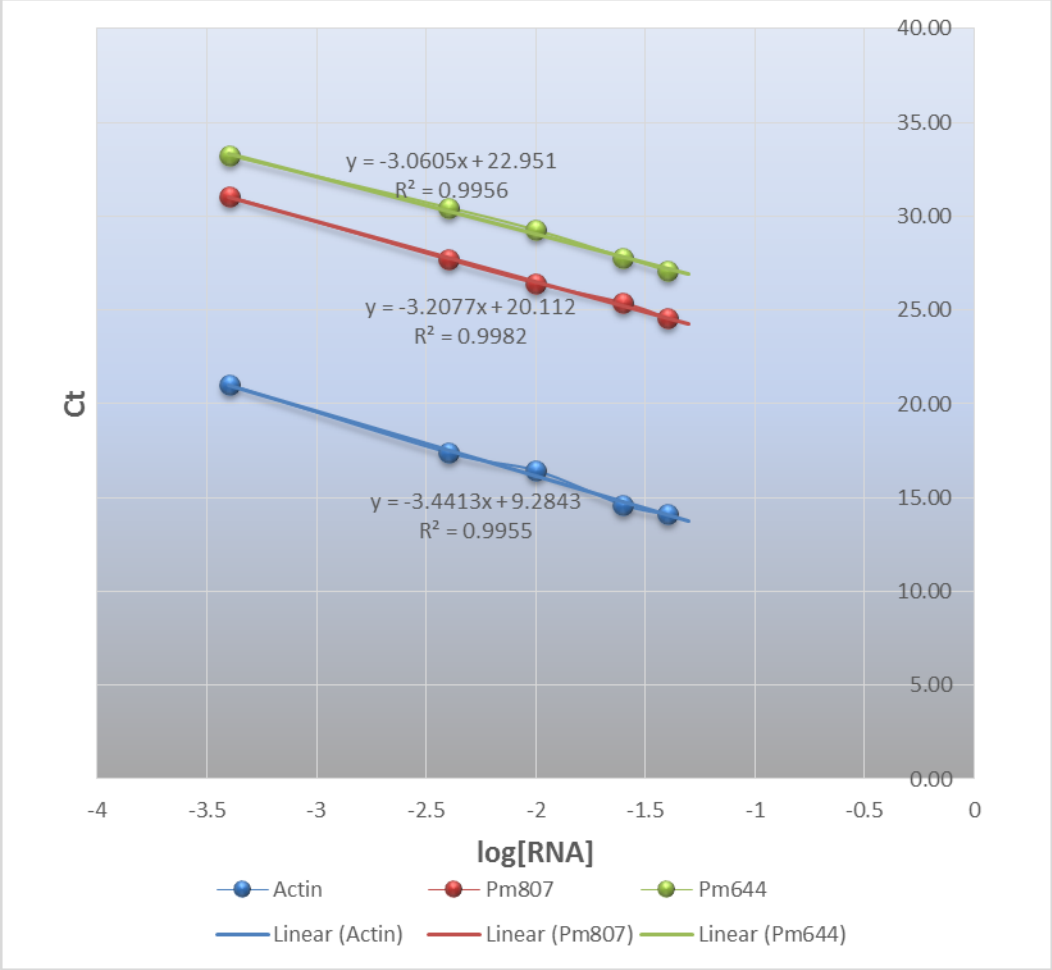
Sorensen, P.W., L.A. Vrieze, and J.M. Fine. 2003. A multi-component migratory pheromone in the sea lamprey. Fish Physiology and Biochemistry 28:253–257.

Sower, S. A., M. Freamat, and S. I. Kavanaugh. 2009. The origins of the vertebrate hypothalamic–pituitary–gonadal (HPG) and hypothalamic–pituitary–thyroid (HPT)

endocrine systems: New insights from lampreys. General and Comparative Endocrinology 161:20–29.

Zhang, GJ, M. M. Miyamoto, and M. J. Cohn. 2006. Lamprey type II collagen and Sox9 reveal an ancient origin of the vertebrate collagenous skeleton. Proceedings of the National Academy of Sciences 103:3180-3185.

Appendix B. Supplemental Data



Supplementary Figure 2-1. Quantitative PCR efficiency and dynamic range were determined on forebrain cDNA for actin, Pm807 and Pm644 using a series of 0.4 ng, 4 ng, 10 ng, 25 ng and 40 ng RNA, in order to select a concentration for the assay (25 ng). Our intention was to quantitatively compare expression between the forebrain and hypothalamus, and between genders within each receptor.

Efficiency= $(10^{(-1/\text{slope})} - 1) \times 100$=

Actin	Pm807	Pm644
95.30%	105.00%	112.20%

Normalization: Mean Relative Expression = triplicate mean actin Ct for each animal – triplicate mean receptor Ct for that animal raised to the 2nd power, or $2^{\Delta Ct}$.

Supplementary Fig. 2-2A through 2-2G. Nucleotide and deduced amino acid sequences with primer locations, intron-exon junctions, transmembrane (TM) domains and structural features noted. Locations of TMs were determined using HMMTOP (http://www.enzim.hu/hmmtop/server/hmmtop.cgi).

Suppl. Fig. 3A. Sea lamprey V1a/OXT receptor Pm807

```
agagatgaaaacggagtggtggtggtgggagagacaacgccagcctataacAltTSSatacacagcacggctccctatctgcaggtctgctgcgtgtgctct
taaagaggcgcagagcctccatcctcctctgcattcaatctggggacagcgacgctgcagctgtcactcagcctctggaagtg38kbattttgggaggaaaa
                          AltTIS                                                        intron
atgccgggaaacgtgacgggcgagatggagggtgccatcctgaatggtaccggcgtggcctacgccccggcaacttcgaacggctcacatgtggtgcccgca
 M  P  G  N  V  T  G  E  M  E  G  A  I  L  N  G  T  G  V  A  Y  A  P  A  T  S  N  G  S  H  V  V  P  A
gaggccttcaccgctgccctcgcctccatcaacgccacgctcaacggctcgtcgcagcacgccctggaccgcaatgaggaggtggccaaagtcgagatcgcc
 E  A  F  T  A  A  L  A  S  I  N  A  T  L  N  G  S  S  Q  H  A  L  D  R  N  E  E  V  A  K  V  E  I  A   TM I
ttactctcgatcatcttgttcgtggccatcgtgggcaacgtctgcgtgctgctggcactcatcaacacccgcaagaagacgtcgcgcatgcatctcttcatc
 L  L  S  I  I  L  F  V  A  I  V  G  N  V  C  V  L  L  A  L  I  N  T  R  K  K  T  S  R  M  H  L  F  I   TM II
atgcacctgagcattgcggacctggtggttgccttcttccaggtgctgccgcagctcatctggaaaatcacctaccgcttcaacggctccgacttcctgtgt
 M  H  L  S  I  A  D  L  V  V  A  F  F  Q  V  L  P  Q  L  I  W  K  I  T  Y  R  F  N  G  S  D  F  L  C
cgcgccatcaagtacctgcagatcctgggaatgttcgcctccacgtacgtgctcatcatgatggggctggaccgctacattgccatctgccacccctctaagg
 R  A  I  K  Y  L  Q  I  L  G  M  F  A  S  T  Y  V  L  I  M  M  G  L  D  R  Y  I  A  I  C  H  P  L  R   TM III
acgctccgccagtcttccaaacagtcctaccagatgatcttcgtcagctggttcctgagcatgctgttcagcctgccgcaggcgttcatcttctcgatgtcc
 T  L  R  Q  S  S  K  Q  S  Y  Q  M  I  F  V  S  W  F  L  S  M  L  F  S  L  P  Q  A  F  I  F  S  M  S   TM IV
gaagtggagaactcgggcatcatcgactgctgggctgagttcatcaagccctggggcaccaaggcgtacatcacctggatgacaggctcagtcttcatcatc
 E  V  E  N  S  G  I  I  D  C  W  A  E  F  I  K  P  W  G  T  K  A  Y  I  T  W  M  T  G  S  V  F  I  I   TM V
ccggtgatcatcctcatctggtgctacggcatgatcaccttcgccatctggaaaaacatcaaggccaagacgcaggaaggtgacagcagacacaacccggcc
 P  V  I  I  L  I  W  C  Y  G  M  I  T  F  A  I  W  K  N  I  K  A  K  T  Q  E  G  D  S  R  H  N  P  A
aagagcagcgccccttccagggtcagctcggtgcgcagcatctccaaggcgaagatccgcaccgccaagatgacgttcgtcatcatcatggtgtacatcatc
 K  S  S  A  P  S  R  V  S  S  V  R  S  I  S  K  A  K  I  R  T  A  K  M  T  F  V  I  I  M  V  Y  I  I   TM VI
tgctggaccccgttcttcttcgtgcagatgtggtcagtctgggacagctccgcgccttttgaa4528bpggcatcccatttgccatcgtcatgctgctggcg
 C  W  T  P  F  F  F  V  Q  M  W  S  V  W  D  S  S  A  P  F  E  intron  G  I  P  F  A  I  V  M  L  L  A   TM VII
agcctcaacagctgcaccaacccctggatttacatgttcttcagcgggcacctactctacgactttgtgcgctacttcccgtgcggggcgcgctcgcgtgcc
 S  L  N  S  C  T  N  P  W  I  Y  M  F  F  S  G  H  L  L  Y  D  F  V  R  Y  F  P  C  G  A  R  S  R  A
cgctcccgcgaggcggacgagtcgcgtgccagcgacagcagccgccggaatcacacgttcgtgagtcgcctgaccaggaggagcctcactctctcgtcgggc
 R  S  R  E  A  D  E  S  R  A  S  D  S  S  R  R  N  H  T  F  V  S  R  L  T  R  R  S  L  T  L  S  S  G
tcgcagcacgaggaggcgtcgtcgcgcaccactagcctctctcccgtcgcacgagtgcccaagacgtactttgcgtga
 S  Q  H  E  E  A  S  S  R  T  T  S  L  S  P  V  A  R  V  P  K  T  Y  F  A  stop codon
```

Primary PCR segment (991 bp) **Nested PCR segment (285 bp)** **AltTSS**=Alternate Transcription Start Site

AltTIS=Alternate Translation Initiation Site

Suppl. Fig. 3B. Sea lamprey V1a/OXT Receptor Pm3133

```
                  atggccaacggcacggcgaatgcatccgagtcgacgctgagccgcgacgaagatctcgccaaggtggagatcacgatcctcgcc
                    M  A  N  G  T  A  N  A  S  E  S  T  L  S  R  D  E  D  L  A  K  V  E  I  T  I  L  A   TM I
atcatcctctgcgcggccgtggtcggcaatgcctgcgtgctcctggcgctgctcagcaccaagaagaagacgtcgcggatgcaccttttatcatgcacctg
 I  I  L  C  A  A  V  V  G  N  A  C  V  L  L  A  L  L  S  T  K  K  K  T  S  R  M  H  L  F  I  M  H  L   TM II
agcatcgccgacctggtagtggccttcttccag2308bpgtgctgccccagctcatctgggaggtgaccttccgcttcaatggctcggacatactgtgccgc
 S  I  A  D  L  V  V  A  F  F  Q intron V  L  P  Q  L  I  W  E  V  T  F  R  F  N  G  S  D  I  L  C  R
accgtcaagtacctgcagatcctcggcatgttcgcctccacctacatgctcatcatgatggccatggaccgctacatcgccatctgccacccgctgcggacg
 T  V  K  Y  L  Q  I  L  G  M  F  A  S  T  Y  M  L  I  M  M  A  M  D  R  Y  I  A  I  C  H  P  L  R  T   TM III
atccagcagtcgtcgacgcagtcctacatgatgatcctcgtcagctggctcgtgagcttcctcttcagcgtcccgcaggcgtacatcttttcgctgcgcgag
 I  Q  Q  S  S  T  Q  S  Y  M  M  I  L  V  S  W  L  V  S  F  L  F  S  V  P  Q  A  Y  I  F  S  L  R  E   TM IV
gtgcagcagggcacgggcgtctacgactgctgggccgcctttgtcgagccgtgggggctcaaggcgtacgtcacgtggaccgccgtcgccgtgttcgtcgcg
 V  Q  Q  G  T  G  V  Y  D  C  W  A  A  F  V  E  P  W  G  L  K  A  Y  V  T  W  T  A  V  A  V  F  V  A   TM V
cccgtcgccgtgctgctctggtgctacggcatgatcaccgccgagatctggaggaacatgcgggccaagatggagaggcgcggcgggcgggacggccg102bpc
 P  V  A  V  L  L  W  C  Y  G  M  I  T  A  E  I  W  R  N  M  R  A  K  M  E  R  R  G  G  R  D  G  Rintron
ccgccgccgaaggggacgccggcgcggtcgtccgcggcggtggcgccggtggcgaccgcgacagcgtcgcgggtcagctcggtgaggaacatctccaaggcg
 P  P  P  K  G  T  P  A  R  S  S  A  A  V  A  P  V  A  T  A  T  A  S  R  V  S  S  V  R  N  I  S  K  A
aagatccgcaccaccaagatgaccttcgtcatcatcgtggtgtacgtcttctgctggacgccgtttttcctcgtccagatgtggtccgtgtgggacgggtct
 K  I  R  T  T  K  M  T  F  V  I  I  V  V  Y  V  F  C  W  T  P  F  F  L  V  Q  M  W  S  V  W  D  G  S   TM VI
gctccctttgaag7161bpggtcggcgttcaccatcgtgatgctgttggccagcctcaacagctgcaccaacccgtggatctacatgttcttcagcgggcat
 A  P  F  E intronG  S  A  F  T  I  V  M  L  L  A  S  L  N  S  C  T  N  P  W  I  Y  M  F  F  S  G  H   TM VII
ctcatttacgacttcgcgcactgcatcccgtgctgtgcgggcgcgctccgtcgcctggaaacgacgggcgccggccacaacgccaccgaccggctcagcgcc
 L  I  Y  D  F  A  H  C  I  P  C  C  A  G  A  L  R  R  L  E  T  T  G  A  G  H  N  A  T  D  R  L  S  A
actccgcagaggcaccgcagcgtcaccctgagctcctgcagccagaagtag
 T  P  Q  R  H  R  S  V  T  L  S  S  C  S  Q  K stop codon
```

Primary PCR segment (829bp) **Nested PCR segment (595bp)**

Suppl. Fig. 3C. Sea lamprey V1B/A/OXT Receptor Pm2017

taatgcactgcccataggaggaggaag**8414bp intron**cgcacatcgcagctcgcagtccccaccgtcgtcgtcgt**27bp deletion**cgtcgtcaaccccgcgcgccaccgac

actgcccgccgccaccctcggcgcctcccgcggcgcgcagcg**atg**gacctcgcgccgtggaggaacgccagccggacgccccgccaccacaacgtctcccac

M D L A P W R N A S R T P R H H N V S H

gcggctcccaacgccgccgaccggcccgaggacgcggcggcggcggcggcggcggcacgcgacgagcgcctcgccgcggcggagatcgccctgctggccgcc

A A P N A A D R P E D A A A A A A A A A R D E R L A A A E I A L L A A TM I

atcgtcgccgtcgccatcgccggcaacggctcggtgctgctggcgctgagccgcacgcgccgcaaggcgtcgcgcatgaacctcttcgtcaagcacctgagc

I V A V A I A G N G S V L L A L S R T R R K A S R M N L F V K H L S TM II

gtcgccgacctggccgtggccgccttccaggtcctgccgcagctgtcgtgggacgtgaccttccgcttccgcgggcccgacgcgctctgccgcctcgtcaag

V A D L A V A A F Q V L P Q L S W D V T F R F R G P D A L C R L V K TM III

tacctgcaggtggtgggca**tgttcgcgtccacctacgtgctcat**cgccatgaccgtcgaccgctacctcgccatctgccacccgctgagcacgctgcgccgg

Y L Q V V G M F A S T Y V L I A M T V D R Y L A I C H P L S T L R R

cggggcgagccgaggaagcaggcgcacgccatggtggcctgcgcgtgggcgctgagcgccctgctcagcacgccgcaggtggtcatcttctcgctgcgggag

R G E P R K Q A H A M V A C A W A L S A L L S T P Q V V I F S L R E TM IV

gtggaggagggcgtcttcgactgctgggccgacttcgggcagccgtggggcatgcgggcgtacgtgacgtggatcacggtgtccgtgtacgtggcgcccgtg

V E E G V F D C W A D F G Q P W G M R A Y V T W I T V S V Y V A P V TM V

ctcatcctggcggcctgctacggggccattctgctcgagatttgcag**2439bp intron**gaacctgcgcctcaagacgcgtggccgataccccaaggacccccggggc

L I L A A C Y G A I L L E I C R N L R L K T R G R Y P K D P R G

tcgcagtctcaatcgcccgtgtgctcgcgcgtgagcagcgtgcgctgcatctcccgcgccaaggtgcgcaccatcaagatgactctcg**tcatcatcgtggtc**

S Q S Q S P V C S R V S S V R C I S R A K V R T I K M T L V I I V V TM VI

tacgtcgtctgctggactcccttcttcgtcatccagatgtgggcggcctgggacgagaccgcaccagacgacg**5835bp intron**actcatccgaccccacctacacc

Y V V C W T P F F V I Q M W A A W D E T A P D D D S S D P T Y T

atcgtgatgctcctttcgagcctcaacagctgcaccaacccgtggatctacatgtcgttcagcgggcacctcc**ttgaggaggtgatgtcctgct**gccgccag

I V M L L S S L N S C T N P W I Y M S F S G H L L E E V M S C C R Q TM VII

gcgcggtgcgccaccgcgtgccgccggcagcagcggcggcggcggcggtgccgaacgccgacgaggcgcggggcgggctcccccaacac**agagcacgtggtg**

A R C A T A C R R Q Q R R R R R C R T P T R R G A G S P N T E H V V

gtgtacgcactgtcgccggcacccacgcagtgccgcagctccccgccaccgccgcaggtggcaccgccacccccccgagacgcggagatccgcagcgccg

V Y A L S P A P T Q C R S S P A T A A G G T A T P P E T R R S A A P

ctgtggtcgccagcgttaccg**tag**

L W S P A L P **stop codon**

Primary PCR segment (796bp) **Nested PCR segment (204bp)**

deleted sequence compared with genome database: cgtcgccgtcatcgtcgtcatcgtcgc

Suppl. Fig. 3D. Sea lamprey V2b/c Receptor Pm644

```
  cacgggaacgatgagtgacgatccattcgttgggaactcctcggacgagagcgccagcacgctggcgagcggtgcctggccctccgccgctcccaacggc
            M  S  D  D  P  F  V  G  N  S  S  D  E  S  A  S  T  L  A  S  G  A  W  P  S  A  A  P  N  G
accgcagcgccgcacccgccgctggcgaggcgccccgagcgtgacgaggcgctggctcgagccgaggtcaccgtgctcggcctgatctttgtgctggcgacg
 T  A  A  P  H  P  P  L  A  R  R  P  E  R  D  E  A  L  A  R  A  E  V  T  V  L  G  L  I  F  V  L  A  T  TM I
agcggcaacgccacgctgctcgtgtcgttgtggaggaggcgcaagcacgcgtcgcgcatgcacgccttcctcgtgcacctcagcgtcgccgacctcgtcgtc
 S  G  N  A  T  L  L  V  S  L  W  R  R  R  K  H  A  S  R  M  H  A  F  L  V  H  L  S  V  A  D  L  V  V  TM II
gccttcttccaggtgctgccgcagctcgcgtgggacgtcaccgacgccttccag547 bpgcatccaacggcgtgtgtaaggtggtgaagtacctgcaggtg
 A  F  F  Q  V  L  P  Q  L  A  W  D  V  T  D  A  F  Q intron A  S  N  G  V  C  K  V  V  K  Y  L  Q  V  TM III
gtcggGATGTTTGCGTCCGCGTACATGATcgtggcgatgaccgtggaccgcttccaggcggtctgctacccgatggtcaccttcaagaagaagcgctcgcgc
 V  G  M  F  A  S  A  Y  M  I  V  A  M  T  V  D  R  F  Q  A  V  C  Y  P  M  V  T  F  K  K  K  R  S  R
tggaacggcctcgtgtgcgccgcgtgggccacctcgctggccctcagcacgccgcagatattcatcttctcgctgagcgagtacgaaaagggcatcttcgat
 W  N  G  L  V  C  A  A  W  A  T  S  L  A  L  S  T  P  Q  I  F  I  F  S  L  S  E  Y  E  K  G  I  F  D  TM IV
tgcaccgcgacgttcgcggcacactggggcgccaaggcgtacgtgacgtgggtgacgctggccgtgttcgtgctgcccacgctcgcagtggtcatttgccag
 C  T  A  T  F  A  A  H  W  G  A  K  A  Y  V  T  W  V  T  L  A  V  F  V  L  P  T  L  A  V  V  I  C  Q  TM V
gctcagatctgccgcatcatccgcctcaacctctacgtgaagacgcaccagggccccgacgaggaggagacgatcggcggggacgagcgccgtgccgggggg
 A  Q  I  C  R  I  I  R  L  N  L  Y  V  K  T  H  Q  G  P  D  E  E  E  T  I  G  G  D  E  R  R  A  G  G
cagctcatgccgtcgcgggccagcagcgtcacgggGATCTCCAG2766bpGGCCATGATCAACAcggtgcgcatgacgctggtgatcgtgctggtgtacgtg
 Q  L  M  P  S  R  A  S  S  V  T  G  I  S  Rintron  A  M  I  N  T  V  R  M  T  L  V  I  V  L  V  Y  V  TM VI
gcgtgctgggcaccgtggttcaccgtgcagctctggtcggcctgggacagcaaggctcccaaggaagG2821bpgccccacgttcgtcatcatcatgctgctg
 A  C  W  A  P  W  F  T  V  Q  L  W  S  A  W  D  S  K  A  P  K  E intronG  P  T  F  V  I  I  M  L  L  TM VII
ggaaacctcaacagctgcaccaaccccctggatctacctgtggttcgcgggggggctgcccccgggcggctgccgggggctcgtgccgcacgcacgcccaccg
 G  N  L  N  S  C  T  N  P  W  I  Y  L  W  F  A  G  G  L  P  P  G  G  C  R  G  L  V  P  H  A  R  P  P
gcgcccgacttcaacgtggacgagtcggtgatgagcacagccacattccgcgaagatcgcgacgactcgctgcgtggcggagacaccccgactcccccgttc
 A  P  D  F  N  V  D  E  S  V  M  S  T  A  T  F  R  E  D  R  D  D  S  L  R  G  G  D  T  P  T  P  P  F
gcatgcctgccgccctcgcaggcgcaggctgccagctcgctgcacgccgtgtcctgcctgccctga
 A  C  L  P  P  S  Q  A  Q  A  A  S  S  L  H  A  V  S  C  L  P stop codon
```

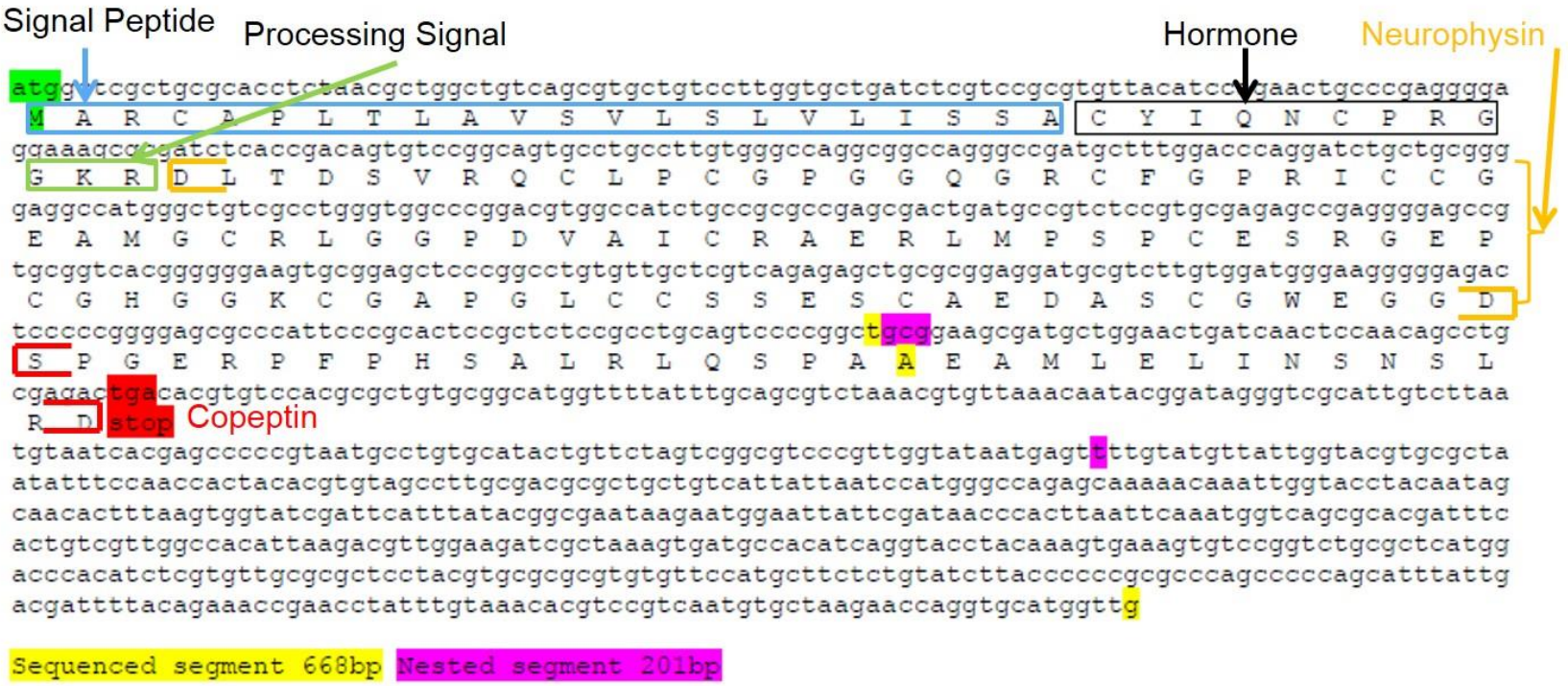
Primary PCR segment (996bp) **Nested PCR segment (461bp)**

Suppl. Fig. 3E. Sea lamprey V2 Receptor pseudogene (Pm4769)

```
acgtcagcggcaccgcggccaggagccaacggctccgcgtcgcgcaacgagcgcaacgaggcgctggcgagagccgaggtggcggtgctgttcgccatgttc
 T  S  A  A  P  R  R  P  G  A  N  G  S  A  S  R  N  E  R  N  E  A  L  A  R  A  E  V  A  V  L  F  A  M  F  TM I
gcatgcaccacgcttggtaactcggtgctgctcacggcgctctaccgacggcgcaagcacgcgtcacgcatgcacgtgttcatggtgcacctgtgcctcgcc
 A  C  T  T  L  G  N  S  V  L  L  T  A  L  Y  R  R  R  K  H  A  S  R  M  H  V  F  M  V  H  L  C  L  A  TM II
gacctggtggtggctttcttccaggtgcttccgcagctcgcgtgggacatcaccgacgtgttcctcgcgtccgacttcgcctgccgcttcatcaagtacatg
 D  L  V  V  A  F  F  Q  V  L  P  Q  L  A  W  D  I  T  D  V  F  L  A  S  D  F  A  C  R  F  I  K  Y  M  TM III
cagatcgtgggcatgtttgcctccacctacatgatcgtgggcacgacggtggaccgataccaggcggtctgcttccccatggtcaccttcaagaagcgtatg
 Q  I  V  G  M  F  A  S  T  Y  M  I  V  G  T  T  V  D  R  Y  Q  A  V  C  F  P  M  V  T  F  K  K  R  M
gcttactggaacgcgcccgtgtgcgccgtgtgggtggtctcgctgctgctcagcgtgccccaggcggtcatcttcgccaaggggggaggtctacccgggcgtc
 A  Y  W  N  A  P  V  C  A  V  W  V  V  S  L  L  L  S  V  P  Q  A  V  I  F  A  K  G  E  V  Y  P  G  V  TM IV
tacgactgctgggggacattccagcccatgtggggcgtcaaggcgtacataacctggatggccatggccatatttgtggtgcccaccgccgtgcttgtgtgg
 Y  D  C  W  G  T  F  Q  P  M  W  G  V  K  A  Y  I  T  W  M  A  M  A  I  F  V  V  P  T  A  V  L  V  W  TM V
tgccaaaccaagatctgccgcgtcatccggctgaacatctacttcaagacgccgcaggtgcacaaagacaagctgctgctgcgtcggcggaagcagcggcgg
 C  Q  T  K  I  C  R  V  I  R  L  N  I  Y  F  K  T  P  Q  V  H  K  D  K  L  L  L  R  R  R  K  Q  R  R
cggcgtcagcagaagcggcaagaggacgaggaggaggtggccgccctgcacaggggtcagacgcacgctgtgaccatgacccgggtgagcagcgtgaccggt
 R  R  Q  Q  K  R  Q  E  D  E  E  E  V  A  A  L  H  R  G  Q  T  H  A  V  T  M  T  R  V  S  S  V  T  G
gtctcgagggcgatggtgaagacggtgaagatgaccctggtcatcgtgatcgtatatgtcctctgctggtcgccgttcttcgtagcgcagctgtggagcgtg
 V  S  R  A  M  V  K  T  V  K  M  T  L  V  I  V  I  V  Y  V  L  C  W  S  P  F  F  V  A  Q  L  W  S  V  TM VI
tgggatcccaatccaccgttcgaagunknbpgtgccttgttcaccatcaccctgctgctggcaagcctcaacagctgcaccaacccgtggatctacatggcg
 W  D  P  N  P  P  F  E  intronG  A  L  F  T  I  T  L  L  L  A  S  L  N  S  C  T  N  P  W  I  Y  M  A  TM VII
ttcagcgggagcactccacgcgccctcttgtcctgcgtgctgtgccgtgcgcggagaggcggcgccggtggcattgcaggcggtggcgttggcagcgatgac
 F  S  G  S  T  P  R  A  L  L  S  C  V  L  C  R  A  R  R  G  G  A  G  G  I  A  G  G  G  V  G  S  D  D
ggcattggtggtggcggtggcactggcggttgggtgcggagcgtgccgggagccacggcgcacgaggactcgatggcgaccagcagcgtgcaactggcggga
 G  I  G  G  G  G  G  T  G  G  W  V  R  S  V  P  G  A  T  A  H  E  D  S  M  A  T  S  S  V  Q  L  A  G
gctctccgctcgcagcacgaggatgacgcggaacaaaccaacgaattaattaggtcgcgcacgattagcgcgtga
 A  L  R  S  Q  H  E  D  D  A  E  Q  T  N  E  L  I  R  S  R  T  I  S  A stop codon
```

Primary PCR segment (368bp) **Nested PCR segment (236bp)** ***Alternative splice intron (110bp)***

Suppl. Fig. 3F. Sea lamprey AVT preprohormone based on Japanese lamprey (*Lethenteron camtschaticum*)



(Sequenced segment in sea lamprey from this study matched that of L. camtschaticum. After Gwee et al., 2009)

Suppl. Fig. 3G. Sea lamprey β-actin, cytoplasmic 2-like (partial cds)

atcatgttcgagaccttcaacacgccggcaatgtacgtcgccatccaggccgtgctgtcgctgtatgcgtccggccgcaccaccggcatcgtgatggactct
 I M F E T F N T P A M Y V A I Q A V L S L Y A S G R T T G I V M D S
ggcgacggcgtgtcgcacaccgtgcccatctacgagggctacgcgctgccccacgccatcctgcgtctcgacctggctggccgcgacctcaccgactacctg
 G D G V S H T V P I Y E G Y A L P H A I L R L D L A G R D L T D Y L
atgaagatcctgacggagcgtggctactcattcaccacgacggccgagcgtgagatcgtgcgtgacatcaaggagaagctctgctacgtggcgctcgacttc
 M K I L T E R G Y S F T T T A E R E I V R D I K E K L C Y V A L D F
gagcaggagatggccactgccgcatcatcctcgtcgctcgagaagagctacgagctgcccgacggccaggtcatcaccattggcaacgagcgcttccgctgc
 E Q E M A T A A S S S S L E K S Y E L P D G Q V I T I G N E R F R C
Cctgaggccctcttccagccgtccttcctcg313 bpgaatggagtcgtgtggcattcacgagaccaccttcaactcgatcatgaagtgcgacgtcgacatc
 P E A L F Q P S F L intronG M E S C G I H E T T F N S I M K C D V D I
cgcaaggacctgtacgccaacaccgtgctgtctggtggcaccaccatgtaccccggcatcgccgaccgcatgcagaaggagatcacagccctggcgcccagc
 R K D L Y A N T V L S G G T T M Y P G I A D R M Q K E I T A L A P S
Accatgaagatcaag238 bpatcatcgccccaccagagcgcaagtactccgtctggatcggaggctccatcctggcgtcactgtccaccttccagcagatg
 T M K I K intron I I A P P E R K Y S V W I G G S I L A S L S T F Q Q M
tggatcagcaagcaggagt
 W I S K Q E

Sequenced segment (721bp)

Supplementary Figure 2-3 A-E. CLUSTAL 2.1 sequence alignments between sea lamprey (*Petromyzon marinus*) and predicted AVT receptor orthologs on *Lethenteron japonicum* scaffolds (Japanese lamprey genome database). Intron insertion points are highlighted.

Fig. 2-3A

```
LjapV1_KE993674  MLGNVTGEMEGDILNGTGVAYAPAASNGSHAAPAEAFTTALASINATLNGSSQNALDRNE
Lamprey_Pm807    MPGNVTGEMEGAILNGTGVAYAPATSNGSHVVPAEAFTAALASINATLNGSSQHALDRNE
                 * ********* ************:******..******:**************:******

LjapV1_KE993674  EVAKVEIALLSIILFVAIVGNVCVLLALINTRKKTSRMHLFIMHLSIADLVVAFFQVLPQ
Lamprey_Pm807    EVAKVEIALLSIILFVAIVGNVCVLLALINTRKKTSRMHLFIMHLSIADLVVAFFQVLPQ
                 ************************************************************

LjapV1_KE993674  LIWKITYRFNGSDFLCRTIKYLQILGMFASTYVLIMMGLDRYIAICHPLRTLRQSSKQSY
Lamprey_Pm807    LIWKITYRFNGSDFLCRAIKYLQILGMFASTYVLIMMGLDRYIAICHPLRTLRQSSKQSY
                 *****************:******************************************

LjapV1_KE993674  QMIFVSWFLSMLFSLPQAFIFSMSEVENSGIIDCWAEFIKPWGTKAYITWMTSSVFIIPV
Lamprey_Pm807    QMIFVSWFLSMLFSLPQAFIFSMSEVENSGIIDCWAEFIKPWGTKAYITWMTGSVFIIPV
                 ************************************************.*******

LjapV1_KE993674  IILIWCYGMITFAIWKNIKAKTQEGDSRHNPAKNSAPSRVSSVRSISKAKIRTAKMTFVI
Lamprey_Pm807    IILIWCYGMITFAIWKNIKAKTQEGDSRHNPAKSSAPSRVSSVRSISKAKIRTAKMTFVI
                 *********************************.**************************

LjapV1_KE993674  IMVYIICWTPFFFVQMWSVWDSSAPFEGIPFAIVMLLASLNSCTNPWIYMFFSGHLLYDF
Lamprey_Pm807    IMVYIICWTPFFFVQMWSVWDSSAPFEGIPFAIVMLLASLNSCTNPWIYMFFSGHLLYDF
                 ************************************************************

LjapV1_KE993674  VRYFPCGARSRARSREADESRASDSSRRNHTFVSRLTRRSLTLSSGSQHEEASSRTTSLS
Lamprey_Pm807    VRYFPCGARSRARSREADESRASDSSRRNHTFVSRLTRRSLTLSSGSQHEEASSRTTSLS
                 ************************************************************

LjapV1_KE993674  PVARVPKTYFA
Lamprey_Pm807    PVARVPKTYFA
                 ***********
```

Fig. 2-3B

```
LjapV1_KE994228  MANGTANASESTLSRDEELAKVEIAILAIILCVAVVGNACVLLALLSTKKKTSRMHLFIM
Lamprey_Pm3133   MANGTANASESTLSRDEDLAKVEITILAIILCAAVVGNACVLLALLSTKKKTSRMHLFIM
                 *****************:******:********.**************************

LjapV1_KE994228  HLSIADLVVAFFQVLPQLIWEVTFRFNGSDILCRTVKYLQILGMFASTYMLIMMAMDRYI
Lamprey_Pm3133   HLSIADLVVAFFQVLPQLIWEVTFRFNGSDILCRTVKYLQILGMFASTYMLIMMAMDRYI
                 ************************************************************

LjapV1_KE994228  AICHPLRTIQQSSTQSYLMILVSWLVSFLFSVPQAYIFSLREVQQGTGVYDCWAAFVEPW
Lamprey_Pm3133   AICHPLRTIQQSSTQSYMMILVSWLVSFLFSVPQAYIFSLREVQQGTGVYDCWAAFVEPW
                 *****************:******************************************

LjapV1_KE994228  GLKAYVTWTAVAVFVAPVVVLIWCYGMITAEIWRNMRAKTERRGGREGRPPPKGMLARSS
Lamprey_Pm3133   GLKAYVTWTAVAVFVAPVAVLLWCYGMITAEIWRNMRAKMERRGGRDGRPPPKGTPARSS
                 ******************.**:***************** ******:*******  ****

LjapV1_KE994228  AAVAPPAAATASRVSSVRNISKAKIRTTKMTFVIIVVYVFCWTPFFLVQMWSVWDGSAPF
Lamprey_Pm3133   AAVAPVATATASRVSSVRNISKAKIRTTKMTFVIIVVYVFCWTPFFLVQMWSVWDGSAPF
                 ***** *:************************************************

LjapV1_KE994228  EGSAFTIVMLLASLNSCTNPWIYMFFSGHLINDFAHCIPCCASSLRRLRTTGAGHNATSQ
Lamprey_Pm3133   EGSAFTIVMLLASLNSCTNPWIYMFFSGHLIYDFAHCIPCCAGALRRLETTGAGHNATDR
                 ****************************** **********.:****.*********.:

LjapV1_KE994228  RSSTPQRHRSITLSSCSQK
Lamprey_Pm3133   LSATPQRHRSVTLSSCSQK
                  *:*******:********
```

Fig. 2-3C

```
LjapV1_KE993677  MDLAPWRNASRTPRHDNISHAAPNDTDRHE---GTAAAARDERLAAAEIALLAAIVAVAI
Lamprey_Pm2017   MDLAPWRNASRTPRHHNVSHAAPNAADRPEDAAAAAAAAARDERLAAAEIALLAAIVAVAI
                 ***************.*:****** :** *   .:************************

LjapV1_KE993677  AGNGSVLLALSRTRCKASRMNLFVKHLSVADLAVAAFQVLPQLSWDVTFRFHGPDALCRL
Lamprey_Pm2017   AGNGSVLLALSRTRRKASRMNLFVKHLSVADLAVAAFQVLPQLSWDVTFRFRGPDALCRL
                 ************** ************************************:********

LjapV1_KE993677  VKYLQVVGMFASTYVLIAMTVDRYLAICHPLSTLRRRGEPRKQAHVMVACAWVLSALLST
Lamprey_Pm2017   VKYLQVVGMFASTYVLIAMTVDRYLAICHPLSTLRRRGEPRKQAHAMVACAWALSALLST
                 *******************************************.******.*******

LjapV1_KE993677  PQVVIFSLQEVEEGVFDCWADFGQPWGMRAYVTWITVSVYVAPVLILAACYGAILLEICR
Lamprey_Pm2017   PQVVIFSLREVEEGVFDCWADFGQPWGMRAYVTWITVSVYVAPVLILAACYGAILLEICR
                 ********:************************************************

LjapV1_KE993677  NLRLKTRGRYPKDPRGSQSQSPVCSRVSSVRCISRAKVRTIKMTLVIIVVYVVCWTPFFV
Lamprey_Pm2017   NLRLKTRGRYPKDPRGSQSQSPVCSRVSSVRCISRAKVRTIKMTLVIIVVYVVCWTPFFV
                 ************************************************************

LjapV1_KE993677  IQMWAAWDETAPDDDSSDPTYTIVMLLSSLNSCTNPWIYMSFSGHLLEEVLACCRQVRCP
Lamprey_Pm2017   IQMWAAWDETAPDDDSSDPTYTIVMLLSSLNSCTNPWIYMSFSGHLLEEVMSCCRQARCA
                 ************************************************::****.**.

LjapV1_KE993677  ATC-RQQ--RRRCRTPTRRGAGSPNTEH-VVYALSPAPTQCRSSPAAVA--TAT-PETRR
Lamprey_Pm2017   TACRRQQRRRRRCRTPTRRGAGSPNTEHVVVYALSPAPTQCRSSPATAAGGTATPPETRR
                 ::* ***  ****************** *****************:.*  *** *****

LjapV1_KE993677  SAAPL-SPALP
Lamprey_Pm2017   SAAPLWSPALP
                 ***** *****
```

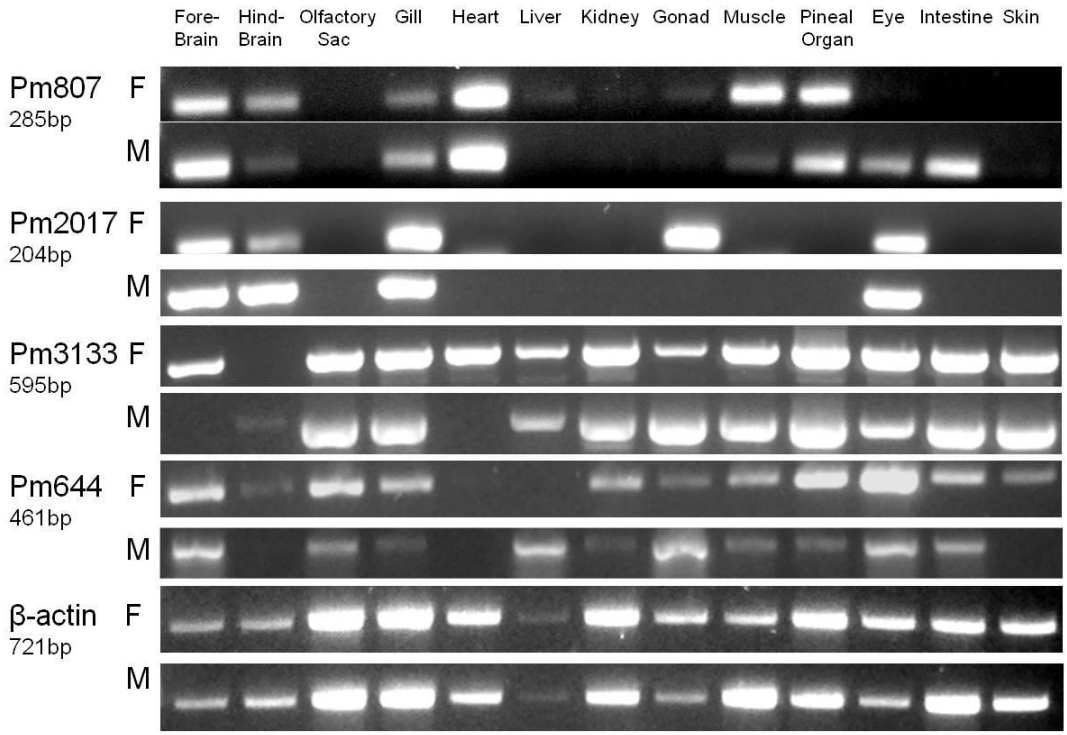
Fig. 2-3D

```
Lamprey_Pm644    MSDDPFVGNSSDESASTLASGAWPSAAPNGTAAPHPPLARRPERDEALARAEVTVLGLIF
LjapV2_KE993674  MSDDAFVGNSSAESASALASGAWSSAAPNGTAAP---LASRPERDEALARAEVTVLGLIF
                 ****.****** ****:******.**********   ** ********************

Lamprey_Pm644    VLATSGNATLLVSLWRRRKHASRMHAFLVHLSVADLVVAFFQVLPQLAWDVTDAFQASNG
LjapV2_KE993674  VLATSGNATLLVSLWRRRKHASRMHAFLVHLSVADLVVAFFQVLPQLAWDVTDAFQASNG
                 ************************************************************

Lamprey_Pm644    VCKVVKYLQVVGMFASAYMIVAMTVDRFQAVCYPMVTFKKKRSRWNGLVCAAWATSLALS
LjapV2_KE993674  VCKVVKYLQVVGMFASAYMIVAMTVDRFQAVCYPMVTFKKKRSRWNGLVCAAWATSLALS
                 ************************************************************

Lamprey_Pm644    TPQIFIFSLSEYEKGIFDCTATFAAHWGAKAYVTWVTLAVFVLPTLAVVICQAQICRIIR
LjapV2_KE993674  TPQIFIFSLSEYEKGIFDCTATFAAHWGAKAYVTWVTLAVFVLPTLAVVVCQAQICRIIR
                 ************************************************:***********

Lamprey_Pm644    LNLYVKTHQGPDEEETIGGDERRAGGQLMPSRASSVTGISRAMINTVRMTLVIVLVYVAC
LjapV2_KE993674  LNLYVKTHQGPDEEEAAGGDERRAGGQLMPSRASSVTGISRAMINTVRMTLVIVLVYVAC
                 ***************: ****************************************

Lamprey_Pm644    WAPWFTVQLWSAWDSKAPKEGPTFVIIMLLGNLNSCTNPWIYLWFAGGLPPGGCRGLVPH
LjapV2_KE993674  WAPWFTVQLWSAWDSKAPKEGPTFVIIMLLGNLNSCTNPWIYLWFAGGLSPGGCRGLVPH
                 ************************************************.***********

Lamprey_Pm644    ARPPAPDFNVDESVMSTATFREDRDDSLRGGDTPTPPFACLPPSQAQAASSLHAVSCLP
LjapV2_KE993674  ARPPAPDFNVDESVMSTATFREDRDDSLRGGDAPTPPLASLPPSQAQAASSPHAVSCLP
                 *********************************:****:*.*********** *******
```

Fig. 2-3E

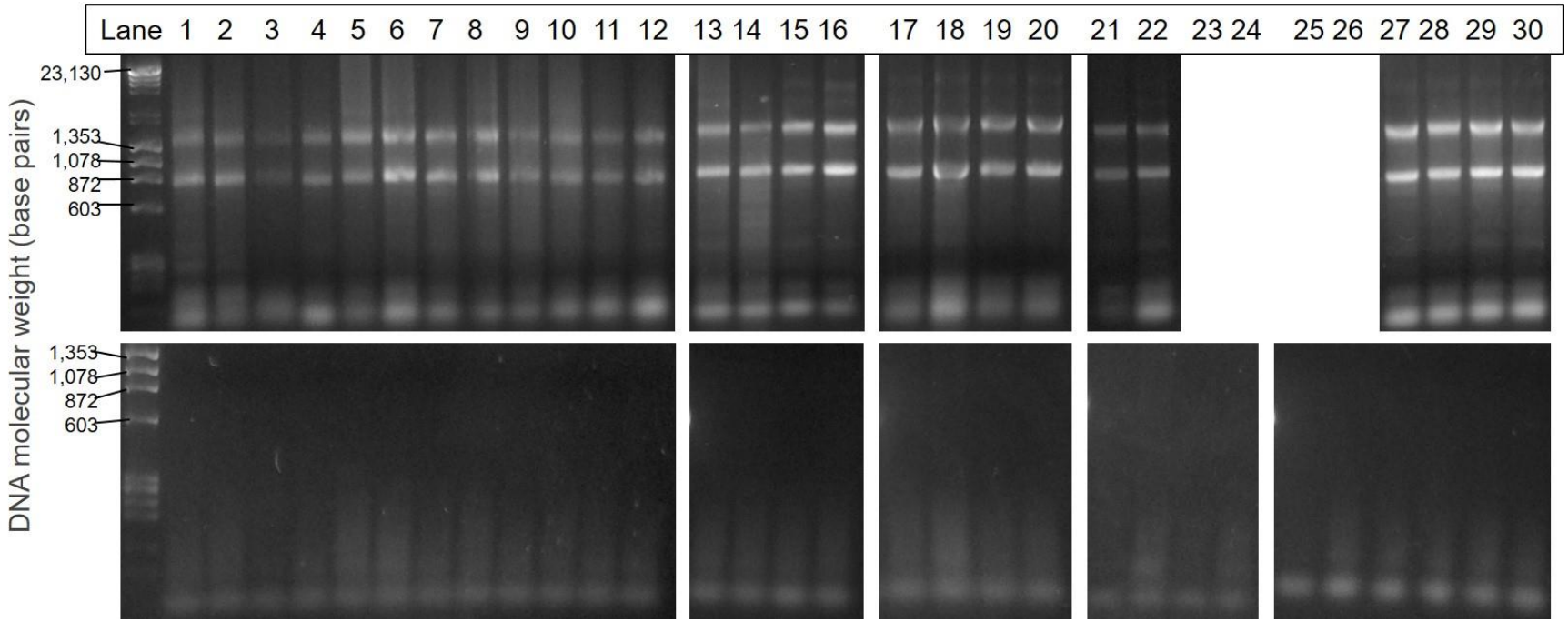
```
Lamprey_Pm4769   -TSAAPRPGANGSASRNERNEALARAEVAVLFAMFACTTLGNSVLLTALYRRRKHASRMH
LjapV2_KE993677  MNVSGTAARANGSASRGERNEALARAEVAVLFAMFACTTLGNSVLLTALFRRRKHASRMH
                  . :.. . *******.*********************************:**********

Lamprey_Pm4769   VFMVHLCLADLVVAFFQVLPQLAWDITDVFLASDFACRFIKYMQIVGMFASTYMIVGTTV
LjapV2_KE993677  VFMVHLCLADLVVAFFQVLPQLAWDITDVFLASDFACRFIKYMQIVGMFASTYMIVGMTV
                 ************************************************************ **

Lamprey_Pm4769   DRYQAVCFPMVTFKKRMAYWNAPVCAVWVVSLLLSVPQAVIFAKGEVYPGVYDCWGTFQP
LjapV2_KE993677  DRYQAVCFPMVTFKKRMAYWNAPVCAVWVVSLLLSVPQAVIFAKGEVYPGVYDCWGTFQP
                 ************************************************************

Lamprey_Pm4769   MWGVKAYITWMAMAIFVVPTAVLVWCQTKICRVIRLNIYFKTPQVHKDKLLLRRRKQRRR
LjapV2_KE993677  MWGVKAYITWMAMAIFVVPTAVLVWCQTKICRVIRLNIYFKTPHVHKDTLLLRRRRKQRR
                 *******************************************:****.******:::**

Lamprey_Pm4769   RQQKRQEDEEEVAALHRGQTHAVTMTRVSSVTGVSRAMVKTVKMTLVIVIVYVLCWSPFF
LjapV2_KE993677  RQQKRQEDEEEVAALHRGQTHAVTMTRVSSVTGVSRAMVKTVKMTLVIVIVYVLCWSPFF
                 ************************************************************

Lamprey_Pm4769   VAQLWSVWDPNPPFEGALFTITLLLASLNSCTNPWIYMAFSGSTPRALLSCVLCRARRGG
LjapV2_KE993677  VAQLWSVWDPNPPFEGALFTITLLLASLNSCTNPWIYMAFSGSTPRALLSCVLCRARRGG
                 ************************************************************

Lamprey_Pm4769   AGGIAGGGVGSDDGIGGGGG-TGGWVRSVPGATAHEDSMATSSVQLAGALRSQHEDDAEQ
LjapV2_KE993677  AGGIGGGGVGSDGGIGGGGGGTGGWVRSVPGATAHEDSMATSSVHLTGALRSQHEDDAEQ
                 ****.*******.******* ***********************:*:*************

Lamprey_Pm4769   TNELIRSRTISA
LjapV2_KE993677  TNELISSRTIS-
                 ***** *****
```

A



Supplementary Figure 2-4. Preliminary (2012) AVT receptor mRNA expression and 2013 RNA quality and negative controls. **(A)** 2012 adult female (F) and male (M) lamprey tissues after one round (Pm807 and β-actin) or nested PCR (Pm2017, Pm3133, Pm644). Products were visualized on 2% agarose gels with SYBR-safe® staining. Amplicon sizes are noted in base pairs (bp).

B



Supplementary Figure 2-4. (B) RNA quality (top) was tested with 8µl RNA on 2% agarose gels showing clear ribosomal bands. Insufficient pineal organ and hypothalamus RNA was available to run the RNA quality check. Negative control PCR (bottom) was conducted using actin primers (721 bp) on non-reverse-transcribed template to demonstrate the absence of genomic DNA. Lanes are as follows: Odd lanes female, even lanes male; 1-2 Liver, 3-4 Muscle, 5-6 Gill, 7-8 Eye, 9-10 Kidney, 11-12 Heart, 13-14 Intestine, 15-16 Skin, 17-18 Olfactory sac, 19-20 Cloaca, 21-22 Gonad, 23-24 Pineal organ, 25-26 Hypothalamus, 27-28 Forebrain, 29-30 Hindbrain.

Supplementary Figure 2-5. ClustalW multiple sequence alignment for AVP/OXT receptor family phylogenetic tree construction (including sea and Japanese lamprey AVT receptors). See Table 2-3 for abbreviations.

```
Cmi_V1B         ------------------------------------------------------------
Lja_2017        ------------------------------------------------------------
Pm2017          ------------------------------------------------------------
Cin_VPR         ------------------------------------------------------------
Ovu_CTR2*       ------------------------------------------------------------
Ovu_CTR1*       ------------------------------------------------------------
Ovu_OPR*        ------------------------------------------------------------


Aoc_V2A         FSSSFVSELNTFNTSHSGGSFFGLFPENGSSTTPFT-----LPQPRVRDLGLARAEIAVL
Ola_V2A2        SSSFFVSDLNSFNGSHSGGSLFGIFAENGSNTTPHA-----LPQPRIRDLTLARAEIAVL
Cne_V2A         NFSVFVTELSSFNASRTGGSIFGIFQENGSNTTPHS-----LPQPRTRDQGLARAEIAVL
Xma_V2A1        NISSFVTELNSFNISRGGGSLFGILIENGSNATSHT-----VPQPRARDQGLARAEIAIL
Dre_V2A2        FTSSLFSDHTGLNSTFRGGSYFWLVSQNSSINPTQK-----PGPVKIRDVALAQAEIGIL
Dre_V2A1        PPDLSSSLFNLSSSYGGTGAFLWLYPSENVTWTTMRPTMQPVTPPRVRDQALAQAEIGVL
Ola_V2A1        ATQRYPSVNGTAIPETPVGSNATIFER---------------------DALLAVAEVVVL
Cpy_V2A         NQSGTYTTFQAPTALQNASNTTIMSDDRD-------------------DNLAKVQIAIL
Tgr_V2A         NQSGTYPTFQTPTTLQNASNTTIMLDDRD-------------------DNLAKVQIAIL
Xtr_V2A         -MSGTQPTFETPNITVGVE-----EDKRD-------------------PIVAQWNIAML
Lch_V2A         NESWNTTMASPCGQRVEQSSYNSSTVDRN-------------------NELAKVEIAIL
Pan_V2A         QRNNSCFETVISTGTFRNGSNTSLLPERD-------------------LELAKVEVAVL
Hsa_V2A         STTSAVPGHPSLPSLPSNSSQERPLDTRD-------------------PLLARAELALL
Mmu_V2A         STTSAVPGALSSPSSPSNSSQEELLDDRD-------------------PLLVRAELALL
Aca_V2A         DKTQGLNLSVTLAGPTASNASLLLSDDRD-------------------VALAQAEIAIL
Aoc_V2B         -----------MRAVNDSLFVDAEDADRD-------------------ESLAKVEIALL
Xma_V2B1        -----------MRPLNESDFVDAYDVPRD-------------------ESLARVEIALL
Gac_V2B1        -----------MCPLNGS--DFAEDVARD-------------------ESLAKLEIALL
Dre_V2B1        ----MPFSFTLNLSNTSVEMDTSADEPRN-------------------EQLAQIEIALL
Xma_V2B2        -----MAWSSANTSNASSEAVPAGDEPRD-------------------ERLARLEIALL
Cmi_V2B1        NSGFSDAMANGFPKNFNFSSKETLPNEKD-------------------YWLAQIEIAVL
Lch_V2C         RSHLEELVFYGNANSSLKERSQNQLFERD-------------------NKLAMIEIAVL
Gga_V2C         DSTHQTESPPHRLLSLTNKSDPVGRPERD-------------------EQLAQVEIAVL
Xtr_V2C         ---------------MTNKSQFIERPDRS-------------------EQLAYVEISVL
Lpl_V2C         SDILSDNIFQHGFTTTANNSGISSRPERD-------------------QQLAWIEIAVL
Cmi_V2C         LTEREMLLKHSRMGATTNRTGEAEINLGK-------------------PTRDEELAKTV
Lja_644         ALASGAWSSAAPNGTAAP---LASRPERD-------------------EALARAEVTVL
Pm644           TLASGAWPSAAPNGTAAPHPPLARRPERD-------------------EALARAEVTVL
Lja_4769        ---------MNVSGTAARANGSASRGERN-------------------EALARAEVAVL
Pm4769          ------------------------------------------------------------VL
Gac_V2C         ------------MGSRSNSSEKATTEEGDV------------------YGRFALVKAGIL
Xma_V2C         -------------MDPSSSSRPAVAADEDE------------------AVRFALLRAGVL
Dre_V2C         --------MADNLNCSAGNCSSAHDAFGLS------------------DGHFALVKAAVL
Ola_OTR         ----------MEIISNESEIWQFNGSWRNSSLGNGTGALNQTN-PLKRNEEVAKVEVTVL
Xma_ITR1        ----------METISNESDIWQLNESWRNSSLLNGTGGLNQTN-PLERNEEVAKVEVTVL
Aoc_ITR         -------------------------MESTSSLINGTGGLNQTN-PLKRNEEVAKVEVTVL
Dre_ITR         ----------MESLLKDVILWPINDSWANSSRGNDSRGLNQTVNPLKRNEEVAKVEVTVL
Lpl_OTR         ---------MEDILFKDQDLWSVNASLENSSLANATREANGTVNPVLRNEEVAKVEVTVL
Cne_ITR         ----------MDELLSAQDAWLQNFSYCNYSHLNKTHPGNNVVNPLKRNEEVAKVEVTVL
Xma_ITR2        ----------MEELLRTQGAWAQNGSWYNSSYLNITNIGNNTVNPLKRNEEVAKVEVTVL
Tgr_MTR         -------MEGLCLKENDCWGLLANRSLDNSTMENRTYAVNKTRDPLKRNEHMARVEVSVL
Xtr_MTR         -------MEPLCAQP-DCPLLLLNGSWGNFSMENRTGPTNTTRDPLKRNEDVAKVEVAVL
Aca_MTR         -------MDPLCFLGNDPWTYNCS--FNNYTLENQTSQTNQT-FPIKRNEDVAKVEVAVL
Lch_OTR         ---------MEGMYMTNTEVWLLNSSWGNASFENGTRNRNG-TDPLVRNEEVAKVEVAVL
Pan_MTR         ----------------------MNSSLENSSLENKTGGRNTSGNPLLRNEDAAKVEVAVL
Hsa_OTR         --------------MEGALAANWSAEAANASAAPPGAEGNRTAGPPRRNEALARVEVAVL
Mmu_OTR         --------------MEGTPAANWSIELDLGSGVPPGAEGNLTAGPPRRNEALARVEVAVL
Gga_MTR         --------MEKLYLAGSGLWANGSLGNGSLQPEDRAAARNGTADPLKRNEDMAKVEVTVL
Cmi_OTR         ------MNKTCCNNVDMDLCVPNASANGSALLENWSRFCNSSADLLKRNETVAKVEVTLL
Lja_807         GDILNGTGVAYAPAASNGSHAAPAEAFTTALASINATLNGSSQNALDRNEEVAKVEIALL
Pm807           GAILNGTGVAYAPATSNGSHVVPAEAFTAALASINATLNGSSQHALDRNEEVAKVEIALL
Lja_3133        --------------------------------MANGTANASESTLSRDEELAKVEIAIL
Pm3133          --------------------------------MANGTANASESTLSRDEDLAKVEITIL
Aoc_V1A         ----------------------------MGTSGNNTVHPNGSDPFGRNEEVAKFEIMVL
Ola_V1A2        ---MYTLSSVLLLSGENCSLDSSLSSDLMMGTSGNGTANSNGSDPFGRNEEVAQIEIMVL
Cne_V1A2        --------------------------MMGTPSNSTVHPNGSDPFARNEDVAQIEIMVL
Xma_V1A         -MRNPSDAPLLSGGNQSLLASSLSGDPTMGTPSNGTPLHANGSDPFARNEDVAQIEIMVL
Dre_V1A         ----------------------------METHSNTSHPPNTSDPFGRDEEVAKIEIAVL
Lpl_V1A         PVNVLFYSGAMRNFTDENQSLGLSATDGSMAGAGNASSSSNASDPFGRNEDVAKIEITVL
Cne_V1A1        ----MLCSSEISCNITEVFNLTQKQEIGAAGSARNLSGKTNDTDPFGRNEEVAKIEITVL
Ola_V1A1        ----MLFPSDSLCNLTGQNCSLNVTQQLGG-------VKKNHSDPFGRNEEVAKIEITVL
Cpy_V1A         PATPWPSSDPLLTRWNSSLTGAPSSPGLASSSLPPTSPSTNSSDLLDRNEALARVEVAVL
Tgr_V1A         PASPWPSSDPLPTRWNSSLTGGPSCPGLASSSLPPTSPSTNSSDLLGRDEALARVEVAVL
Hsa_V1A         SAGPDAGPSGNSSPWWPLATGAGNTSREAEALGEGNGP-----PRDVRNEELAKLEIAVL
Mmu_V1A         PRGSHDLPAGNSSPWWPLTTEGANSSREAAGLGEGGSP-----PGDVRNEELAKLEVTVL
Gga_V1A         GSPRAAGPPGNGSRWRGAAEDGSSPSPEAWSGAPNGSAGD--WDPFGRDEELAKLEIAVL
Pan_V1A         --------------MWNCSLEIGTSACGLLRIKENASSASNTTDLYGRDEELAKIEITVL
Xtr_V1AR        ---MDLTHSSESQQDSNLAAPTEDCSPNSTLFNSSSSQIDPTADPYVRNEALAKIEIAIL
Cmi_V1A         -----MGVTETSGWDWGIWNFTRAGVSLSRLQNESFTKSRNKTDPFGRNENLAKTEISVL
```

```
Lch_V1A         -------MKLYTGPEWNSTKETLKIELNATVAPPADKNYSNKTNSYGRNEELAKVEIAVL
Aca_V1B         ------------MEKNSTYFNSFTDSSLLDNTDPRNEATNQTALYTRDEELAKAEIGVL
Gga_V1B         ---------MEPGWSWNSTHSSQTTHSTELQSSQPPAGTPNRTLLQGRDEQLARAEVGVL
Cpy_V1B         --------MDLDKYLYRNQTHFLRMDAVTIIAATNASHQSSGEQDDPRDETLAKAEIAVL
Tgr_V1B         --------MDLGKDLHRNQTQFLRMDDVSTIAATNASHQSSGEHDDPRDATLAKIEIAVL
Lch_V1B         ---------------------------NQKPDIHSYPYLGNKTFLLTRDESLAKVEVGIL
Pan_V1B         DASLTGNCTVADYYQVSCRVAGILNQSTSAPDGRSISCFHNESSLLPRNENLAKVEIAVL
Hsa_V1B         -------------------MDSGPLWDANPTPRGTLSAPNATTPWLGRDEELAKVEIGVL
Mmu_V1B         -------------------MDSEPSWTATPSPGGTLFVPNTTTPWLGRDEELAKVEIGIL
Cmi_V1B         ----------MLTGQNCSWSSGFGADRQWPNQTAAKPTQRSADFLLSRDEELARAEISVL
Lja_2017        --------MDLAPWRNASRTPRHDNISHAAPNDTDR---HEGTAAAARDERLAAAEIALL
Pm2017          --------MDLAPWRNASRTPRHHNVSHAAPNAADRPEDAAAAAAAARDERLAAAEIALL
Cin_VPR         --------------MTSSIPLLAMHRVRNG---------------TALQRQIRITELYII
Ovu_CTR2*       -------MYQAMEVESTSPSGFFLDFYTQSTIPTTDFLNNTNSSHPIRDEKLVKIEIAVL
Ovu_CTR1*       --------MRYITTHPNEISTQIWNNFSSTEIWSNFSAAKNETQPIRRNQDLANAEVITL
Ovu_OPR*        -------------------MENFTEENLHPWITTTTRVYNNVTIFPQYDDELGKFEIMVL

Aoc_V2A         GLVLALTTLGNSFVLWVLLRR--RKHNAPMHVFMVNLCIADLVVALFQVLPQLIWDITER
Ola_V2A2        GIVLALTTLGNSFVLWVLLRR--RKHNAPMHLFMVNLCVADLVVALFQVLPQLIWDITER
Cne_V2A         GMVLALTTLGNSFVLWVLLRR--RKHNAPMHLFMVNLCVADLVVALFQVLPQLIWDITER
Xma_V2A1        GVVLALTTLGNSFVLWVLLRR--RKHNAPMHLFMVNLCVADLVVALFQVLPQLIWDITER
Dre_V2A2        GLVLALTTLGNSFILWVLLRR--RKYNAPMHQFMVNLCVADLVVGFFQVLPQMLWDITER
Dre_V2A1        GLVLALTALGNGFVLWVLLRR--KKNHAPMHLFMVNLCVADLVVAFFQVLPQLVWDITER
Ola_V2A1        AVILVMALLSNGMVLMVLLRR--GKRHSPLHQFMLNLCVADLVVALFQVLPQLVWDAKGR
Cpy_V2A         AILFVCTTLSNVLILFVLLKR--RKHNALMHTFMINLCIADLVVAFFQVLPQLVWVITDR
Tgr_V2A         AILFVCATLSNFLILFVLLKR--RKHNALMHTFMINLCIADLVVAFFQVLPQLVWVVTDR
Xtr_V2A         AVIFSFATFGNCLVLFTLLRR--RKHNALMHTFMIHLCLADLVVSFFQVLPQLVWDITDR
Lch_V2A         VIIFACATLSNSVVLVTLFRH--RKQNALMHIFMVNLCIADLMVAFFQVLPQLMWDITDR
Pan_V2A         AAVFVLATFSNLVVLYVLIKR--RKYNTPMHAFMTNLCIADLVVAFFQVLPQLLWDVTDQ
Hsa_V2A         SIVFVAVALSNGLVLAALARRGRRGHWAPIHVFIGHLCLADLAVALFQVLPQLAWKATDR
Mmu_V2A         STIFVAVALSNGLVLGALIRRGRRGRWAPMHVFISHLCLADLAVALFQVLPQLAWDATDR
Aca_V2A         ASIFLLATLSNGLVLGALFLRGHRAAPTPMHRFILHLCLADLTVALFQVLPQLIWDITDR
Aoc_V2B         SAIFVSAAILNTALLLILWRQ--RKQMTRMRIFVFHLCLADLVVAFFQVCPQLMWDITDR
Xma_V2B1        GAIFVTATILNTALLLLLWRQ--RKQMSRMRVFVFHLCLADLVVAFFQVCPQLMWDITDR
Gac_V2B1        SVVFIGAAILNTSLLLVLWGQ--RKHMSRMRVFVFHLCLADLVVAFFQVCPQLIWDITDR
Dre_V2B1        SVIFLCASTLNFSLLLVLWRK--RQQMSRMRVFVTHLCLADLVVAFFQVCPQLMWDITDR
Xma_V2B2        SIIFVAAGLLNFGLLLALWKR--RKQLSRMRVFVLHLCVADLVVAFFQVCPQLMWDITDR
Cmi_V2B1        GAIFALALVTNSILLLLLWRR--HKQVSRMHVFVIHLCLADLVVALFQVFPQLFWDITDR
Lch_V2C         SLIFFCASATNSILLMTLWKR--RKQVSRMHVFVLHLCLADLVVAFFQVFPQLFWDITER
Gga_V2C         GVIFLTASVGNFILILVLWRR--RKKLSRMYVFMLHLSIADLVVAFFQVLPQLIWDITDV
Xtr_V2C         GVIFMVASAGNLTLILVLWNK--RKKLSRMYVFMLHLSLADLVVAFFQVLPQMIWDITDV
Lpl_V2C         GIIFFTASIGNSVLVMVLWRK--RKRMSRMYVFMMHLSIADLVVAGFQVLPQLIWDITDI
Cmi_V2C         SLLFLTAAVGNCLLIRSLWDR--RK---CMYAFLMHLSIADLVVAVFQVLPQLVWDITEI
Lja_644         GLIFVLATSGNATLLVSLWRR--RKHASRMHAFLVHLSVADLVVAFFQVLPQLAWDVTDA
Pm644           GLIFVLATSGNATLLVSLWRR--RKHASRMHAFLVHLSVADLVVAFFQVLPQLAWDVTDA
Lja_4769        FAMFACTTLGNSVLLTALFRR--RKHASRMHVFMVHLCLADLVVAFFQVLPQLAWDITDV
Pm4769          FAMFACTTLGNSVLLTALYRR--RKHASRMHVFMVHLCLADLVVAFFQVLPQLAWDITDV
Gac_V2C         GLVFLLATCGNIFFLCILWKR--RKRNTKTQLFLLHLCLADLVVAFFQVLPQFLMEITHR
Xma_V2C         GLIFVFATCGNLFFLVTLCKR--RKRNSRTQLFLLHLCLADLVVAFFQVLPQLSMEITHR
Dre_V2C         GCVFVLATCSNLFLLHALWKR--RKRHTRTQLFLLHLCLADLVVAFFQVLPQLSMEITHR
Ola_OTR         ALVLFLALAGNLCVLLAIHTT--KHSQSRMYYFMKHLSIADLVVAIFQVLPQLIWDITFR
Xma_ITR1        ALVLFLALAGNLCVLLAIHTT--KHSQSRMYYFMKHLSIADLVVAIFQVLPQLIWDITFR
Aoc_ITR         ALVLFLALAGNLCVLLAIHTT--KHSQSRMYYFMKHLSIADLVVAIFQVLPQLIWDITFR
Dre_ITR         VLILLLALAGNLCVLVAIQTS--KHGQSRMYYFMKHLSIADLVVAVFQVLPQLIWDITFR
Lpl_OTR         ALILLLALTGNLCVLLAIYTT--KNNQSRMYFFMKHLSIADLVVAIFQVLPQLIWDITFR
Cne_ITR         VLVLLLALTGNLCVLWAIHTT--KHSKSRMYYFMKHLSIADLVVAVFQVLPQLIWDITFR
Xma_ITR2        VLVLLLALTGNLCVLCAIHTT--KHSHSRMYYFMKHLSIADLVVAVFQVLPQLIWDITFR
Tgr_MTR         ALILFLALVGNVCVLLAIHIN--RNKHSRMYFFMKHLSIADLVVAIFQVLPQLIWDITFR
Xtr_MTR         ALILFLALAGNICVLIAIHIN--RHKHSRMYFFMKHLSIADLVVALFQVLPQLIWDITFR
Aca_MTR         CLIFFLALTGNLCVLMAIYTT--RHKHSRMYFFMKHLSIADLVVAIFQVLPQLIWDITFR
Lch_OTR         SLILFLALTGNICVLLAIHTT--SHKKSRMYFFMKHLSIADLVVAIFQVLPQLIWDITFR
Pan_MTR         SVILFFALTGNICVLLAIHTS--TNKQPRMYYFMKHLSIADLVVAIFQVLPQLIWDITFR
Hsa_OTR         CLILLLALSGNACVLLALRTT--RQKHSRLFFFMKHLSIADLVVAVFQVLPQLLWDITFR
Mmu_OTR         CLILFLALSGNACVLLALRTT--RHKHSRLFFFMKHLSIADLVVAVFQVLPQLLWDITFR
Gga_MTR         CLILFLALSGNLCVLLAIHTT--RQKHSRMYFFMKHLSIADLVVAVFQVLPQLIWDITFR
Cmi_OTR         AVILILALTGNIAVLLAIHSN--RHKPSRMYFFVKHLSIADLVVAVFQVLPQLIWEITFR
Lja_807         SIILFVAIVGNVCVLLALINT--RKKTSRMHLFIMHLSIADLVVAFFQVLPQLIWKITYR
Pm807           SIILFVAIVGNVCVLLALINT--RKKTSRMHLFIMHLSIADLVVAFFQVLPQLIWKITYR
Lja_3133        AIILCVAVVGNACVLLALLST--KKKTSRMHLFIMHLSIADLVVAFFQVLPQLIWEVTFR
Pm3133          AIILCAAVVGNACVLLALLST--KKKTSRMHLFIMHLSIADLVVAFFQVLPQLIWEVTFR
Aoc_V1A         SITFVVAVIGNVSVLLAMHNT--KKKMSRMHLFIKHLSLADLVVAFFQVLPQLCWEITYR
Ola_V1A2        SITFVVAVIGNVSVLLAMYNT--KKKMSRMHLFIKHLSLADLVVAFFQVLPQLCWEITFR
Cne_V1A2        SITLVVAVIGNVSVLLAMYNT--KKKMSRMHLFIKHLSLADLVVAFFQVLPQLCWKITDR
Xma_V1A         SITFVVAVIGNVSVLLAMYHT--KKKTSRMHLFIKHLSLADLVVAFFQVLPQLCWEITFR
Dre_V1A         SVTFAVAVIGNYSVLVAIHNT--KKKTSRMHLFIKHLSLADLVVAFFQVLPQLCWEITFR
Lpl_V1A         SITFVVAVIGNLSVLLALYRT--KKKMSRMHLFIKHLSLADLVVAFFQVLPQLCWEITFR
Cne_V1A1        SLAFVAAVVGNLSVLLAMYRS--RRKLSRMHLFMKHLSLADLVVAFFQVLPQLCWEVTFR
```

```
Ola_V1A1        SLAFVAAVVGNVSVLLAMHRT--RRKLSRMHLFMKHLSLADLVVAFFQVLPQLCWEITFR
Cpy_V1A         ALIFVAAVVGNSCVLLALQRS--RRKSSRMHLFILHLSLADLVVALFQVLPQLCWEVTYR
Tgr_V1A         ALIFVAAVVGNGCVLLALQRS--RRKTSRMHLFIRHLSLADLVVALFQVLPQLCWEVTYR
Hsa_V1A         AVTFAVAVLGNSSVLLALHRT--PRKTSRMHLFIRHLSLADLAVAFFQVLPQMCWDITYR
Mmu_V1A         AVIFVVAVLGNSSVLLALHRT--PRKTSRMHLFIRHLSLADLAVAFFQVLPQLCWDITYR
Gga_V1A         AVTFAVAVLGNGSVLLALRRT--PRKASRMHLFIRHLSLADLVVAFFQVLPQLCWEVTHR
Pan_V1A         AVIFLVAVIGNISVLLALYKS--KKKMSRMHLFIKHLSLADLVVAFFQVLPQLCWEVTYR
Xtr_V1AR        AIIFVAAVLGNCSVLIGLYKS--KKKTSRMHLFIKHLSLADLAVAFFQVLPQLCWEVTYR
Cmi_V1A         GIIFLVAVIGNLSVLMALYKT--KKKMSRMHLFIKHLSVADLVVAVFQVLPQFIWDITYR
Lch_V1A         AAIFAMAVIGNSSVLLALYKT--KKKASRMHFFIKHLSLVDLAVAFFQLLPQLCWDITGR
Aca_V1B         AAILAVTLVGNLGVLLTMYHL--RKKMSRMHLFILHLALTDLVVALFQVFPQMIWEVTYR
Gga_V1B         AAILLVATTGNLAVLLAVCRR--GRKLSRMHLFVLHLALSDLGVALFQVLPQMLWEVTYR
Cpy_V1B         AVILAVTTVGNLVVLFTMYRI--RRKMSRMHLFIMHLGLTDLVVAGFQVLPQMIWDITFR
Tgr_V1B         AVILAVTTVGNLVVLFAMYRI--RRKMTRMHLFIMHLGLTDLVVAGFQVLPQMIWDITFR
Lch_V1B         AAILVMATVGNLGVLLAMYRI--RRKMTRMHLFITHLGLTDLVVALFQVLPQLIWKVTYR
Pan_V1B         ATIFVVAMASNLGVLLAMYRM--RKKMTRMHIFIMHLGFTDLVVALFQVLPQMIWEVTYR
Hsa_V1B         ATVLVLATGGNLAVLLTLGQL--GRKRSRMHLFVLHLALTDLAVALFQVLPQLLWDITYR
Mmu_V1B         ATVLVLATGGNLAVLLILGLQ--GHKRSRMHLFVLHLALTDLGVALFQVLPQLLWDITYR
Cmi_V1B         AAVLIVAVAGNLAVLTALCRM--KRKLSRTHLFVLHLALTDLCVALFQVLPQLCWEITYR
Lja_2017        AAIVAVAIAGNGSVLLALSRT--RCKASRMNLFVKHLSVADLAVAAFQVLPQLSWDVTFR
Pm2017          AAIVAVAIAGNGSVLLALSRT--RRKASRMNLFVKHLSVADLAVAAFQVLPQLSWDVTFR
Cin_VPR         ATVFILGLIGNSCVLVALWQR--YSKHTRMHILIFHLAVADLIVVLFEMLPEWIIRFGFG
Ovu_CTR2*       GTCFTLAIINNLCVLLVLLWR--RKKVRRMQMFILHLSIADLIVAFFNILPQLIWDITFR
Ovu_CTR1*       AVVIIITVIGNSIVLITLFQR--RKKLTRMHLFILHLSVTDLFVAFFNNLPQMIWDITFL
Ovu_OPR*        CILCFMALFGNAVVLIVLRIK--KTTLTRMQLLIVYLSVTDISVALFHILPTIILKINVY

Aoc_V2A         FQGPDFLCRSIKYLQIVGMFASSYMIVAMTVDRHHAICCPLQAYRG--GAMSRWNTPVMV
Ola_V2A2        FQGPDFLCRSVKYLQIVGMFASSYMIVAMTVDRHRAICCPLQAYRV--GAMSRWNTPVMV
Cne_V2A         FQGPDLLCRSVKYLQIVGMFASSYMIVAMTVDRHHAICCPLQAYRV--GAMSRWNTPVMV
Xma_V2A1        FQGPDLLCRSVKYLQIVGMFASSYMIVAMTVDRHHAICCPLQAYRV--GAMSRWNTPVIV
Dre_V2A2        FQGPDVLCRSVKYLQIVGMFASSYMIVAMTVDRHNAICCPLQAYKG--GAVSRWNTPIMV
Dre_V2A1        FHGPDALCRSVKYLQIVGMFASSYMIVAMTVDRHYAICCPLQAYRG--GATSRWNTPIMV
Ola_V2A1        FPGPDFLCRLVKYLQVLGMFASSYMIVAMTVDRHYAICCPLQAHRS--WATKRWNTPIAV
Cpy_V2A         FRGPDILCRLIRYLQVVGMFASSYMIVAMTFDRHQAICRPMMTYKK--GVA-RWNVPVFV
Tgr_V2A         FRGPDILCRLIKYLQVVGMFASSYMIVAMTFDRHQAICRPMMTYKK--GMA-RWNVPVFV
Xtr_V2A         FRGPDVLCRGVRYLQVVGMYASSYMIVAMTFDRHQAICRPMMTFKK--GSA-RWNIPVSL
Lch_V2A         FHGPDFLCKAVKYLQVVGMFASSYMIMAMTFDRHQAICQPMGSFHK--GGS-RWNIPVAV
Pan_V2A         FLGPDLLCRAVKYFQVVGMFASSYMIVAMTFDRHQAICRPMMTYRK--GIA-RWNIPVIV
Hsa_V2A         FRGPDALCRAVKYLQMVGMYASSYMILAMTLDRHRAICRPMLAYRH--GSGAHWNRPVLV
Mmu_V2A         FHGPDALCRAVKYLQMVGMYASSYMILAMTLDRHRAICRPMLAYRH--GGGARWNRPVLV
Aca_V2A         FQGPDILCRAITYLQVVGMFASSYVIVAMTYDRHRAICQPMLVFRR--GP-GTWYRPVMV
Aoc_V2B         FVGPDLVCRVVKYLQVLGMFSSTYMIVVMTVDRYQAVCNPMVKFQR--TRT-RLSIPVCV
Xma_V2B1        FVGPDLVCRLVKYLQVLGMFSSTYMIVVMTVDRYQAICNPMVKFQR--ARS-RLNVPVCV
Gac_V2B1        FVGPDPVCRLVKYLQVLGMFSSTYMIVAMTVDRYQAVCNPMVKFQR--SHT-GLNVPVCS
Dre_V2B1        FVGPDLVCRLVKYLQVVGMFASTYMIVVMTVDRYQAICNPMVTFQR--RRA-RWNLPVCV
Xma_V2B2        FVGPDILCRLVKYLQVVGMFASTYMIVVMTIDRYQAICNPMVTFQR--RRA-RWNGPVCA
Cmi_V2B1        FVGPDAICRLVKYLQIVGMLASTYMIVVMTIDRYQAICNPMVTFQR--KRA-RWNLPVCI
Lch_V2C         FIGPDIVCRLIKYLQIVGMFASTYMIVVMTLDRYQAICNPMVTFQR--RRT-RWNIPICI
Gga_V2C         FIGPDFLCRIIKYLQLLGMFASTYMIVVMTVDRYQAVCYPMVTFQK--KRA-LWNIPICT
Xtr_V2C         FFGPDPMCRLIRYLQLVGMFASTYMIVVMTLDRFQAICYPMVTFQK--KRA-LWNAPICA
Lpl_V2C         FLGPDFLCRVVKYLQLVGMFASTYMTVAMTVDRFQAVCYPMVTFQK--KRA-FWNAAICS
Cmi_V2C         FLGLDTLCRAVTYLQLVGMLASSYTLLVMIMDRFQAVCYPMVTFQK--RRT-FWNGATCT
Lja_644         FQASNGVCKVVKYLQVVGMFASAYMIVAMTVDRFQAVCYPMVTFKK--KRS-RWNGLVCA
Pm644           FQASNGVCKVVKYLQVVGMFASAYMIVAMTVDRFQAVCYPMVTFKK--KRS-RWNGLVCA
Lja_4769        FLASDFACRFIKYMQIVGMFASTYMIVGMTVDRYQAVCFPMVTFKK--RMA-YWNAPVCA
Pm4769          FLASDFACRFIKYMQIVGMFASTYMIVGMTVDRYQAVCFPMVTFKK--RMA-YWNAPVCA
Gac_V2C         FRGTDFLCRLVKYLQVVGMFASTYMIVAMTIDRYHAVCRPMVSFLK--GSFRRYLS-IGA
Xma_V2C         FRGTDFLCRLVKYLQVVGMFASAYMIVAMAIDRYHAVCKPMVSFLE--GSFRRYIS-IGA
Dre_V2C         FRGSDLVCRSVKYLQVVGMFASSYMIVAMTIDRYHAVCKPMVSFFR--GSFRRYVS-ISA
Ola_OTR         FYGPDILCRLVKYLQVVGMFASTYMLVLMSIDRCLAICQPLHSLHR---RKDR--IYVIL
Xma_ITR1        FYGPDILCRLVKYLQVVGMFASTYMLVLMSVDRCLAICQPLRSLHR---RKDR--FYVIF
Aoc_ITR         FYGPDILCRLVKYLQVVGMFASTYMLVLMSVDRCLAICQPLRSLHR---RKDR--FYVLI
Dre_ITR         FYGPDFLCRLVKYLQVVGMFASTYMLVLMSIDRCLAIWQPLRSLRR---RKDR--FYVLA
Lpl_OTR         FYGPDFLCRLVKYLQVVGMFASTYMLVLMSIDRCLAICQPLRSLHR---RADR--FYIIT
Cne_ITR         FYGPDIICRLVKYLQVVGMFASTYMLVLMSIDRCLAVCQPLRSVHK---GKDR--FCVIG
Xma_ITR2        FYGPDFLCRLVKYLQVVGMFASTYMLVLMSIDRCLAICQPLRSVHK---GKDR--FCVIA
Tgr_MTR         FYGPDFLCRAVKYLQVVGMFASTYMLLLMSMDRCLAICQPLRSLHR---RSDR--VYVIV
Xtr_MTR         FYAPDFVCRIVKYLQVVGMFASTYMLLLMSLDRCLAICQPLRSLHR---RSDC--VYVVL
Aca_MTR         FYGPDYLCRLVKYLQVVGMFASTYMLLLMSLDRCLAICQPLRSLHR---RSDR--LSVLL
Lch_OTR         FYGPDFLCRLVKYLQVVGMFASTYMLLLMSIDRCVAICQPLRSLHQ---RSDR--VYVFF
Pan_MTR         FYAPDILCRLVKYLQVVGMFASTCMLLLMTIDRCLAICQPLRSLHR---RSDR--VSVIV
Hsa_OTR         FYGPDLLCRLVKYLQVVGMFASTYLLLLMSLDRCLAICQPLRSLRR---RTDR--LAVLA
Mmu_OTR         FYGPDLLCRLVKYLQVVGMFASTYLLLLMSLDRCLAICQPLRSLRR---RTDR--LAVLA
Gga_MTR         FYGPDFLCRLIKYLQVVGMFASTYMLLLMSLDRCLAICQPLRSLHR---RADR--VSVLL
Cmi_OTR         FHGPDFLCRIVKYLQIVGMFASTYLLIVMSIDRCLTICQPLKSLHQ---RSNS--MYVIS
Lja_807         FNGSDFLCRTIKYLQILGMFASTYVLIMMGLDRYIAICHPLRTLRQ---SSKQSYQMIFV
Pm807           FNGSDFLCRAIKYLQILGMFASTYVLIMMGLDRYIAICHPLRTLRQ---SSKQSYQMIFV
```

```
Lja_3133        FNGSDILCRTVKYLQILGMFASTYMLIMMAMDRYIAICHPLRTIQQ---SSTQSYLMILV
Pm3133          FNGSDILCRTVKYLQILGMFASTYMLIMMAMDRYIAICHPLRTIQQ---SSTQSYMMILV
Aoc_V1A         FYGPDFLCRIVKHPQVMGMFASTYMMVMMTLDRYIAICHPLKTLQQ---STQRSYIMIIS
Ola_V1A2        FYGSDFLCRIVKHLQVMGMFASTYMMVMMTLDRYIAICHPLKTLQQ---PTKRSYIMITS
Cne_V1A2        FYGTDALCRIVKHLQVMGMFASTYMMVMMTLDRYIAICHPLKTLQQ---STKRSYVMIIS
Xma_V1A         FYGSDFLCRVVKHLQVMGMFASTYMMVMMTVDRYIAICHPLKTLQQ---PTKRSYVMIVS
Dre_V1A         FYGPDFLCRIVKHLQVMGMFASTYMMVMMTLDRYIAICHPLKTLQQ---STRRSQVMIGG
Lpl_V1A         FSGPDSLCRIVKHLQVLGMFASTYMMVMMTVDRYIAICHPLKTLQQ---PTQRSYIMIIG
Cne_V1A1        FYGPDFLCRIVKHLQVLGMFASTYMMVMMTLDRYIAICHPLQTLQQ---PTRRAYIMISS
Ola_V1A1        FYGPDFLCRIVKHLQVLGMFASTYMMVMITLDRYIAICHPLKTLQQ---PTQRAYIMIGS
Cpy_V1A         FRGTDALCRVVKHLQVFGMFASAYMLVAMTADRYIAVCHPLKTLQQ---PGRRSHAMIGS
Tgr_V1A         FRGTDALCRVVKHLQVFGMFASAYMLVAMTADRYIAVCHPLKTLQQ---PGRRSHAMIGG
Hsa_V1A         FRGPDWLCRVVKHLQVFGMFASAYMLVVMTADRYIAVCHPLKTLQQ---PARRSRLMIAA
Mmu_V1A         FRGPDWLCRVVKHLQVFAMFASSYMLVVMTADRYIAVCHPLKTLQQ---PARRSRLMIAA
Gga_V1A         FHGPDGLCRVVKHLQVFGMFASAYMLVAMTADRYIAVCHPLKTLQQ---PTKRSYAMIAA
Pan_V1A         FYGPDILCRIVKHLQVMGMFASTYMLVIMTVDRYIAICHPLKTLQQ---PTKRSYMMIIT
Xtr_V1AR        FRGPDILCRIVKHLQVFGMFASAYMLVVMTADRYIAICHPLKTLQQ---PTKRSYVMIIS
Cmi_V1A         FNGPDFLCRIVKHLQVLGMFASTYMMVMMSVDRYIAICHPLKTLQQ---ATKRSYLMIIT
Lch_V1A         FLGPDYLCKIVTHLQVLSMFISTFMLLIMTADRYIAICRPLKTLQL---SKKRPYVVVLI
Aca_V1B         FQGPDPLCKLVKYLQVLSMFASTYMLIAMTLDRYMAVCHPLRTLQQ---TSCQAYLMIGS
Gga_V1B         FAGPDLLCRAVKYLQVLSMFASTYMLMAMTLDRYVAVCHPLRTLRR---PGRQPCAMVGA
Cpy_V1B         FVGSDLLCRVVKYFQVLSMFASTYMLIMMTVDRYIAVCHPLKTLQQ---PSMQAYLMIGA
Tgr_V1B         FVGSDLLCRAVKYFQVLSMFASTYMLIMMTVDRYIAVCHPLKTLQQ---PSRQAYLMIGA
Lch_V1B         FYGTDFICRTVKYLQVLSMFASTYMLIMMTVDRYIAVCHPLKTLQQ---PTKQAYIMISA
Pan_V1B         FVGPDFLCRVVRYMQLLCMFASTYMLIMMTVDRYIAVCHPLKTIQQ---PSKQANLMIAV
Hsa_V1B         FQGPDLLCRAVKYLQVLSMFASTYMLLAMTLDRYLAVCHPLRSLQQ---PGQSTYLLIAA
Mmu_V1B         FQGSDLLCRAVKYLQVLSMFASTYMLLAMTLDRYLAVCHPLRSLQQ---PSQSTYPLIAA
Cmi_V1B         FQGPDALCRSVKYLQVVGMFASTNMLLVMTADRYLAVCHPLRTFQQ---PVRQVYVMIGA
Lja_2017        FHGPDALCRLVKYLQVVGMFASTYVLIAMTVDRYLAICHPLSTLRRRGEPRKQAHVMVAC
Pm2017          FRGPDALCRLVKYLQVVGMFASTYVLIAMTVDRYLAICHPLSTLRRRGEPRKQAHAMVAC
Cin_VPR         FFASDAMCKFVKYMQILGMYGSTYVLLCAAFDRYRAIRYPMQSFQLT---AKRVHTSVLI
Ovu_CTR2*       FMAGDAMCRFIKYAQMFSLYLSTYILIMTAVDRYRAICHPLSNQTWT---PCMVYCKIFI
Ovu_CTR1*       FLGTDLLCRLVTYLQSVAMYASSYVLVATAIDRYFAICHPLSSHKWT---TARVHVMVFI
Ovu_OPR*        FLGDISACRVYQFITVAELYASSFVLIVTALDRYISICHPLAAHMWT---NRRVHMTTAL


Aoc_V2A         AWGLALVLSIPQIFIFSRSEVAP----GEFECWGHFAEPWGLKAYVTWMTVAVFLLPALI
Ola_V2A2        AWVLALVLSIPQVFIFSRSEVAP----GEFECWGHFTEPWGLKAYVTWMTMAVFLLPALI
Cne_V2A         AWGLALVLSIPQVFIFSRSEVAP----GEFECWGHFTEPWGLKAYVTWMTVAVFLLPALI
Xma_V2A1        AWALALVLSIPQVFIFSRSEVAP----GEFECWGHFTEPWGLKAYVTWMTVAVFLLPALI
Dre_V2A2        AWGLSLILSVPQVFIFSRSEVSP----GVYECWGNFAEPWGLKAYVTWMTVAVFVLPTFI
Dre_V2A1        AWGLAFVLSLPQVFIFSQSEVSP----GVFECWGHFAEPWGLKAYVTWMTVAVFVLPALI
Ola_V2A1        AWGLSLLLSLPQVFIFSRSEVAP----GVYECWGNFAESWGLKAYVTWMTLAVFLLPVLI
Cpy_V2A         AWAASLVLSIPQIFIFSKTKLST----GTYDCYGAFIEPWGTKAYITWMTTAVFVLPTLV
Tgr_V2A         AWAASLVLSIPQIFIFSKTKLPT----GTYECNGAFIEPWGTKAYITWVTAAVFVLPTLV
Xtr_V2A         AWLASAIFSLPQGFIFSRIEVHP----GVFDCWATFIEPWGLKAYVTWITLAVFILPALF
Lch_V2A         AWVSALLLSLPQVFIFSKVKLKD----NIFECWACFVEPWGVKAYVTWITLMVFIGPMVF
Pan_V2A         AWTFSFLLSLPQIVIFSKKEIKP----GVFQCWAHFQEPWGLRTYVTWVTVMVFILPAVI
Hsa_V2A         AWAFSLLLSLPQLFIFAQRNVEGG--SGVTDCWACFAEPWGRRTYVTWIALMVFVAPTLG
Mmu_V2A         AWAFSLLLSLPQLFIFAQRDVGNG--SGVFDCWARFAEPWGLRAYVTWIALMVFVAPALG
Aca_V2A         AWTASFLFSLPQLFIFSKVELP----SGAHECWATFAEPWGARAYVTWVTLMVFILPTVF
Aoc_V2B         AWGISLLGSLPQVFIFSQVEVEP----GVFDCWANFIQPWGLQTYITWTTLVIFVLPVIT
Xma_V2B1        AWGVSLVGSLPQVFIFSQVEVAP----GVFDCWADFIQPWGLQTYITWTTLVIFILPVST
Gac_V2B1        AWGVSLLGSFPQIFIFSQIEVAP----GVFDCWATFVQPWGLQTYITWTTLVIFILPVIT
Dre_V2B1        AWLVSFVLSVPQLFIFSRVQIAP----GVYDCWAEFVQPWGPKAYVTWTTLVIFLVPVVM
Xma_V2B2        AWCVSFVGSLPQVFIFSRVEVAP----GVYDCWGQFVAPWGLRAYVTWTTLVIFVLPILI
Cmi_V2B1        AWCISLIGSLPQVFIFSKTEIYP----GVFECWAHFIEPWGLKAYVTWTSMVVFVIPALT
Lch_V2C         AWTISLIGSLPQVFIFSKVEILP----GVFDCWAHFLEPWGLKAYITWTALVVFILPTLI
Gga_V2C         SWSISLILSLPQVFIFSKIEISP----GIFECWAEFIQPWGPRAYVTWILVVIFFIPSTI
Xtr_V2C         SWCISLLFSIPQVFIFSKTEIYP----GIFECWAKFMEPWGSKAYVTWIFVAIFFIPINI
Lpl_V2C         SWCIAFIFSLPQVFIFSKTEISP----GVNECWAQFITPWGLEAYVTWVFTVIFFIPTII
Cmi_V2C         SWAMSLLLSLPQVFIS-KKEMAP----GVLQCWANFIQPEGLEAFMTWIFLIIFFAPLWS
Lja_644         AWATSLALSTPQIFIFSLSEYEK----GIFDCTATFAAHWGAKAYVTWVTLAVFVLPTLA
Pm644           AWATSLALSTPQIFIFSLSEYEK----GIFDCTATFAAHWGAKAYVTWVTLAVFVLPTLA
Lja_4769        VWVVSLLLSVPQAVIFAKGEVYP----GVYDCWGTFQPMWGVKAYITWMAMAIFVVPTAV
Pm4769          VWVVSLLLSVPQAVIFAKGEVYP----GVYDCWGTFQPMWGVKAYITWMAMAIFVVPTAV
Gac_V2C         AWLISLAFSAAQISIFSLQEVQE----NQYECWATFIEPWGSRIYITWITMSVFVIPAVI
Xma_V2C         AWLISLVFSSPQLFIFSLQEVEA----NLYDCWATFVEPWGSRIYISWITLAVFALPAVI
Dre_V2C         AWLISLAFSAPQIFIFSLQEVEK----DVFDCWATFIEPWGGRIYITWITLSVFVLPAVI
Ola_OTR         SWLLSLIFSIPQMFIFSLREVGSA-GSGVYDCWGDFVKPWGAKAYITWISLTIYIIPVAI
Xma_ITR1        SWLLSLLFSTPQIFIFSLREVG----SGVYDCWGDFVKPWGAKAYITWISLTIYIIPVAI
Aoc_ITR         SWILSLLFSIPQMFIFSLREVGLA-GSGVYDCWGDFVKPWGAKAYITWISLTIYIIPVAI
Dre_ITR         SWIISLLFSLPQVYIFSLREVG-D-G--VFDCWGDFVQPWGAKAYVTWISLTIYIIPVAI
Lpl_OTR         SWVLSLLFSIPQVYIFSLREIP-A-GSGVYDCWGDFIQPWGAKAYVTWISLTIYIIPVAI
Cne_ITR         SWILSLVFSTPQAYIFSLREVG----NGVYDCWGDFVQPWGAKAYITWMSLSIYIIPVAI
Xma_ITR2        SWILSLIFSTPQAYIFSLREVG----NGVYDCWGDFVQPWGAKAYITWMSLSIYILPVAI
Tgr_MTR         TWILSFLVSIPQIHIFSLKEVG----NGVQDCWADFIQPWGVKAYVTWITVMVYIIPVLI
Xtr_MTR         TWILSFLLSVPQIHIFSLKEVG----NKVYDCSASFIEPWGLKTYITWITITVYILPVMI
```

```
Aca_MTR         TWIVCLLFSIPQLQIFSLKEVA----HGGVDCWATFILPWGPKAYVTWITLTVYIIPMFV
Lch_OTR         SWFISLFFSIPQVHIFSLKEVG----AGVYDCWAGFIEPWGLKAYVTWITLTVYIVPVVI
Pan_MTR         SWIISLLFSIPQIHIFSLKHMG----AGVYDCWADFIQPWGVKAYVTWITLTVYIIPVVI
Hsa_OTR         TWLGCLVASAPQVHIFSLREVA----DGVFDCWAVFIQPWGPKAYITWITLAVYIVPVIV
Mmu_OTR         TWLGCLVASVPQVHIFSLREVA----DGVFDCWAVFIQPWGPKAYVTWITLAVYIVPVIV
Gga_MTR         TWLLCLLVSIPQIHIFSLRDVG----NGVYDCWADFIQPWGPKAYVTWITLMVYIIPVLM
Cmi_OTR         SWVISLAFSIPQIFIFSFREVG----PEVYDCWADFIQPWGAKAYITWITVSVYIIPVLI
Lja_807         SWFLSMLFSLPQAFIFSMSEVEN---SGIIDCWAEFIKPWGTKAYITWMTSSVFIIPVII
Pm807           SWFLSMLFSLPQAFIFSMSEVEN---SGIIDCWAEFIKPWGTKAYITWMTGSVFIIPVII
Lja_3133        SWLVSFLFSVPQAYIFSLREVQQG--TGVYDCWAAFVEPWGLKAYVTWTAVAVFVAPVVV
Pm3133          SWLVSFLFSVPQAYIFSLREVQQG--TGVYDCWAAFVEPWGLKAYVTWTAVAVFVAPVAV
Aoc_V1A         TWMCSLVLSTPQYFIFSLSEIEN--GSDVYDCWAHFIEPWGAKAYITWITVGIFLVPVVI
Ola_V1A2        TWMCSLVLSTPQYFIFSLSEIKN--GSDVYDCWAHFIQPWGAKAYITWITVGIFLVPVVI
Cne_V1A2        TWICSLVLSSPQYFIFSMSEIKN--GSEVYDCWAHFIEPWGSKAYITWMTVGIFLVPVVI
Xma_V1A         TWMCSLVLSVPQYFIFSLSEIKN--GSEVYDCWAHFVEPWGSKAYVTWMTVGIFLVPVVI
Dre_V1A         TWVCSLVLSSPQYFIFSLSEIQN--GSEVFDCWAHFIQPWGVRAYITWITAGIFLVPVLI
Lpl_V1A         TWICSLVLSTPQYFIFSLSEIKN--GSDVYDCWAHFIKPWGIKAYITWITVGIFLIPVVI
Cne_V1A1        TWACSLVLSTPQYCIFSLSEVRP--GSAVYDCWGHFVEPWGPRAYITWITAGIFLVPVAV
Ola_V1A1        TWACSLVLSTPQYFIFSLSEVRP--GSAVYDCWGHFMEPWGLRAYITWITAGIFLLPVAI
Cpy_V1A         AWALSLLLSTPQYAIFSMCEVRS--GSQVYDCWAHFIEPWGARAYVTWIAVSIFVVPVLI
Tgr_V1A         AWALSLLLSTPQYAIFSMREVRE--HSQVYDCWADFIEPWGARAYVTWIVVSIFVVPVLI
Hsa_V1A         AWVLSFVLSTPQYFVFSMIEVNN--VTKARDCWATFIQPWGSRAYVTWMTGGIFVAPVVI
Mmu_V1A         SWGLSFVLSIPQYFIFSVIEFEVNNGTKAQDCWATFIPPWGTRAYVTWMTSGVFVVPVII
Gga_V1A         AWALSLLLSTPQYFIFSLSEVER--GSRVYDCWAHFIMPWGPRAYITWITGGIFVAPVLI
Pan_V1A         AWVGSFVLSTPQYFIFSLTEVKN--GSDVHDCWANFIMPWGAKAYITWITFGIFIIPVII
Xtr_V1AR        AWIISFLLSIPQYFIFSFSPVN---GSEVYDCWAHFIQPWGARAYITWMTASIFVVPVAI
Cmi_V1A         TWMGSFILSAPQSFIFSLSEIET--GSGVYDCWANFILPWGIKAYITWITVSVFIIPVLT
Lch_V1A         AWAVSFILSTPQYFIFSVRKIEE---SEEYDCWANFVVPWGTKAYVTWITISIFIVPVII
Aca_V1B         TWLLSCLLSLPQIFIFSVREVRQG--SGVLDCWAEFRYPWGSKAYVTWMTLCVFVLPVAI
Gga_V1B         AWLLSCLLSTPQIFIFSLREVQPG--SGVLDCWADFGYPWGARAYITWTTLCIFLLPVGI
Cpy_V1B         TWLISSLLSLPQIFIFSMREVSKG--SGIIDCWAEFHFTWGAKAYVTWTTMSIFVIPVIV
Tgr_V1B         TWLISSLLSLPQIFIFSLREVSKG--SGIIDCWAEFHFNWGAKAYVTWTTMSIFVIPVIV
Lch_V1B         TWFLSSILSIPQLFIFSLTEVDQG--SGVYDCWAKFQLPWGAKAYITWTTLSIFIVPVAI
Pan_V1B         TWLVSGALSIPQMFIFSLKEVKEG--SCVFDCWATFEFDWGIQAYITWTTLSIFVIPVAI
Hsa_V1B         PWLLAAIFSLPQVFIFSLREVIQG--SGVLDCWADFGFPWGPRAYLTWTTLAIFVLPVTM
Mmu_V1B         PWLLAAILSLPQVFIFSLREVIQG--SGVLDCWADFYFSWGPRAYITWTTMAIFVLPVVV
Cmi_V1B         AWLLSCVLSLPQVFIFSLREVDEG--SGVYDCWAVFDVSWGAKAYITWTTASIFILPVLV
Lja_2017        AWVLSALLSTPQVVIFSLQEVEEG----VFDCWADFGQPWGMRAYVTWITVSVYVAPVLI
Pm2017          AWALSALLSTPQVVIFSLREVEEG----VFDCWADFGQPWGMRAYVTWITVSVYVAPVLI
Cin_VPR         AWGVSAFLSTPQLFLFEKPETRN------MCRMKRFPSAAFDQAYTTWYFFVIMFIPTCF
Ovu_CTR2*       AYAIATIFSIPQAILFQMQEVNEG--SGIYDCWVHFEPAWVLTAYALYIFFALYLIPILI
Ovu_CTR1*       AWMLSFLFSTPQLFIWSMQFSNIG-----LTCQATFDPEWTLKFYITWLTVAIWILPTIA
Ovu_OPR*        ALFLALMCSLPQLDAVLVDFHGG------KLCRPNLTTELANIAYSWWAFCSVFFVPLLL


Aoc_V2A         ITICQIRIFREIHNNIYLKSERMVMAELKKSEILFRFHGFKKEDERARERGR-RASGGGG
Ola_V2A2        ITICQIRIFREIHNNIYLKSERTVMAEVKKSDILLRFHGFKKEEERAKETQTSREAAMDE
Cne_V2A         ITICQIRIFREIHNNIYLKSERIVMAELKRSENLFRIHGFRTEEDRARERERGRQVPGRG
Xma_V2A1        ITICQIRIFREIHNNIYLKSERIMMAELKRSEILLRFQGLRKEDE--REGERGRQATGRE
Dre_V2A2        ITVCQVRIFKEIHDNIYLKSERVVSVDLKKNR----------------------------
Dre_V2A1        ITVCQVRIFREIHDNIYLKSERVVTAEFKKNSVFFHFP-----------------------
Ola_V2A1        IAFCQVRIFREIHHNLYLKSESRPSRKSPSHRGGRNVREKGGGAG---------------
Cpy_V2A         IATCQVLIFKEIHDSIYLK---SEKIVAMVKKKTQLVNGKDTQKS---------------
Tgr_V2A         IATCQVLIFKEIHDSIYLK---SERITAMVKKKTQLVNGKDTQKS---------------
Xtr_V2A         IATCQVLIFREIHNSINMGPGHSPRPRRKAKLINTRNGARSQSDT---------------
Lch_V2A         ITICQVRIFKEIYNNIYLKSERVVAEAKKLHQG----SKKGNDTS---------------
Pan_V2A         IAICQFRIFKEIHDNLYLKSERTIAQVKKQQQQQQQTSRKNSDDS---------------
Hsa_V2A         IAACQVLIFREIHASLVPGPSERPGGRRRGRRTGSPGEGA----------------------
Mmu_V2A         IAACQVLIFREIHASLVPGPSERAGRRRRGHRTGSPSEGA----------------------
Aca_V2A         IATCQGMIFHEVHRSLHLGPEGALKGR------GSLRGSS----------------------
Aoc_V2B         VTVCQVRIFRAIQINLYQKT-------------------------------------------
Xma_V2B1        VVVCQVRICRAIQTNLSRKT-------------------------------------------
Gac_V2B1        VVFCQVRICRAIQMNLHQKT-------------------------------------------
Dre_V2B1        VTACQVRICRALQINLYLKTQQQ----------------------------------------
Xma_V2B2        VIVCQVRICRAVHANIHLKARHH----------------------------------------
Cmi_V2B1        LIVCQVRICRAIQMNLYSKTHQ-----------------------------------------
Lch_V2C         VIVCQVRIYRAIQTNLYIKTHQ-----------------------------------------
Gga_V2C         LITCQVKICKIIKRNIYVKKQN-----------------------------------------
Xtr_V2C         LIVCQIKICRTIKTNIYVKQHS-----------------------------------------
Lpl_V2C         LIICQVQICRILQMNIHAKTHG-----------------------------------------
Cmi_V2C         SPICQMKICRAIMNIHVKTCQD-----------------------------------------
Lja_644         VVVCQAQICRIIRLNLYVKT--------------------HQGPD----------------
Pm644           VVICQAQICRIIRLNLYVKT--------------------HQGPD----------------
Lja_4769        LVWCQTKICRVIRLNIYFKTPHVHKDTLLLRRRRKQRRRQQKRQE----------------
Pm4769          LVWCQTKICRVIRLNIYFKTPQVHKDKLLLRRRKQRRRRQQKRQE----------------
Gac_V2C         VLYCQIRICTTIYFNMKRKALQA----------------------------------------
Xma_V2C         LLFCQMRICTTIYFNMKRKALQS----------------------------------------
Dre_V2C         LMFCQIKICAGIYFNMKRKALQG----------------------------------------
```

```
Ola_OTR         LSICYGLISFKIWQNFKLKTKREQC----------------------------------
Xma_ITR1        LSICYGLISFKIWQNFKLKTRREQC----------------------------------
Aoc_ITR         LSICYGLISFKIWQNFKMKTRRERC----------------------------------
Dre_ITR         LSVCYGLISYKIWQNFRLKTRRDQC----------------------------------
Lpl_OTR         LSVCYGLISFKIWQNFKLKTRKDQS----------------------------------
Cne_ITR         LSICYGLICFKIWQNINMKTKREHF----------------------------------
Xma_ITR2        LSICYGLICFKIWQNINMKTRRDHV----------------------------------
Tgr_MTR         LSICYGLISFKIWQNFRLKTVCET-----------------------------------
Xtr_MTR         LSVCYGLISYKIWQNIRLKTMCES-----------------------------------
Aca_MTR         LSVCYGLISFKIWQNIKLKTVHET-----------------------------------
Lch_OTR         LSVCYGLIVFKIWQNVKLKTRQET-----------------------------------
Pan_MTR         LTVCYGLISFKIWQNAKLKTRREP-----------------------------------
Hsa_OTR         LATCYGLISFKIWQNLRLKTAAAAAAEA-------------------------------
Mmu_OTR         LAACYGLISFKIWQNLRLKTAAAAAAAE-------------------------------
Gga_MTR         LSVCYGLISFKIWQNVKLKTAHGPPGGQ-------------------------------
Cmi_OTR         LTVCYGLISFKIWQNVKLKTRRETKTP--------------------------------
Lja_807         LIWCYGMITFAIWKNIKAKTQEGDSRHN-------------------------------
Pm807           LIWCYGMITFAIWKNIKAKTQEGDSRHN-------------------------------
Lja_3133        LIWCYGMITAEIWRNMRAKTERRGGREGRPPPKGMLA-----------------------
Pm3133          LLWCYGMITAEIWRNMRAKMERRGGRDGRPPPKGTPA-----------------------
Aoc_V1A         LMMCYGFICHTIWKNIKFKKRK---TMSGAAS---------------------------
Ola_V1A2        LMLCYGFICHSIWKNIKYKKRK---TTAGASN---------------------------
Cne_V1A2        LILCYGFICHSIWKNIKYKKKK---TVAGAAG---------------------------
Xma_V1A         LILCYGFICRTIWRNIKYKKKKK--TVAGAAG---------------------------
Dre_V1A         LMTCYGFICHSIRMSIRYKSK---------------------------------------
Lpl_V1A         LVVCYGFICHSIWKNIRYKTKK---HVGETAS---------------------------
Cne_V1A1        LVFCYGFICRAIWRNLKCKTRRKSADAVVEAT---------------------------
Ola_V1A1        LVFCYGFICRTIWMSIKYKTRRKRTDAAEGAT---------------------------
Cpy_V1A         LATCYGFICYIIWTNIRGKTRPSN------------------------------------
Tgr_V1A         LATCYGFICYIIWTNIRGKTRPSN------------------------------------
Hsa_V1A         LGTCYGFICYNIWCNVRGKTASRQSKGAEQAG---------------------------
Mmu_V1A         LGTCYGFICYHIWRNVRGKTASRQSKGGKGSGEAA-------------------------
Gga_V1A         LATCYGFICFRIWRSARGRARPGEAAGG--------------------------------
Pan_V1A         LTTCYGFICHSIWRNIKCKTRQGMSEYALKNG----------------------------
Xtr_V1AR        LTTCYGFICYNIWRNIQCKTKREENEGRKSHG----------------------------
Cmi_V1A         LAACYICVCYNISKNVKYKTTNNSSESAVKNG----------------------------
Lch_V1A         LLTCYGSICYNISRNIKCKTRKDLNDNALKKG----------------------------
Aca_V1B         LTLCYGLICHKICKNLRGKTK-----SGGSCRATTSIS----------------------
Gga_V1B         LTACYSLICREICKNLKGKTQSCTPVAGGLLAASPSAP----------------------
Cpy_V1B         LVVCYSLITYEICKNLKGKTQTGR------------------------------------
Tgr_V1B         LVVCYSLITYEICKNLKGKTQTGR------------------------------------
Lch_V1B         LIVCYSLICYEISKNLKYKTQTS-------------------------------------
Pan_V1B         LIVCYSLICYEIYKNFRCKTQTRK------------------------------------
Hsa_V1B         LTACYSLICHEICKNLKVKTQAWR---VGGGGWRTWDR----------------------
Mmu_V1B         LTACYGLICHEIYKNLKVKTQAGR---EER---RGWPK----------------------
Cmi_V1B         LALCYSLICLEICRNLRWKSAGSE------------------------------------
Lja_2017        LAACYGAILLEICRNLRLKTRGRYPKDPRGSQ----------------------------
Pm2017          LAACYGAILLEICRNLRLKTRGRYPKDPRGSQ----------------------------
Cin_VPR         LIYLYGMIAVLIMRNLKQKNKAGSSTSP--------------------------------
Ovu_CTR2*       LFFTYGSICYTIWAKYRHAIKTKKDANTRYPQ----------------------------
Ovu_CTR1*       LTLFYGMMCFAVWKRGRSTLGSSRTRNRSFLTNRVSTRIG--------------------
Ovu_OPR*        LIFFYGRICFVVWQSMRGRECTQSVGSSASRYVRKPIKCRISSQTS--------------


Aoc_V2A         RGGHLLKGANNSPHNNTHSSQVGECYDYVPNAIQYNSCHSEHVTTTTATTTSLMHQQTPS
Ola_V2A2        SEGQLLRGVNNNPHNNSNSAQVGECYDCAPSALQYSSCCGERMTTTSP-----IQQQTLH
Cne_V2A         GGGQLLK------DSNLHDCEVGQHYDYVPSVIQYSSCCG--QTEGTS----LSQEQIPK
Xma_V2A1        GGGRLLK------DNDAHDCKVGEHYDYVPSVIQYSSCCGEHQTAATT----LMQEQIQN
Dre_V2A2        ----------------------------------------------------------
Dre_V2A1        ----------------------------------------------------------
Ola_V2A1        ----------------------------------------------------------
Cpy_V2A         ----------------------------------------------------------
Tgr_V2A         ----------------------------------------------------------
Xtr_V2A         ----------------------------------------------------------
Lch_V2A         ----------------------------------------------------------
Pan_V2A         ----------------------------------------------------------
Hsa_V2A         ----------------------------------------------------------
Mmu_V2A         ----------------------------------------------------------
Aca_V2A         ----------------------------------------------------------
Aoc_V2B         ----------------------------------------------------------
Xma_V2B1        ----------------------------------------------------------
Gac_V2B1        ----------------------------------------------------------
Dre_V2B1        ----------------------------------------------------------
Xma_V2B2        ----------------------------------------------------------
Cmi_V2B1        ----------------------------------------------------------
Lch_V2C         ----------------------------------------------------------
Gga_V2C         ----------------------------------------------------------
Xtr_V2C         ----------------------------------------------------------
```

```
Lpl_V2C         ------------------------------------------------------------
Cmi_V2C         ------------------------------------------------------------
Lja_644         ------------------------------------------------------------
Pm644           ------------------------------------------------------------
Lja_4769        ------------------------------------------------------------
Pm4769          ------------------------------------------------------------
Gac_V2C         ------------------------------------------------------------
Xma_V2C         ------------------------------------------------------------
Dre_V2C         ------------------------------------------------------------
Ola_OTR         ------------------------------------------------------------
Xma_ITR1        ------------------------------------------------------------
Aoc_ITR         ------------------------------------------------------------
Dre_ITR         ------------------------------------------------------------
Lpl_OTR         ------------------------------------------------------------
Cne_ITR         ------------------------------------------------------------
Xma_ITR2        ------------------------------------------------------------
Tgr_MTR         ------------------------------------------------------------
Xtr_MTR         ------------------------------------------------------------
Aca_MTR         ------------------------------------------------------------
Lch_OTR         ------------------------------------------------------------
Pan_MTR         ------------------------------------------------------------
Hsa_OTR         ------------------------------------------------------------
Mmu_OTR         ------------------------------------------------------------
Gga_MTR         ------------------------------------------------------------
Cmi_OTR         ------------------------------------------------------------
Lja_807         ------------------------------------------------------------
Pm807           ------------------------------------------------------------
Lja_3133        ------------------------------------------------------------
Pm3133          ------------------------------------------------------------
Aoc_V1A         ------------------------------------------------------------
Ola_V1A2        ------------------------------------------------------------
Cne_V1A2        ------------------------------------------------------------
Xma_V1A         ------------------------------------------------------------
Dre_V1A         ------------------------------------------------------------
Lpl_V1A         ------------------------------------------------------------
Cne_V1A1        ------------------------------------------------------------
Ola_V1A1        ------------------------------------------------------------
Cpy_V1A         ------------------------------------------------------------
Tgr_V1A         ------------------------------------------------------------
Hsa_V1A         ------------------------------------------------------------
Mmu_V1A         ------------------------------------------------------------
Gga_V1A         ------------------------------------------------------------
Pan_V1A         ------------------------------------------------------------
Xtr_V1AR        ------------------------------------------------------------
Cmi_V1A         ------------------------------------------------------------
Lch_V1A         ------------------------------------------------------------
Aca_V1B         ------------------------------------------------------------
Gga_V1B         ------------------------------------------------------------
Cpy_V1B         ------------------------------------------------------------
Tgr_V1B         ------------------------------------------------------------
Lch_V1B         ------------------------------------------------------------
Pan_V1B         ------------------------------------------------------------
Hsa_V1B         ------------------------------------------------------------
Mmu_V1B         ------------------------------------------------------------
Cmi_V1B         ------------------------------------------------------------
Lja_2017        ------------------------------------------------------------
Pm2017          ------------------------------------------------------------
Cin_VPR         ------------------------------------------------------------
Ovu_CTR2*       ------------------------------------------------------------
Ovu_CTR1*       ------------------------------------------------------------
Ovu_OPR*        ------------------------------------------------------------

Aoc_V2A         SSDCQEPYSNSPRCSLDYVPPPLPATPPPSITKAMSKTVRMTLVIVLVYTICWSPFFIVQ
Ola_V2A2        NSDFQEPHTNSPRCSLEYAPPHSPATPPPSITKAMSKTIRMTLVIVLVYTICWSPFFIVQ
Cne_V2A         GSDFRESCTTSPRCSLDCAPPPHPVTPPHSITKAMSKTVRMTLVIVLVYTICCSPFFIDQ
Xma_V2A1        GSDFRESCANSARCSLDYAPPPHQVTPVPSITKAMSKTVRMTLVIVLVYTVCWSPFFIVQ
Dre_V2A2        ----------------PSFQPCPLHPPLTGVSKAMSKTVKMTLVIVLVYTVCWSPFFIVQ
Dre_V2A1        -------------------------RREAGVTKAMSKTVRMTLVIVLVYTICWSPFFIVQ
Ola_V2A1        -----------------RFCLPLGVSSCGQVTAAMSKTVRMTLVIVVVYSVCWAPFFSVQ
Cpy_V2A         ----------------------------EVTTAMSKTIKMTMVIVLVYVVCWAPFFLIQ
Tgr_V2A         ----------------------------EVTTAMSKTIKMTMVIVLVYVVCWAPFFLIQ
Xtr_V2A         ----------------------------GVSSAMAKTVRMTLVIVLVYVVCWAPFFIAQ
Lch_V2A         ----------------------------SVSSAMSKTIKMTFVIVLVYTACCAPFFIVQ
Pan_V2A         ----------------------------GVSTAMSKTIRMTFVIVVIYIACWAPFFITQ
Hsa_V2A         ----------------------------HVSAAVAKTVRMTLVIVVVYVLCWAPFFLVQ
Mmu_V2A         ----------------------------HVSAAMAKTVRMTLVIVIVYVLCWAPFFLVQ
Aca_V2A         ----------------------------PISCALTKTLKMTLVIVLTYVFCWAPFFLVQ
```

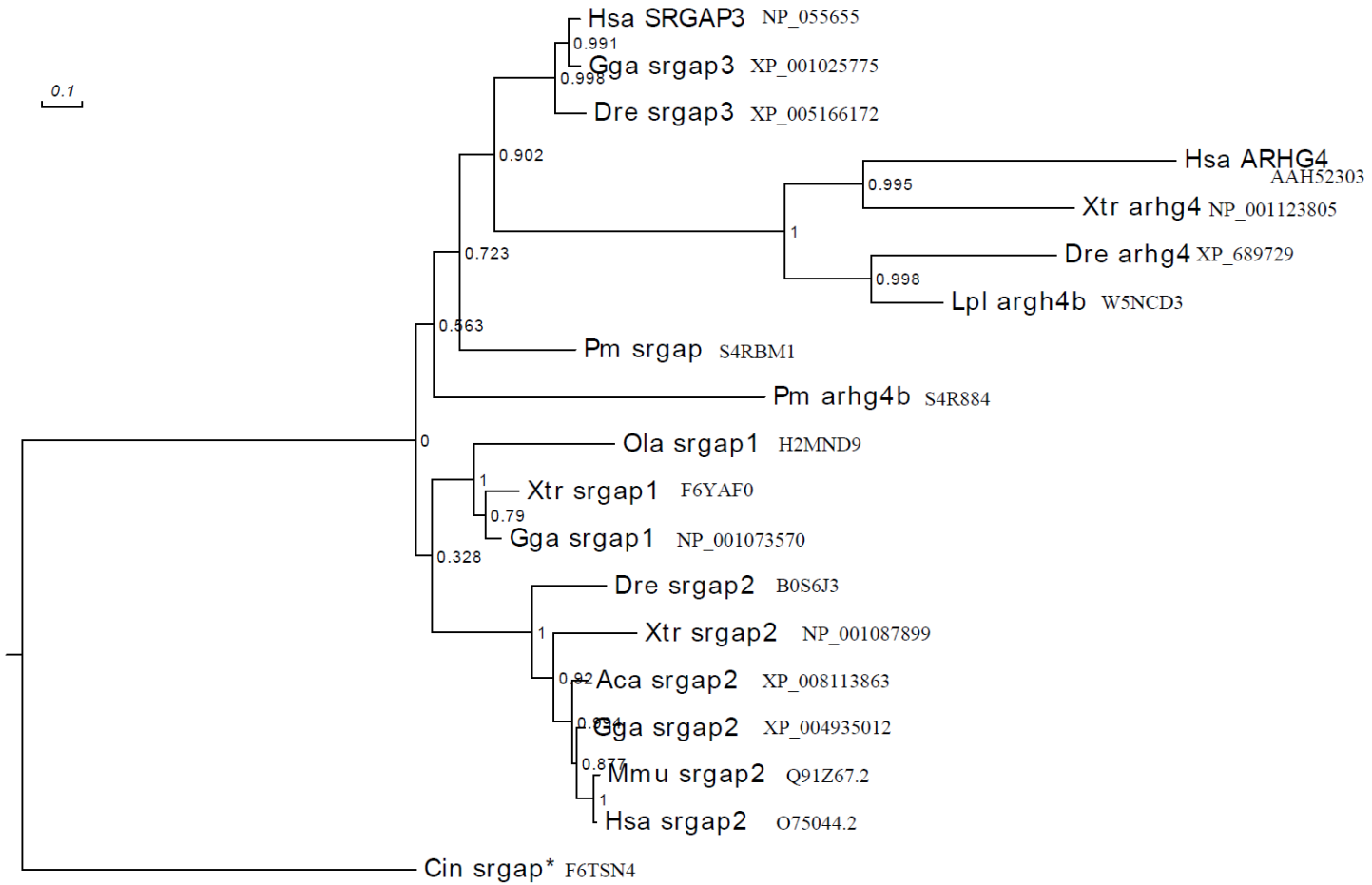
```
Aoc_V2B          ----------QQLGS-VGVPLPSRASGVAGMSKARVKTLKMTVVIVLAYIICWAPFFTVQ
Xma_V2B1         ----------HRQGVSVGAPLPARASGAAGMSKARVKTLKMTVVIVLVYVMCWAPFFTVQ
Gac_V2B1         ----------AGLQGRAGHPAPSRGSGVAGMSKARVKTLKMTVVIVVAYIVCWAPFFTVQ
Dre_V2B1         ---------Q-RGDEGHSHPLSSRASNVVSVSKSRIKTVKMTVVIVLAYTVCWAPFFTVQ
Xma_V2B2         ---------AGAAGEAASRPLSSRASCVVGLSKARVKTVKMTVVIVLAYVVCWTPFFTVQ
Cmi_V2B1         ----------QSGESKKELKTWRASSVAGVSKARIKTVKMTLVIVLVYILCWSPFFTVQ
Lch_V2C          ----------DSENTKKTTLPSRASSVAGVSKARIKTVKMTVVIVVAYIACWAPFFTVQ
Gga_V2C          ----------EYQVTNQKQVLPSRASSVNCISKAMIKTVKMTIVTVVAYVLCWSPFFIAQ
Xtr_V2C          ----------DLELQEQHRIGASRASSINCISKAMIKTVKMTVVTVVAFFLCWTPFFIVH
Lpl_V2C          ----------DILRHNQMQVIPSRASNFNCFSKAMIKTVKMTIITVVVYILCWAPFFVVQ
Cmi_V2C          ----------DPQMGGSVLQSHQHQPRLECDE---------------------------
Lja_644          ----EEEAAGGDERRAGGQLMPSRASSVTGISRAMINTVRMTLVIVLVYVACWAPWFTVQ
Pm644            ----EEETIGGDERRAGGQLMPSRASSVTGISRAMINTVRMTLVIVLVYVACWAPWFTVQ
Lja_4769         ----DEEEVAALHRGQTHAVTMTRVSSVTGVSRAMVKTVKMTLVIVIVYVLCWSPFFVAQ
Pm4769           ----DEEEVAALHRGQTHAVTMTRVSSVTGVSRAMVKTVKMTLVIVIVYVLCWSPFFVAQ
Gac_V2C          -------------------VRAGTKGVSKGVSSAMLKTVKMTFVIIMTYTVCWSPFFVVQ
Xma_V2C          --------------------GRAGTKGVSS----AMLKTVKMTFVIIMAYTVCWSPFFVVQ
Dre_V2C          -------------------SAGDGRSSTKGISSAMLKTVKMTFIIILVYTLCWSPFFVVQ
Ola_OTR          --------INLTPKTSQ-NNTLARVSSVKLISKAKITTVKMTFVIVVAYIVCWTPFFSVQ
Xma_ITR1         --------ISLTPKTCK-SNTLARVSSVRLISKAKITTVKMTFVIVVAYIVCWTPFFSVQ
Aoc_ITR          --------ISLTPKTSK-GNTLARVSSVKLISKAKITTVKMTFVIVVAYIVCWTPFFSVQ
Dre_ITR          --------LSLTPRPTK-GAALSRVSSVKLISKAKIRTVKMTFVIVMAYIICWTPFFFVQ
Lpl_OTR          --------IPLTPKAFK-GSALSRVSSVKLISKAKIRTVKMTFVIVLAYIVCWTPFFFVQ
Cne_ITR          --------LALTPRPSKSAHPLSRVSSVRLISKAKIRTVKMTFVVVIAYIVCWTPFFFVQ
Xma_ITR2         --------LALTPRPSKGAHPLSRVSSVRLISKAKIRTVKMTFVVVIAYIVCWTPFFFVQ
Tgr_MTR          --------NVALTSTSTRRTTLSRVSSVKLISKAKIRTVKMTFIIVLAYIVCWTPFFFVQ
Xtr_MTR          --------SVRLS--SNKRATLSRVSSVRLISKAKIRTVKMTFIIVLAYIVCWTPFFFVQ
Aca_MTR          --------NVNLTSSHYHGGTLSRVSSIKLISKAKIRTVKMTFIIVLAFVMCWTPFFIVQ
Lch_OTR          --------HVNVASKSSREAVFSRVSSVKLISKAKIRTVKMTFVIVLAYILCWTPFFFVQ
Pan_MTR          --------NMVMATKPARGAMLSRVSSVKLISKAKIRTVKMTFVIVLAYIVCWTPFFSVQ
Hsa_OTR          --------PEGAAAGDGGRVALARVSSVKLISKAKIRTVKMTFIIVLAFIVCWTPFFFVQ
Mmu_OTR          --------GSDAAGG-AGRAALARVSSVKLISKAKIRTVKMTFIIVLAFIVCWTPFFFVQ
Gga_MTR          --------SSTARGG----AAFARVSSTRLISKAKIRTVKMTFIIVLAFIVCWTPFFFVQ
Cmi_OTR          ----------------QGGSMLARVSSIKIISRAKIRTVKMTFVIVLAYIICWTPFFYVQ
Lja_807          ----------------PAKNSAPSRVSSVRSISKAKIRTAKMTFVIIMVYIICWTPFFFVQ
Pm807            ----------------PAKSSAPSRVSSVRSISKAKIRTAKMTFVIIMVYIICWTPFFFVQ
Lja_3133         --------RSSAAVAPPAAATASRVSSVRNISKAKIRTTKMTFVIIVVYVFCWTPFFLVQ
Pm3133           --------RSSAAVAPVATATASRVSSVRNISKAKIRTTKMTFVIIVVYVFCWTPFFLVQ
Aoc_V1A          ----------------KNGLIGKNSVSSVTTISRAKLRAVKMTFVIVLAYIICWAPFFIVQ
Ola_V1A2         ----------------KNGLIGKNSVSSVTTISRAKLRTVKMTFVIVLAYIVCWSPFFIVQ
Cne_V1A2         ----------------KNGLIGKCSVSSITTISRAKLRTVKMTFVIVLAYIICWAPFFTVQ
Xma_V1A          ----------------KSGLIGKSSVSSIMMISRAKLRTVKMTFVIVLAYIVCWAPFFTVQ
Dre_V1A          ----------------RTQVFGKQSVSSVSSISRAKLRTVKMTFVIVLAYVICWAPFFTVQ
Lpl_V1A          ----------------KNGLIARSSVSSVSTISRAKLRTVKMTFVIVLAYVVCWAPFFIVQ
Cne_V1A1         ----------------KSGILGRSSVSSVSTISRAKLRTVKMTFVIVVAYVVCWAPFFTVQ
Ola_V1A1         ----------------TNGVLSRSSVSSVSTISRAKLRTVKMTFVIVLAFVVCWAPFFTVQ
Cpy_V1A          ---------------GGAQLLSTSAVSSVKTISRAKIRTVKMTFVIVSAYVICWAPFFTVQ
Tgr_V1A          ---------------GGAQLLSTSAVSSVKTISRAKIRTVKMTFVIVSAYVICWAPFFIVQ
Hsa_V1A          ------------VAFQKGFLLAPCVSSVKSISRAKIRTVKMTFVIVTAYIVCWAPFFIIQ
Mmu_V1A          ------------GPFHKGLLVTPCVSSVKSISRAKIRTVKMTFVIVSAYILCWTPFFIVQ
Gga_V1A          -------------GPRRGLLLAPCVSGVKTISRAKIRTVKMTFVIVSAYVVCWAPFFTIQ
Pan_V1A          -------------------LMPSCVSSVRTISRAKIRTVKMTFVIVVAYIVCWAPFFIVQ
Xtr_V1AR         -------------------LLSTSVSSVKTISRAKIRTVKMTLVIVTAYILCWAPFFIVQ
Cmi_V1A          -----------------LITSGVNVSNVKTISKAKIRTVKMTLVIVLAYIVCWAPFFSVQ
Lch_V1A          --------------------LLSCVSNVKIMSQAKLKTVKMTFVIVLTYIVCWAPFFTIQ
Aca_V1B          --------YTGQKGAECHGG--SRVSSVRTISRAKIRTVKMTFVIVMAYVACWAPFFSVQ
Gga_V1B          --------CCSQKGSQCPSGQPSRVSSVRTISRAKIRTVKMTFVIVVAYVACWAPFFSMQ
Cpy_V1B          ----------SENGCYRGQGMPSRVSSIRTISRAKIRTVKMTFVIVLAYVFCWTPFFSVQ
Tgr_V1B          ----------SENGCYRGQGMPSRVSSVRTISRAKIRTVKMTFVIVLAYVLCWTPFFSVQ
Lch_V1B          ----------VESGCHSGPVMPSRVSSVRTISRAKIRTVKMTFVIVFTYIACWSPFFSVQ
Pan_V1B          ----------SKNGLQNSQLAASRVSSVRSISRAKIRTVKMAFVIVLTYIACWTPFFSVH
Hsa_V1B          --------PSPSTLAATTRGLPSRVSSINTISRAKIRTVKMTFVIVLAYIACWAPFFSVQ
Mmu_V1B          --------SSSSAAAAATRGLPSRVSSISTISRAKIRTVKMTFVIVLAYIACWAPFFSVQ
Cmi_V1B          -----------GEGRPGGEMGPGRVSSIKSMSRAKIRTVKMTFVIVLAYVVCWSPFFSVQ
Lja_2017         ----------------SQSPVCSRVSSVRCISRAKVRTIKMTLVIIVVYVVCWTPFFVIQ
Pm2017           ----------------SQSPVCSRVSSVRCISRAKVRTIKMTLVIIVVYVVCWTPFFVIQ
Cin_VPR          --------------VRQGGNFAPRASSTAGISRAKIKTVQMTLVIVITYVVLWTPFFTVQ
Ovu_CTR2*        --------------RRKKKGVILRTHSVHGFSKAKLNSVKLTFAVIVTYIICWSPFFVSQ
Ovu_CTR1*        -------QSHLARGFSEEDMEGQSVNYNRGISRAKVRSVALTLSVVACCFICWSPFFVCQ
Ovu_OPR*         ----SENRVKNYSDARDKDSSRNPRAICRGVSKSKIKTIKLTFSVVACFIICYTPFFTVL
```

```
Aoc_V2A         LWAAWDPN---PPDQGVAFTILMLLASLNSCTNPWIYTAFSSSVSRELQNLLHCRSRPSR
Ola_V2A2        LWAAWDPN---PPDQGVAFTILMLLASLNSCTNPWIYTAFSSSVSRELQNLLHCRSRLGR
Cne_V2A         LWPAWDPD---PPDQGVAFTILMLLASLNSCTNPWIYTAFSSSVSRELQNLLQCRSRLGR
Xma_V2A1        LWAAWDPD---PPDQGVAFTILMLLASLNSCTNPWIYTAFSSSVSRELQNLLQCRSRLGR
Dre_V2A2        LWAAWDPN---PPDQGVAFTILMLLASLNSCTNPWIYTAFSSSVSRELLALLRCRPKVPR
Dre_V2A1        LWAAWDPN---PPIQGAVFTILMLLASLNSCTNPWIYTAFSSSVSRELLALLRCHRRSPR
Ola_V2A1        MWAAWYPD---PPQRDAAFTLLMLLASLNSCTNPWIYSAFSSSVSPELR-LLLCRGSFHH
Cpy_V2A         MWSVWDPK---AETERPVFVILMMLASLNSCTNPWIYTTFSSSVSNDLHQMFCCARRG-L
Tgr_V2A         MWSVWDPK---AETERPVFVILMMLASLNSCTNPWIYTTFSSSVSNDLHQMFCYVRRG-I
Xtr_V2A         LWNVWNQE---SAAGSSAIKVLMILASLNSCTNPWIYTIFSSSVSKDVKEILCFACRPRQ
Lch_V2A         LWSVWDPK---APKEGVTFVLLMLLASLNSCTNPWIYTAFSNSVSRELQLLLCCQGQNVR
Pan_V2A         LWSVWDPH---APKEGVAFTILMLLASLNSCSNPWIYTAFSSSVSQELWTLLCCVHNKFR
Hsa_V2A         LWAAWDPE---APLEGAPFVLLMLLASLNSCTNPWIYASFSSSVSSELRSLLCCARGRTP
Mmu_V2A         LWAAWDPE---APLERPPFVLLMLLASLNSCTNPWIYASFSSSVSSELRSLLCCAQRHTT
Aca_V2A         LWSVWDPE---APIDGPAFTLLMLVASLNSCTNPWVYAAFSSSISNELCRIFCPRLARQR
Aoc_V2B         LWSAWDIN---APKETATFTILMLLASLNSCANPCIYLLFSGQFPKRLVTLLCR-RHSEG
Xma_V2B1        LWSAWDAN---APKETAIFTILMLLASLNSCANPCIYLLFCGEFPKRLVKLLRPPALSAG
Gac_V2B1        LWSAWDTN---APKESATFTVLMLLASLNSCANPCIYLLFSGQFPKKMRTFLWQWRPNG-
Dre_V2B1        LWSVWDDD---APTETATFTILMLLASLNSVANPCIYLLFTVKFPELLGSLLCMKHAEK-
Xma_V2B2        LWSVWDAE---APTQSATFTILMLLASLNSCVNPCIYLVFSGKLPERLVALMCVDEADL-
Cmi_V2B1        LWSVWDPK---APRETATFTILMLLASLNSCANPFIYMFFSGQMPKGLVETLCRKNPSG-
Lch_V2C         LWSVWDHN---APKETAVFTILMLLANLNSCTNPCIYLFFSGQLPKSLIAMVCGKRSDI-
Gga_V2C         LWSVWFPS---GITEGSAFTIIMLLGNLNSCTNPWIYMYFCGHIPYCTNKQLENTSAQED
Xtr_V2C         LWVAWSSE---DVTEGAAFTIIMLLGNLNSCANPWIYMYFSGHIPRCVS-HRETLSTRED
Lpl_V2C         LWSVWFPS---AITEGSVFTIIMLLGNLNSCANPWIYMYFCGHIP-CCKKHQGSSSIRED
Cmi_V2C         ------------------------------------------------------------
Lja_644         LWSAWDSK---APKEGPTFVIIMLLGNLNSCTNPWIYLWFAGGLSPGGCRGLVPHARPPA
Pm644           LWSAWDSK---APKEGPTFVIIMLLGNLNSCTNPWIYLWFAGGLPPGGCRGLVPHARPPA
Lja_4769        LWSVWDPN---PPFEGALFTITLLLASLNSCTNPWIYMAFSGSTPRALLSCVLCRARRGG
Pm4769          LWSVWDPN---PPFEGALFTITLLLASLNSCTNPWIYMAFSGSTPRALLSCVLCRARRGG
Gac_V2C         LWSAWSPSS--TPTQGPVFAIIMLLASLNSCTNPWIYLYYS--------------------
Xma_V2C         LWSAWSPSS--APTQGPIFSIIMLLASLNSCTNPWIYLYYS--------------------
Dre_V2C         LWSAWSPSS--APTQGPVFVTIMLLASLNSCTNPWIYLYYS--------------------
Ola_OTR         MWSAWDQAA---PREAMPFIISMLLASLNSCCNPWIYMCFAGHLFHDLRQNLLCCSTRYL
Xma_ITR1        MWSAWDPAA---PREAMPFIISMLLASLNSCCNPWIYMCFAGHLFHDLRQNFLCCSARYL
Aoc_ITR         MWSAWDPAA---PREAMPFIISMLLASLNSCCNPWIYMCFAGHLFQDLRQNFLCCSAHYL
Dre_ITR         MWSAWDPMA---PREAMAFIIAMLLASLNSCCNPWIYMFFAGHLFRDLMQRCLTVSH---
Lpl_OTR         MWSAWDPAA---PREELAFIIAMLLASLNSCCNPWIYMFFAGHLFHDLMQHFLCCSASYL
Cne_ITR         MWSAWDPAA---PREDMAFIISMLLASLNSCCNPWIYMLFAGHLFHDLIKSFCCCCRNYV
Xma_ITR2        MWSAWDPAA---PREDMAFIISMLLASLNSCCNPWIYMFFAGHLFHDLMRSFFCCCRHYL
Tgr_MTR         MWSVWDINA---PKEASLFIIAMLLGSLNSCCNPWIYMLFTGHLFHDLLQRFLCCSAHYL
Xtr_MTR         MWSVWDPDA---PKEDSLFIIAMLLGSLNSCCNPWIYMLFTGHLFHDLLQRFLCCSARYL
Aca_MTR         MWIVWDENA---PKEELAFIITTLLASLNSCCNPWIYMLFTGHLFHDLLHRFLCCSSQYL
Lch_OTR         MWSVWDPKP---PREAFAYVIAMLLASLNSCCNPWIYMFFTGHLFQDLVQRFLCCSAQYL
Pan_MTR         MLSAWDPDV---PQEALAYIIAMLLASLNSCYNPWIYMYFARHLFHDLAQQFLCCSTHYL
Hsa_OTR         MWSVWDANA---PKEASAFIIVMLLASLNSCCNPWIYMLFTGHLFHELVQRFLCCSASYL
Mmu_OTR         MWSVWDVNA---PKEASAFIIAMLLASLNSCCNPWIYMLFTGHLFHELVQRFLCCSARYL
Gga_MTR         MWSVWDTNA---PQEASPFIIAMLLASLNSCCNPWIYMLYTGHLFHDLMRRFLCCSARYL
Cmi_OTR         MWTAWDPHA---PKEDLPFVLAMLLASLNSCCNPWIYMFFTGRLFQDLMHRFLCCSLLLA
Lja_807         MWSVWDSSA---PFEGIPFAIVMLLASLNSCTNPWIYMFFSGHLLYDFVRYFPCGARSRA
Pm807           MWSVWDSSA---PFEGIPFAIVMLLASLNSCTNPWIYMFFSGHLLYDFVRYFPCGARSRA
Lja_3133        MWSVWDGSA---PFEGSAFTIVMLLASLNSCTNPWIYMFFSGHLINDFAHCIPCCASSLR
Pm3133          MWSVWDGSA---PFEGSAFTIVMLLASLNSCTNPWIYMFFSGHLIYDFAHCIPCCAGALR
Aoc_V1A         MRSVWDENFQWADSVDTAVTLSALLASLNSCCNPWIYMIFSGHLLQDFVHCFSCCRKVNT
Ola_V1A2        MWSVWDENFLWDDSDNTAVTLSALLASLNSCCNPWIYMIFSGHLLQDFVHCFSCCRRVNA
Cne_V1A2        MWSVWDKNFKWDDSENTAVSLSALLASLNSCCNPWIYMIFSGHLLQDFMHCFSCCRKLNA
Xma_V1A         MWSVWDRNFQWDDSENTAVTLSALLASLNSCCNPWIYMIFSGHLLQDFVHCFSCCCRLSG
Dre_V1A         MWSVWDQSFNWDDSKNTAVTLSALLASLNSCCNPWIYMVFSGHLLQDFALCFPCCQKIRK
Lpl_V1A         MWSVWDENFSWDESENTAVTLSALLASLNSCCNPWIYMIFSGHLLQDFALCFPCCQRLQQ
Cne_V1A1        MWSVWDKTFSWDDSENTAVSLSALLASLNSCCNPWIYMTFSGHLLQDFMHCFSCCYRLRD
Ola_V1A1        MWPVWDQTFSWDDSENTTVTLSALLASLNSCCNPWIYMIFSGHLLSDFFSGLPCCRRLTI
Cpy_V1A         TWSVWDKNFMWIDSENTAVTVTSLLASLNSCCNPWIYMFFSGHLFQDFIHSFICWRKMQH
Tgr_V1A         TWSVWDKNFTWIDSENTSVTVTSLLASLNSCCNPWIYMFFSGHLFQDFIQSFVCWRKMQH
Hsa_V1A         MWSVWDPMSVWTESENPTITITALLGSLNSCCNPWIYMFFSGHLLQDCVQSFPCCQNMKE
Mmu_V1A         MWSVWDTNFVWTDSENPSTTITALLASLNSCCNPWIYMFFSGHLLQDCVQSFPCCQSIAQ
Gga_V1A         MWSVWDQHFPWVDSENTATTVTALLASLNSCCNPWIYMFFSGHLLQDCVQSFPCCQKIKQ
Pan_V1A         MWSVWDDQFSWDESENTAVTVSALLASLNSCCNPWIYMFFSGHLLQDFIQCFPCCQKLQQ
Xtr_V1AR        MRSVWDKNFEWTDSEDIATTVTALLGSLNSCCNPWIYMFFSGHLLQDFIHSFPCFQKIKQ
Cmi_V1A         MWSVWDQKAPKDDSTDTAFTLTMLLASLNSCCNPWIYMFFSGHLLSDVSKFFPCCHKFTQ
Lch_V1A         MLAVWVKHFEWEESVNPIVTITALLASLNSCTNPWIYMFFSGHLLKDFTEGFSCCYKLQQ
Aca_V1B         MWSVWDKDAPNDESSNVTFTITMLLASLSSCCNPWIYMFFSGHLFHDVFHFFACCGSLHS
Gga_V1B         MWSVWDEDAPDDESADAAFTITMLLASLSSCCNPCIYMFFSGRPLQDTARCLALWGCPPS
Cpy_V1B         MWSVWDANAPNEESSDFAFTITMLLASLSSCCNPWIYMFFSGRVLQDVLGNVSCCRSSRR
Tgr_V1B         MWSVWDENAPNEESSDFAFTITMLLASLSSCCNPWIYMFFSGRVLQDVLGYVSCCRSSRR
Lch_V1B         MWTVWDENSPNEDSSDFAYTINMLLASLSSCCNPWIYMFFSGHLVHDVLNYFPCCGRFQP
```

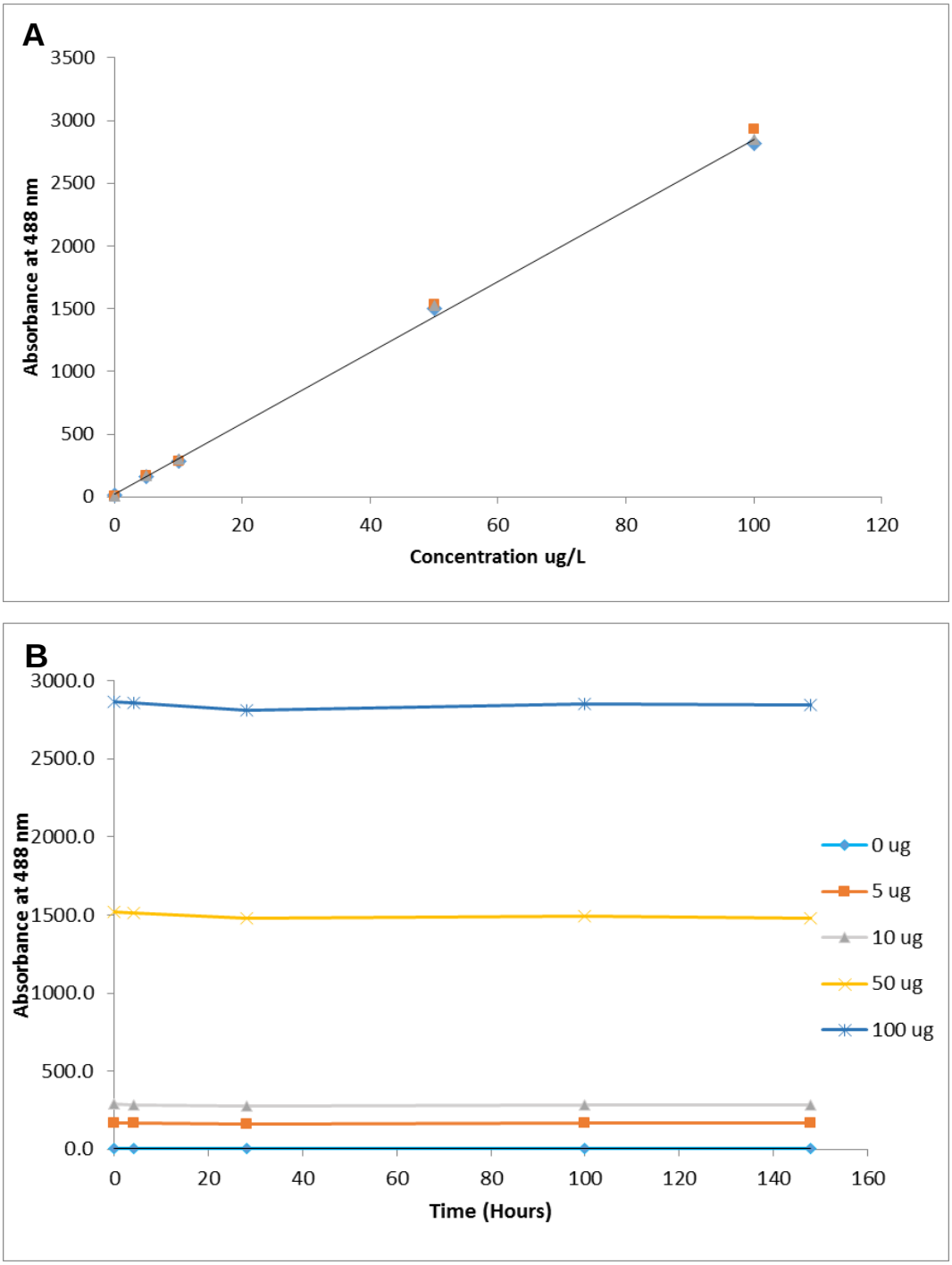
```
Pan_V1B         MWSVWDVNAPKEDSDDTAFTITMLLGSLSSCCNPWIYMAFSGHFIQDIFRYIPCCRRFQY
Hsa_V1B         MWSVWDKNAPDEDSTNVAFTISMLLGNLNSCCNPWIYMGFNSHLLPRPLRHLACCGGPQP
Mmu_V1B         MWSVWDENAPNEDSTNVAFTISMLLGNLSSCCNPWIYMGFNSHLLPRSLSHRACCRGSKP
Cmi_V1B         MWSVWDLAAPDDGSSNSLFTIAMLLANLNSCCNPWIYLVFSGNLRQSAHRYLLPCQGQPR
Lja_2017        MWAAWDETAPDDDSSDPTYTIVMLLSSLNSCTNPWIYMSFSGHLLEEVLACCRQVRCPAT
Pm2017          MWAAWDETAPDDDSSDPTYTIVMLLSSLNSCTNPWIYMSFSGHLLEEVMSCCRQARCATA
Cin_VPR         MLAAWGAIN----EDHIAASCIKLLASLNSCTNPWIYMAFSGNLLGDIKRFILCQKKKKS
Ovu_CTR2*       IWWLFDET----VVGNAGVVVILLMACLNSCTNPWIYLIFNRNYISNVLPCKCLRRHRVE
Ovu_CTR1*       MWAAWDEN---APYSGAIYTILLLLSSLNSCTNPWIYMIFSVFQHRAKTSRFVNDEETTS
Ovu_OPR*        MARTYDAELS--SAQTPALVILSLLPSLNSCTNPWIYLAFSGKVWCRQQSQNFPRTWTQT


Aoc_V2A         RGSLPDDSTTTHTSTTKDSLY----------------------------------------
Ola_V2A2        RGSLHDDSTTTHTFTTKDNLY----------------------------------------
Cne_V2A         RGSLPDDSTATHTSTTKDNLY----------------------------------------
Xma_V2A1        RGSLPDDSTATHTSTTKDNLY----------------------------------------
Dre_V2A2        RSSLYDHSSDINTSTTKDNQY----------------------------------------
Dre_V2A1        RGSLPDDSTATHTSTNTKDSIY---------------------------------------
Ola_V2A1        RGSLPGDSNTTHTSNA---------------------------------------------
Cpy_V2A         RKNSNPEDSCFTGTSTVPKDSLY--------------------------------------
Tgr_V2A         RKNSIPEDSCLTGTSTVPKDSLY--------------------------------------
Xtr_V2A         RKNSLPEDSCFTGSTTVPKECLY--------------------------------------
Lch_V2A         EAAHPEESSCSTAISSLPKGAPC--------------------------------------
Pan_V2A         RKSIGEDS-CITASSSLPKESLY--------------------------------------
Hsa_V2A         PSLGPQDESCTTASSSLAKDTSS--------------------------------------
Mmu_V2A         HSLGPQDESCATASSSLMKDTPS--------------------------------------
Aca_V2A         MGSLRED-SVLTASFSLGRDTLS--------------------------------------
Aoc_V2B         KESMHEEATLVSTLYMSFKNVSESK------------------------------------
Xma_V2B1        KNPMREEVTLVSTLYMSFKNASESKGL----------------------------------
Gac_V2B1        KDSIREEATMVSTLYMSFKNVSEAN------------------------------------
Dre_V2B1        KEPLPEETTMVSSLYLSFKSHSDIR------------------------------------
Xma_V2B2        KESMQEEATMVSSLYISLKGLSDCR------------------------------------
Cmi_V2B1        GDNVIEEPSLISTLHLTQRSSCESRIVTFLN------------------------------
Lch_V2C         RDSMPEEPSVVSTLYVSQKSLS---------------------------------------
Gga_V2C         SVVTGSIHLVDRDPEENSTCA----------------------------------------
Xtr_V2C         SLNTASIDLGDRDCEDRTTAV----------------------------------------
Lpl_V2C         SVITASIHIGDKGPTDHSTSL----------------------------------------
Cmi_V2C         ----------------------------------------------------------
Lja_644         P------DFNVDESVMSTATFREDRDDSLRGGDAPTPPLASLPPSQAQAASSPHAVSCLP
Pm644           P------DFNVDESVMSTATFREDRDDSLRGGDTPTPPFACLPPSQAQAASSLHAVSCLP
Lja_4769        AGGIGGGGVGSDGGIGGGGGGTGGWVRSVPGATAHEDSMATSSVHLTGALRSQHEDDAEQ
Pm4769          AGGIAGGGVGSDDGIGGGGG-TGGWVRSVPGATAHEDSMATSSVQLAGALRSQHEDDAEQ
Gac_V2C         ----------------------------------------------------------
Xma_V2C         ----------------------------------------------------------
Dre_V2C         ----------------------------------------------------------
Ola_OTR         KSPECRCERDFNSSHKSNSSNFAIKSTSSSRSITQTSTT----------------------
Xma_ITR1        KSSQCRCEHDFDSSHKSNSSTFAIKSTSSQRSITQTSTT----------------------
Aoc_ITR         KSSQCRCERDFDSSHKSNSSTFAIKSTNSQRSITQTSTT----------------------
Dre_ITR         ----CGCKRHWR--NKNHSGTGAMRSSSSQKSVTQSSTT----------------------
Lpl_OTR         KATQCSCELDSS--RKSNSSTYIIKNSSSQKSITQTSST----------------------
Cne_ITR         ADSSCHVNQECR--HKRGSSTFDIKNGSSMRSLTHTSSLGGPTH----------------
Xma_ITR2        ADSSCRCDQQCR--HKSGSSTFVIKNGSSMRSLTHTSSTGGPAQ----------------
Tgr_MTR         KSRQQGADLSSASRKSNSTVVLSRKSSS-QKSIAQPSMA----------------------
Xtr_MTR         KSSQQGSDMSTSRKSNSSTFVLSRKSSS-QKSITQPSIA----------------------
Aca_MTR         KSRQ-GCDLSVSKKSNSSTFVLSLKSSS-QRSFTQPSTA----------------------
Lch_OTR         KSRQHGCDISVSRKSNSSTFALSLKSSSSQKSITQTSMT----------------------
Pan_MTR         KSKQRSCELSSSRRGSLSTYVFSRKSSSSQKSITQTSST----------------------
Hsa_OTR         KGRR-LGETSASKKSNSSSFVLSHRSSS-QRSCSQPSTA----------------------
Mmu_OTR         KGSR-PGETSISKKSNSSTFVLSRRSSS-QRSCSQPSSA----------------------
Gga_MTR         KARP-ACELSVGRKSHSSSFVLSCRSSS-QRSLAQPPAT----------------------
Cmi_OTR         KPEPASGEITYSKKSNSSTFALSLKSSGSQKSIAVQQQQE----------------------
Lja_807         RSREADESRASDSSRRNHTFVSRLTRRSLTLSSGSQHEEASSRTTSLSPVARVPKTYFA-
Pm807           RSREADESRASDSSRRNHTFVSRLTRRSLTLSSGSQHEEASSRTTSLSPVARVPKTYFA-
Lja_3133        RLRTTG----AGHNATSQRSSTPQRHRSITLSSCSQK------------------------
Pm3133          RLETTG----AGHNATDRLSATPQRHRSVTLSSCSQK------------------------
Aoc_V1A         DFKKEDSDSSLRRTTLLTKMSNRSPTGSTGNWRELDNSPKFSVQAE--------------
Ola_V1A2        DKK-EDSDSSIRRTTLLTKMTNRSPTGSTGNWKELDNSPKNSAQAE--------------
Cne_V1A2        DYKKEDSDSSIRRTTLLTRITNRSPTGSSSNWRELDNSPKTSAQTE--------------
Xma_V1A         DYK-EDSDSSIRRTTLLTRITNRSPTGSTGNWRELDNSPKTSAQTE--------------
Dre_V1A         SFRKEDSDSSIRRTTLLTRMANRSPTCSSSTWRELQHSPKTDRSTPANT-----------
Lpl_V1A         KFRKEDSDSSLRRNTLLTRINNRSPTCSVGTWKDYDNSPKSYRSIPAES-----------
Cne_V1A1        KLKKQDSDSSIRRTTLLSRLQGPRLSKPFRELNYTIKNCPQATPAS--------------
Ola_V1A1        QLHQQDSDSSVRRTTLLSRLQGPRLSEPFRDLNLTGKSCPQVPSASYQTDLSRAES----
Cpy_V1A         NTQKEDSDSSCRRQTSFTRINNRSPTNSMEAWKESPKSIRSTRFLPIQT-----------
Tgr_V1A         NTQKEDSDSSCRRQTSFTRINNRSPTNSTDTWKESPKSIRSTRFLPIQT-----------
Hsa_V1A         KFNKEDTDSMSRRQTFYS--NNRSPTNSTGMWKDSPKSSKSIKFIPVST-----------
Mmu_V1A         KFAKDDSDSMSRRQTSYS--NNRSPTNSTGTWKDSPKSSKSIRFIPVST-----------
Gga_V1A         TLSKEDSNSNSRRQTSFT--NNRSPTHSLNTWRESP-HSKSTSFIPVPT-----------
```

```
Pan_V1A         ALHEEESDSSTRRQTSLTKMNSRSPTCSIGTWKESPKSARSIPVES--------------
Xtr_V1AR        TFSKEDSDSSTRRQTSFTRIQTRSPTNSTHTLKESPKSSKSIKFLPIQT-----------
Cmi_V1A         SLKKEDSDSITKRHTLLTRLSHRSSTLSSCNWKDTETSPQLLRFIPIET-----------
Lch_V1A         SPKKEDSDSCSRPHSISSKLSTQTQICG-TMWKESPKLNTSTPIES--------------
Aca_V1B         NFKRQFSNGSLCSRKTTILTHSPQSTPA----AAAAAGLEMQLGSLKDFYQPYEETVSDS
Gga_V1B         APRRQGSAGSLCSRRTTILSHG------------LSAGIPLHRGSGTDLSPPEEEAVTES
Cpy_V1B         SLRRQNSNGSLSSRRVTILTRIHQQT-------ISSNVRPMQSESLKDFYPPEEETVMVS
Tgr_V1B         SLRRQNSNGSLSSRRVTILTRIHQRT-------ISSNVRPMQSESLKDFYPPEEETVMVS
Lch_V1B         NLKRQGSNGSLSSRRNTLLTKISHRS-------IASNGAHMKEENFLPNMKEIYTE----
Pan_V1B         KLRRQGSNGSLSCRRNTLLTKLNH-----------------------------------
Hsa_V1B         RMRRRLSDGSLSSRHTTLLTRSSCPATLSLSLSLTLSGRPRPEESPRDLELADGEGTAET
Mmu_V1B         RVHRQLSNSSLASRRTTLLTHTCGPSTLRLSLNLSLHAKPKPAGSLKDLEQVDGEATMET
Cmi_V1B         RESSSHSPPTHLSPRYAASTAHKHHRPHPPPPEASCTVFIEENMMVSGVLPLSLA-----
Lja_2017        CRQQ---RRRCRTPTRRGAGSP-NTEHVVYALSPAPTQCRSSPAAVAT---ATPETRRSA
Pm2017          CRRQQRRRRRCRTPTRRGAGSPNTEHVVVYALSPAPTQCRSSPATAAGGTATPPETRRSA
Cin_VPR         SFTNKSRQGDTGKKTRNYSAFDPTTKETDMTATTNMELSVAKPKARNANNKKKYQTSELS
Ovu_CTR2*       VAATTETERLSLGSVRRDSRKTSDPKRISESRRISDARRISGKTQKNNSSSPRKTSDQFI
Ovu_CTR1*       VTVLSSRNDIRLMSMKKKLEQTARN-----------------------------------
Ovu_OPR*        TNTYLVELEAKKRTSFGAEHVTFASNSTARKTLNVDDTNTTALMSSSPC-----------


Aoc_V2A         --------------
Ola_V2A2        --------------
Cne_V2A         --------------
Xma_V2A1        --------------
Dre_V2A2        --------------
Dre_V2A1        --------------
Ola_V2A1        --------------
Cpy_V2A         --------------
Tgr_V2A         --------------
Xtr_V2A         --------------
Lch_V2A         --------------
Pan_V2A         --------------
Hsa_V2A         --------------
Mmu_V2A         --------------
Aca_V2A         --------------
Aoc_V2B         --------------
Xma_V2B1        --------------
Gac_V2B1        --------------
Dre_V2B1        --------------
Xma_V2B2        --------------
Cmi_V2B1        --------------
Lch_V2C         --------------
Gga_V2C         --------------
Xtr_V2C         --------------
Lpl_V2C         --------------
Cmi_V2C         --------------
Lja_644         --------------
Pm644           --------------
Lja_4769        TNELISSRTIS---
Pm4769          TNELIRSRTISA--
Gac_V2C         --------------
Xma_V2C         --------------
Dre_V2C         --------------
Ola_OTR         --------------
Xma_ITR1        --------------
Aoc_ITR         --------------
Dre_ITR         --------------
Lpl_OTR         --------------
Cne_ITR         --------------
Xma_ITR2        --------------
Tgr_MTR         --------------
Xtr_MTR         --------------
Aca_MTR         --------------
Lch_OTR         --------------
Pan_MTR         --------------
Hsa_OTR         --------------
Mmu_OTR         --------------
Gga_MTR         --------------
Cmi_OTR         --------------
Lja_807         --------------
Pm807           --------------
Lja_3133        --------------
Pm3133          --------------
Aoc_V1A         --------------
Ola_V1A2        --------------
Cne_V1A2        --------------
Xma_V1A         --------------
```

```
Dre_V1A         --------------
Lpl_V1A         --------------
Cne_V1A1        --------------
Ola_V1A1        --------------
Cpy_V1A         --------------
Tgr_V1A         --------------
Hsa_V1A         --------------
Mmu_V1A         --------------
Gga_V1A         --------------
Pan_V1A         --------------
Xtr_V1AR        --------------
Cmi_V1A         --------------
Lch_V1A         --------------
Aca_V1B         GIL-----------
Gga_V1B         GTL-----------
Cpy_V1B         GVL-----------
Tgr_V1B         GVL-----------
Lch_V1B         --------------
Pan_V1B         --------------
Hsa_V1B         IIF-----------
Mmu_V1B         SIS-----------
Cmi_V1B         --------------
Lja_2017        APLSPALP------
Pm2017          APLWSPALP-----
Cin_VPR         DIASSGLMENESEN
Ovu_CTR2*       YSDKTT--------
Ovu_CTR1*       --------------
Ovu_OPR*        --------------
```



Supplementary Figure 2-6. Phylogeny of Slit-Robo GTPase-activating proteins (srgap) and Rho GTPase-activating protein (arhgap) 4 to aid in identification of the sea lamprey (Pm) "arhgap4b" gene neighboring the Pm807 AVT receptor. Accession or UniProt number is shown near the gene. Abbreviations: Hsa-Human, Gga-Chicken, Dre-zebrafish, Xtr-African clawed frog, Lpl-Spotted gar, Pm-sea lamprey, Ola-Medaka fish, Aca-Anole lizard, Mmu-mouse, Cin-sea squirt (root).



Supplementary Figure 5-1. Fluorescein dye was dissolved in de-chlorinated tap water for normalization of gill exudate sample volumes collected for pheromone analysis **(A)** Fluorescein dye standard curve. **(B)** Timecourse showing fluorescein dye stability over a period of at least six days (144 hours) when held at 4°C in darkness.

Supplemental Table 5-1. Analytical results (ng/ml) for pheromones from *ex vivo* male lamprey gills incubated in arginine vasotocin (AVT) in buffer only (B), or with antagonist Vaptan OPC-21268 (O). Pre-incubation (P) in either buffer only or buffer and OPC-21268.

Sample ID	[AVT] nM	Gill Weight mg	3kPZS	PZS	ACA	3kACA	PADS
PB1-1	0	preinc	30.3229	437.076	5.3708	5.2833	84.5501
B1-1	0	350	127.9268	555.2571	15.1632	11.9093	71.8161
B1-2	1	380	169.5865	763.0075	20.2764	13.4707	108.116
B1-3	10	400	133.9453	590.7529	24.2535	11.7161	70.5864
B1-4	50	400	104.5753	508.7914	18.8448	10.3101	71.4408
B1-5	100	420	95.0547	406.6663	19.7927	9.2163	73.2784
PO1-1	0	preinc	70.9506	672.8325	11.3948	11.1397	133.6535
O1-1	0	350	111.8003	610.4502	16.6856	11.2178	76.2704
O1-2	1	420	120.4688	550.2345	21.6903	10.9637	86.3594
O1-3	10	330	127.8598	588.731	21.3546	9.9789	79.04
O1-4	50	390	146.9977	770.2155	23.6788	12.1351	133.3305
O1-5	100	370	330.3221	725.1079	22.8475	15.8041	121.3606
PB2-1	0	preinc	36.0924	155.9288	1.4909	2.5835	169.0301
B2-1	0	390	18.6994	162.9883	3.5716	2.2606	96.1649
B2-2	1	360	31.3794	120.1713	4.5184	1.9155	86.3131
B2-3	10	310	20.6456	102.2479	1.6869	1.1699	61.3047
B2-4	50	420	30.2644	154.0743	2.1472	1.9267	64.0368
B2-5	100	320	34.2048	109.7753	2.731	2.1528	75.12
PO2-1	0	preinc	33.4701	221.0656	2.022	4.9537	202.4177
O2-1	0	420	36.1986	162.4876	3.883	2.6935	101.7426
O2-2	1	290	36.2432	111.4672	3.0402	2.4017	77.2576
O2-3	10	340	50.214	131.647	3.3766	3.0144	116.2145
O2-4	50	390	28.1618	110.4302	3.7288	2.2642	67.6495
O2-5	100	420	33.9697	115.6901	4.2401	2.6609	68.8977
PB3-1	0	preinc	0	0	0.316	0	6.9828
B3-1	0	340	0	0	0.7859	0.017	8.3945
B3-2	1	390	0	0	0	0	3.9648
B3-3	10	390	0.0829	0	0	0	5.119
B3-4	50	400	0	0	0.0995	0	4.5915
B3-5	100	330	0	0	0	0	5.237
PO3-1	0	preinc	0	0	0	0	8.8457
O3-1	0	400	0	0	0	0	3.9658
O3-2	1	290	0	0	0.1234	0	2.9447
O3-3	10	370	0	0	0	0	1.9047
O3-4	50	360	0	0	0.0685	0	4.0861
O3-5	100	400	0.043	0	0	0	5.5666

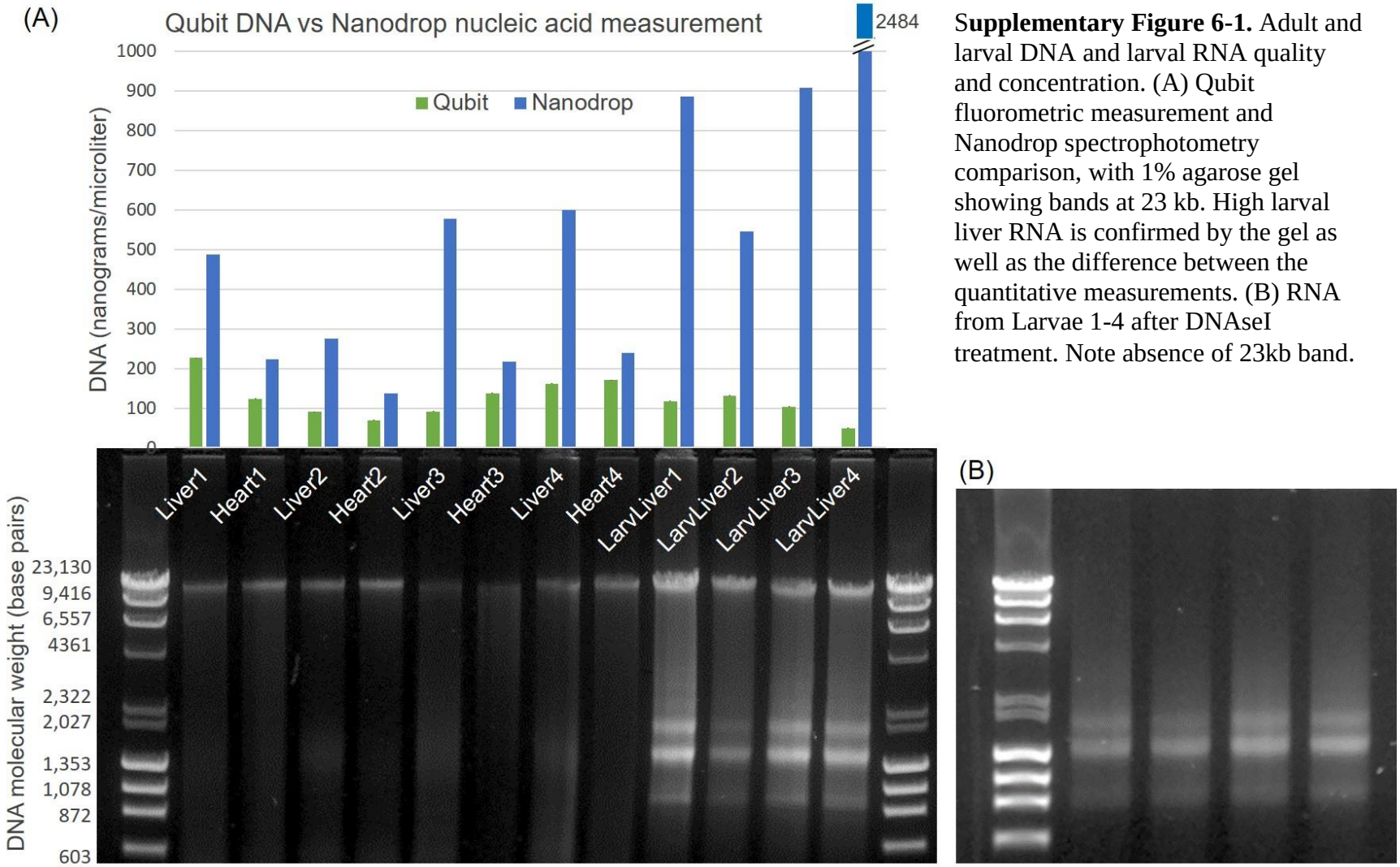
Sample ID	[AVT] nM	Gill Weight mg	3kPZS	PZS	ACA	3kACA	PADS
PB4-1	0	preinc	47.0736	241.6596	1.7675	2.1203	84.3977
B4-1	0	260	82.5728	109.58	2.5354	3.5739	28.8001
B4-2	1	270	63.7149	119.179	4.0378	4.8744	24.2023
B4-3	10	270	89.389	129.6548	4.2688	3.9725	26.5848
B4-4	50	260	72.1653	110.2935	3.4939	4.4715	27.6515
B4-5	100	210	62.713	144.8073	3.3383	3.2465	37.2509
PO4-1	0	preinc	29.7417	276.786	0.7026	1.485	61.3742
O4-1	0	300	108.961	168.4912	3.6386	5.1839	52.1726
O4-2	1	220	57.0791	125.4012	3.1936	2.5128	33.931
O4-3	10	250	69.0699	156.4248	3.626	3.3412	38.4917
O4-4	50	300	77.7611	158.7136	4.6831	5.0387	33.3741
O4-5	100	280	60.0067	160.2597	2.8837	3.58	34.1858
PB5-1	0	preinc	0.0905	16.6496	0	0	68.7873
B5-1	0	590	0.2041	5.0692	0	0	26.1798
B5-2	1	520	0.3641	9.2024	0	0	34.0033
B5-3	10	410	0.368	5.5092	0	0	54.4501
B5-4	50	550	0.0941	6.2055	0	0	23.3822
B5-5	100	510	0.0474	5.198	0	0	32.1737
PO5-1	0	preinc	0.093	18.5322	0	0	91.9712
O5-1	0	430	0.3118	6.1298	0	0	31.396
O5-2	1	500	0.3386	6.4873	0	0	47.8557
O5-3	10	530	0.3211	5.8222	0	0	27.3942
O5-4	50	530	0.4494	8.1259	0	0	30.7691
O5-5	100	540	0.3106	6.5159	0	0	37.9766
PB6-1	0	preinc	0.1647	20.4047	0	0	77.2641
B6-1	0	410	0.7025	7.148	0	0	29.4674
B6-2	1	520	0.6562	5.7238	0	0	30.7394
B6-3	10	520	0.4259	6.2125	0	0	17.055
B6-4	50	510	0.5359	7.2844	0	0	20.1223
B6-5	100	480	0.4653	7.6031	0	0	25.5457
PO6-1	0	preinc	0.4698	32.4045	0	0	124.3359
O6-1	0	370	1.5609	11.0169	0.416	0.0138	59.6354
O6-2	1	480	0.7959	9.2863	0	0	54.3988
O6-3	10	530	1.1238	10.7476	0	0	52.8092
O6-4	50	510	2.0202	8.5005	0	0	61.1508
O6-5	100	540	1.02	9.0937	0	0	51.305
PB7-1	0	preinc	0.1005	6.1317	0	0	39.2456
B7-1	0	360	0.4987	1.5971	0	0	8.0358
B7-2	1	310	0.626	1.918	0	0	18.125
B7-3	10	380	0.2859	1.7946	0		7.3023

Sample ID	[AVT] nM	Gill Weight mg	3kPZS	PZS	ACA	3kACA	PADS
B7-4	50	370	0.4601	1.4066	0	0	6.6564
B7-5	100	250	0.0727	0.6254	0	0	2.3054
PO7-1	0	preinc	0.7025	7.6651	0	0	36.4266
O7-1	0	310	0.8951	1.6082	0	0	12.3467
O7-2	1	280	0.309	1.0052	0	0	8.0307
O7-3	10	330	0.5895	1.2892	0	0	6.314
O7-4	50	370	1.0359	2.4208	0	0	15.9893
O7-5	100	330	0.9046	2.2445	0	0	13.8196

Supplemental Table 5-2. Analytical results (ng/ml) (not normalized for fluorescein) for pheromones from intact male lampreys injected with arginine vasotocin (AVT), oxytocin (OXT) or saline control. Blank space = not detected.

ACA	3kACA	PZS	3kPZ	PZ	3kPZS	PSDS	PADS	Lamprey	Treatment	Date tested	Weight
2.213	6.451	0.09	0.029	0.039	0.119	0.008	0.161	1 Before A	Saline	6/12/2014	
2.237	12.474	0.106		0.099	0.095	0.01	0.163	1 Before B			
0.122	0.414	0.12	0.023	0.097	0.035	0.006	0.121	1 Post A			
0.147	0.87	0.097		0.144	0.141	0.003	0.15	1 Post B			
		0.059	0.093		0.002	0.005	0.064	2 Before A	AVT	6/12/2014	249.4
0.027	0.535	0.078	0.039			0.005	0.036	2 Before B			
0.124		0.103	0.004		0.049		2.153	2 Post A			
0.126	0.273	0.073	0.093		0.027	0.003	1.821	2 Post B			
0.013	0.463	0.047	0.025	0.06	0.031	0	0.203	3 Before A	Saline	6/13/2014	220.9
0.016		0.073	0.025	0.157	0.019	0.002	0.227	3 Before B			
0.308	0.685	0.184	0.005	0.187	0.063	0.003	5.623	3 Post A			
0.292	0.569	0.13	0.004		0.172	0.001	7.001	3 Post B			
0.001		0.065	0.044		0.021		0.054	4 Before A	AVT	6/14/2014	318.5
	0.242	0.057	0.021	0.138	0.037		0.048	4 Before B			
0.237	0.221	0.128	0.009	0.009	0.058	0	1.183	4 Post A			
0.174	0.409	0.098	0.068	0.075	0.04	0.003	1.499	4 Post B			
0.012	0.029	0.054	0.057	0.091	0.072			5 Before A	Saline	6/16/2014	111.3
		0.062	0.017		0.018	0.003	0.019	5 Before B			
0.204		0.168	0.002		0.019	0.001	0.34	5 Post A			
0.171	0.031	0.307			0.004	0.008	0.246	5 Post B			
0.056	0.477	0.067			0.081	0.001	0.023	6 Before A	AVT	6/16/2014	224.7
0.025	0.165	0.049	0.044		0.044	0	0.018	6 Before B			
0.069	0.298	0.045	0.007		0.021	0.003	0.14	6 Post A			
0.104	0.281	0.042			0.006	0.002	0.185	6 Post B			
0.02	0.478	0.069	0.03		0.096	0.007	0.06	7 Before A	AVT	6/16/2014	246.3
0.034	0.179	0.049	0.054		0.065		0.032	7 Before B			
0.113	0.277	0.357	0.044	0.008	0.056	0.006	2.912	7 Post A			
0.086	0.142	0.249		0.081	0.001	0.005	2.802	7 Post B			
0.153	0.07	0.086	0.008	0.182	0.023		0.054	8 Before A	AVT	6/16/2014	221.5
0.222	0.862	0.171	0.007	0.035	0.056	0.001	0.064	8 Before B			
0.029	0.264	0.062	0.068	0.006	0.028		3.052	8 Post A			
0.029	0.336	0.062			0.044	0.009	2.316	8 Post B			
0.854	2.234	0.077	0.004	0.014	0.064	0.004	0.133	9 Before A	Saline	6/18/2014	218
0.908	2.799	0.072	0.007	0.164	0.151	0.004	0.139	9 Before B			
0.036	0.205	0.051	0.007		0.001			9 Post A			
0.024		0.066	0.004	0.034	0.014			9 Post B			

ACA	3kACA	PZS	3kPZ	PZ	3kPZS	PSDS	PADS	Lamprey	Treatment	Date tested	Weight
0.015	0.136	0.056	0.008		0.042		0.046	10 Before A		6/19/2014	262.9
0.036	0.418	0.037			0.052	0	0.093	10 Before B	OXT		
0.05	0.232	0.061	0.006		0.002	0.002	0.744	10 Post A			
0.036	0.339	0.046	0.078	0.038	0.003		0.71	10 Post B			
	0.846	0.037	0.002	0.149	0.005	0.005	0.005	11 Before A		6/19/2014	189.3
0.021	2.132	0.066	0.014	0.582	0.013	0.003		11 Before B	OXT		
0.046		0.133	0.005		0.002		1.366	11 Post A			
0.1	0.652	0.198	0.463	6.852	0.045		1.717	11 Post B			
0.237	0.466	0.091		0.009	0.048	0.004	0.27	12 Before A		6/19/2014	262.2
0.217	0.766	0.058	0.076	0.156	0.043	0.001	0.381	12 Before B	OXT		
0.332	0.762	0.503	0.035		0.057		6.456	12 Post A			
0.427	0.866	0.386	0.011	0.036	0.069	0.008	6.03	12 Post B			
0.171	1.433	0.284	0.017		0.014	0.007		13 Before A		6/19/2014	188.4
0.195	2.083	0.063			0.006	0.001		13 Before B	OXT		
0.038	0.649	0.122		0.011	0.03	0.011		13 Post A			
0.052	0.699	0.12			0.003		0.073	13 Post B			
		0.076	0.003	0.091	0.01			14 Before A		6/19/2014	214.6
0.006	0.048		0.011	0.016	0.002			14 Before B	OXT		
0.112	0.053	0.224			0.014	0.006	0.035	14 Post A			
0.134	0.038	0.2	0.015	0.152	0.008	0	0.087	14 Post B			
		0.078	0.008	0.077	0.005	0.003		15 Before A		6/19/2014	169.2
0.001		0.04	0.006	0.009	0.001			15 Before B	OXT		
	0.219	0.053	0.013	0.16	0.001	0	0.772	15 Post A			
0.005	0.128	0.072	0.029		0.002	0	0.885	15 Post B			
	0.202	0.065	0.007		0.012	0.001		16 Before A		6/25/2014	188.1
		0.056	0.002	0.015	0.018	0.001		16 Before B	Saline		
0.001	0.078	0.046	0.013	0.006	0.005		0.124	16 Post A			
	0.356	0.07	0.005	0.055	0.005	0.004	0.047	16 Post B			
0.035	0.029	0.06	0.014	0.008	0.003	0.003		17 Before A		41815	304.2
0.015	0.507	0.057	0.012		0.004			17 Before B	AVT		
0.011	0.14	0.064	0.008	0.01	0.006	0.001		17 Post A			
0.001	0.709	0.064	0.041	0.077	0.006			17 Post B			
0.013	0.006		0.012					18 Before A		6/25/2014	133
0.017	0.145	0.065	0.009	0.125	0	0.008	0.028	18 Before B	AVT		
0.006		0.05	0.046	0.089	0.011	0.001	0.236	18 Post A		Mean Wt	218.97
	0.258	0.063		0.008	0.002		0.282	18 Post B		± SD	52.543



Supplementary Figure 6-1. Adult and larval DNA and larval RNA quality and concentration. (A) Qubit fluorometric measurement and Nanodrop spectrophotometry comparison, with 1% agarose gel showing bands at 23 kb. High larval liver RNA is confirmed by the gel as well as the difference between the quantitative measurements. (B) RNA from Larvae 1-4 after DNAseI treatment. Note absence of 23kb band.

Supplementary Table 6-1. Genomic sequences of lamprey Pm807 putative vasotocin V1a/OXT receptor gene promoter region segments. Numbers indicate base pairs upstream of transcription start site. C-G dinucleotides (CpG sites) are highlighted in green, simple sequence repeats in gray, and 5'-untranslated region in yellow.

Segment J
-2304
GTGGGGAAGCTCCTGGAACCCGCAATTGGGCTCGAGACGGGAGGCGGAGCCAGGGCGGAGCCAGGGCGGAG
CCAGCGAGGAGTAGAATAAAAGCAGCGGAGAGCGAGAGCTCCGGGCTGCCGTTTCTGGGAGTGGACTCGTC
TCGCCGTCGCTCTCGCCTTCGACGCCTTCGTCGCTGCCGCCTTCACTCGATCATTGCGTGGCCAGACAGCT
GTTGGCAATGC -2082

Segment I
-2087
CAATGCAGTTAGGGTTGTGTCTAGGCTGCGCTGTACGGCAACGGGAGCGAGAAGGAGGCGATCCTAGGTGT
GACCGGCCGAAGGGTGTTACAAGAAAAGAAGGCGCCTCCTGCAACTACATCGTCTCCGACTTGTGCTACAC
TACATTTCTATGCACGTTGAACTTATTTAGCAGCGTACGTAGCCAACCGTGGTAGTCGGGGGTGCGGGAAA
AGGACAACAATATTAAGCGTAAGATAAGCAATTTTAAGGAAATACAAAAAGCAGTTCTAAAGAAATTGTTT
TTAAGAAGCGAATCTTAACTATACTGTGTGTGCAAT -1768

Segment H
-1847
GCAATTTTAAGGAAATACAAAAAGCAGTTCTAAAGAAATTGTTTTTAAGAAGCGAATCTTAACTATACTGT
GTGTGCAATCACCACCAACAACAATGGCCATAGCCTAATACCTCGGCATCGCGAATACATGTTTAACGTCA
CGCAAAGGTAGGCGTAAGTCTGCACATTGCATTGTTACAGTTGATGGCCCGCAGCTTTGCGTTATTGGCAT
TGTGTGGTTGGAATCGGCTCTAATGCAAGGGATGAAATGGGACCACCCTGCCCAA -1580

Segment G
-1605
GGATGAAATGGGACCACCCTGCCCAATATCCGCCCTCTGGGCCACGGGTCCCAGTGCAGTTGCACCGCCTG
CACACTCCACACCGACATTCCACAGTTGGACGCTCTGCGGATTTGTGGTAGAGGATGAGAAGCGAGCGGAA
TATTCATTGTGTGGAGCCCGGCCACTACGCACGTGGATGTGATAACTCCACCCTCCCCCGCCACAATGAGC
AGCATGCCAATTGACTCGTTCGCCAAATGTTATGTGCACGTTATGTAATGTGTAACTTTGGCAAATTATCC
CATTTTATAAGACACATTTGTTGTGTAAGAATCAGTGACGATCAAAGAAAGCTTGTGTATCCCTAAGGGAC
TGGACCTC -1243

Segment F
-1218
TCATTGTCTTGTGTGCATTTACCAACGTATCCCCCCCAGTATGTGCACATGCCTACGAAATGTGTGTGCGT
GCGTGTACGTGCGAAATGCCTGTGTGTGTATGTGCACATAGCCTGGCACAAGCGTCTGTGACGACGTCACC
GTCACTTCGTGACGAATGTTGTGTGTGTGCGTATGCCTACGAGTCTGCCAGTCTGTGTGTGTGTGTGAG
AGAAGCGGTGGCAACAGGAATGAGTGAACCACACACTGCGATCCCTGGTGCCGTGAGTTCGATTTCTGGTC
ACGGCTAAAATCAGCAGTGAATGATATC -907

Segment E
-993
AACAGGAATGAGTGAACCACACACTGCGATCCCTGGTGCCGTGAGTTCGATTTCTGGTCACGGCTAAAATC
AGCAGTGAATGATATCCGAACCGGTCCCTGGTTTCCGCCCACCCATCCATCCACCCACCCACCCCTTCAGC
AAAAGCAACAAGAATAATTGCGGTGAAGTATGGTCAAAACATAATTCCGTGACAAACTCAGCGTGGAGAGA
CTTCATGCAGTGTGTAGTGCATTGGCTAATGGCTGTGATTTATCACCATTACTATGATTAGTCGAGCGTGG
TCAAAAGTATTTCACAAACGGGTCCCTTAACAGAGTTGCATCCC -666

Segment uncharacterized (No reaction)
-748
CTGTGATTTATCACCATTACTATGATTAGTCGAGCGTGGTCAAAAGTATTTCACAAACGGGTCCCTTAACA
GAGTTGCATCCCCATAATTTCCACGCCTCTGCGCCGTGTCTCTTCATGGATTATATGCACCGATGTTTAGT
TTGTGGATGCATTTTCCCTTTCACTCTCAAATGCAAAAATAAAACAATGCACATACATCTCAAACTACAAT
ATAGATCTTCTCCCCGTATCGTTTTTTTTAACGACGATATTTGTTTCGGTTTGACAAAACAAACTTAGGAA
AC -463

Segment D
-488
GGTTTGACAAAACAAACTTAGGAAACACTTCACAAAAATATCTCTGAAGAAACCCCATTGTGATCGATGCG
GTGACAGCAGAGAATGAAAAAAAGCGACAACTTTATAGACAAAAACCACTCAACAAAATGTGTAAAACGAA
GCGCAGCAGTTTTTGTGCCAAGATTATATTTAGTTGCAAAGTTAGTTGCAAAGTTTGCC -287

Segments A-B-C
-314
AGTTGCAAAGTTAGTTGCAAAGTTTGCCCGAGTGCGCGCGGCGAGCGAGGCGTGAGTGGTTGAAATCAGCA
GTCTCGGAGTTGGTGATGAGTGAGCGAGAGGCAAGGGACGTTTATAATGAGCCGCAACGCGAAGTTGCAGA
GCGCAATCTTCAGACGCGATGTTGGTGGCGCGTGTGTTCGGGAAAGTGGGGTGGGATGTGGGGGGGAGGGG
GGTGGTGCGTGACCGAGCGAGGTAGGTGTGACGTTTAGGAGGAGGGATGGGTGAGGAGTGGTGGTGGTGGT
GGGTGGTGCGGGAGTGGTGGGTGGTGCGGGagagatgaaaacggagtggtggtggtgggagagacaacgcc
agcctataacatacacagcacggctccctatctgcagg +79