ABSTRACT

SCHILLING, JUSTIN DALE. Proteomic and Machine Learning Analyses of White Perch (*Morone americana*) Plasma and Ovary. (Under the direction of Harry V. Daniels III).

Three studies were conducted to characterize the plasma and ovary proteomes of and vitellogenesis in white perch (*Morone americana*). First, a discovery proteomics study employed a simple fractionation method and filter-aided sample preparation (FASP) to characterize the cytosolic and membrane fractions of white perch ovary tissues by semiquantitative tandem mass spectrometry (MS/MS). A total of 882 unique proteins, 114 found only in the cytosolic fraction, 169 proteins found only in the membrane fraction, and 599 found in both fractions, were identified using the striped bass ovary transcriptome as the reference database. Of these 882 proteins, a majority was from mitochondria and other membrane-bounded organelles. Support vector machines (SVMs) were able to perfectly (100% correct) classify samples as either membrane or cytosolic fraction during cross-validation based on the expression data of 242 proteins with the highest ANOVA p-values (i.e. those that were not significantly enriched in either fraction) as measured by MS/MS. SVMs offer categorical classification of proteomics data superior to that of parametric statistical methods such as analysis of variance (ANOVA).

A second study was carried out in order to measure differences within and between the global plasma proteomes of sexually mature male and female white perch before (Initial Control, IC) and after 17β-estradiol (E_2) induction. Semiquantitative protein expression data were analyzed by SVMs and by two-way ANOVA. The expression levels of 44, 77, and 57 proteins varied significantly ($p \leq 0.05$) by gender (male/female), treatment (IC/E_2), and the interaction of gender and treatment, respectively. SVMs were able to perfectly (100% correct) classify male and female perch IC and E_2-induced plasma samples during cross-

validation using the MS/MS protein expression data. Following E_2 induction, both the male and female perch plasma proteomes contained significantly higher levels of vitellogenin Aa and Ab (VtgAa, VtgAb), latrophilin and seven transmembrane domain-containing protein 1 (Eltd1), and kininogen 1 (Kng1) than IC plasmas. This is the first report that Eltd1 and Kng1 may be E_2-responsive proteins in fishes.

In the third study, selected reaction monitoring (SRM) tandem mass spectrometry was employed to accurately quantify the three white perch vitellogenins (VtgAa, VtgAb, VtgC) in the liver, plasma, and ovary during pre-, early-, mid-, and post-vitellogenic oocyte growth. Both SRM and immunohistochemistry confirmed that VtgC is present in the pre-vitellogenic perch ovary and was the only quantifiable vitellogenin within the ovary at this time point. Only VtgAb could be confidently quantified in the pre-vitellogenic perch liver. VtgAb was found to be the predominant Vtg during vitellogenesis in perch liver, plasma, and ovary.

Observed differences in the proportional accumulation of Vtgs suggest that there is considerable plasticity in the Vtg-Vtgr system in white perch that may allow fine-tuning of egg buoyancy based upon the salinity of the estuarine water into which white perch and striped bass spawn their eggs. It appears that VtgC is the most variable form of vitellogenin within the post-vitellogenic oocytes of Acanthomorph fishes, ranging from ~2.5% in perch to 26% in striped bass, suggesting that VtgC composition may relate to other aspects of early life history in these fishes. While the implications upon early life history remain to be fully elucidated, the proportion of VtgC in post-vitellogenic oocytes varies considerably among Acathomorph fishes, even between closely related species of the genus *Morone*.

As a result of these studies, the molecular identities of the white perch vitellogenin receptors, LR8 and Lrp13, have been confirmed. Additionally, these studies demonstrated the utility of machine learning in general, and SVMs in particular, for categorical

classification of proteomics data, and also paved the way for future investigation of Eltd1 and Kng1 as potential biomarkers of EDC exposure in vertebrates.

Proteomic and Machine Learning Analyses of White Perch (*Morone americana*) Plasma and Ovary

by
Justin Dale Schilling

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Zoology

Raleigh, North Carolina

2015

APPROVED BY:

_____	_____
Harry V. Daniels Committee Chair	Robert M. Kelly
_____	_____
Jeffrey A. Yoder	Antonio Planchart

BIOGRAPHY

After graduating from the Pennsylvania State University with a B.A. in Classics and Ancient Mediterranean Studies, a B.S. in Biology, and a minor in International Studies, Justin began eight years of work in the biotechnology industry. In 2008, Justin transitioned back into academia as a Research Specialist in a laboratory at the University of Michigan specializing in negative stain and cryo-electron microscopy single particle analysis of macromolecular protein complexes. After two fruitful years, Justin moved to Raleigh, NC where he began a doctoral program in Zoology at the North Carolina State University in August of 2010. Justin married Emily Sluzas in July of 2014. After graduation, Justin will join his wife in their native city of Philadelphia, where she is a medical student.

ACKNOWLEDGMENTS

I thank my Doctoral Advisory Committee, Drs. Harry V. Daniels (Major Professor), Robert M. Kelly, Jeffrey A. Yoder, and Antonio Planchart. I am especially grateful for the generous support of Dr. Steven Lommel, associate dean and director of the N.C. Agricultural Research Service in N.C. State University's College of Agriculture and Life Sciences. I thank the staff, students, and colleagues from both the North Carolina State University Departments of Applied Ecology and Chemistry and the University of Michigan Life Sciences Institute for all of the assistance and guidance they have provided:

Dr. Benjamin J. Reading

Dr. David C. Muddiman

Dr. Angelito I. Neopmuceno

Mr. Philip L. Loziuk

Dr. Naoshi Hiramatsu

Dr. Valerie N. Williams

Dr. Kevin Gross

Dr. Georgios Skiniotis

Dr. William Clay Brown

Dr. Min Su

Mr. Michael Hopper

Dr. Andy McGinty

Dr. Scott A. Salger

Mrs. Susan Marschalk

Mr. Tyler Nethers, Mr. Brad Ring, and Mr. John Davis

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1

Vertebrate Vitellogenins and Their Receptors: A Comprehensive Review of the Literature and Recent Insights

Justin Schilling[1]

[1]Department of Applied Ecology, North Carolina State University, Raleigh, North Carolina 27695, United States

Correspondence: Justin Schilling, 127 David Clark Laboratories, Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695-7617, USA Telephone: 919-515-3830

E-mail: jdschill@ncsu.edu

Abbreviations: Vtg(**s**): vitellogenin(s); **Vtgr**(**s**): vitellogenin receptor(s); **LvH**: lipovitellin heavy chain; **LvL**: lipovitellin light chain; **β'-c**: beta component; **CT**: C-terminal domain; **LLT**: Large lipid transfer domain; **LLTP**: large lipid transfer protein; **LDL**: low-density lipoprotein; **LDLR**: low density lipoprotein receptor; **VLDL**: very low-density lipoprotein; **VLDLR**: very low-density lipoprotein receptor; **LBR**: ligand-binding repeat; **AP-MS/MS**: affinity-purification and tandem mass spectrometry; **PV**: Phosvitin domain; **VRBP**: vitellogenin receptor binding protein; **LvA**: VtgAa lipovitellin; **LvB**: VtgAa lipovitellin; **LvC**: VtgAa lipovitellin.

Justin Schilling (JS) was the primary author of this manuscript, which was a collaborative effort that also involved Naoshi Hiramatsu (NH) and Benjamin J. Reading (BJR). Figure 1 was composed by BJR and NH based upon data collected by both individuals. Figure 3 was composed by JS and BJR based upon data collected by both individuals. Figure 4 was composed by JS and BJR based upon data collected by both individuals. Figure 5 was composed by NH, JS, and BJR based upon data collected by all three individuals. Figure 6 was composed by BJR and NH. Figure 7 was composed by JS and BJR. Figure 8 was composed by JS and BJR.

Vitellogenins and their receptors have crucial and diverse physiological roles

Vitellogenins (Vtgs) and their receptors (Vtgrs) are crucial components of a dynamic receptor-mediated system that underlies egg buoyancy and the deposition of nutrients and structural components in growing oocytes of oviparous vertebrates. Produced by the liver in response to estrogen, Vtgs are secreted as 300-600 kDa homodimers into the bloodstream and then bind Vtgrs that are expressed exclusively on the surface of growing oocytes. These Vtgrs mediate endocytosis of Vtgs into the ooplasm where they are processed into yolk proteins and deliver structural and metabolic substrates required during embryogenesis.

Vtgs are members of the large lipid transfer (LLT) protein (LLTP) superfamily, which includes insect apolipoprotein [apolipophorin-II/I (apoLp-II/I)], microsomal triglyceride transfer protein (MTP), and apolipoproteins B and E (apoB, apoE). Vtgs and other LLTPs play crucial roles in immunity, reproduction, development, aging, lifespan regulation, and diseases in vertebrates (Brandt, Zwaan, Beekman, Westendorp, & Slagboom, 2005). Dysfunctional lipidation, receptor binding, and lipid cargo processing of apoB can lead to metabolic disorders such as atherosclerosis, obesity, and type 2 diabetes (Boren et al., 1998; Mahley & Rall, 2000; Olofsson & Boren, 2005; Taskinen, 2005).

While all LLTPs bind and transport lipids, their cargoes can differ in composition. For instance, apoB and apoLp-II/I primarily bind neutral lipids, while MTP and Vtg predominantly bind phospholipids. In addition, these lipid cargoes are delivered to different targets and contribute to different cellular and developmental pathways. For instance, vertebrate apoB is largely responsible for postprandial lipid transport and distribution, while apoLp-II/I plays an additional role in transport of stored lipids in insects (Smolenaars, Madsen, Rodenburg, & Van der Horst, 2006). MTPs are present in both invertebrates and vertebrates and play a role in assembly, lipidation, and secretion of all other LLTPs (Smolenaars et al., 2006). Vtgs

are complex lipoprotein yolk precursors that are loaded primarily with phospholipids in addition to carbohydrates, vitamins, and metal ions that are used as structural and metabolic substrates for the embryo before it can feed on its own (Finn, 2007a). As a shuttle of metal ions, Vtgs play a central role in embryonic trace mineral homeostasis in oviparous vertebrates (Montorzi, Falchuk, & Vallee, 1995; Richards, 1997).

Vtgrs are members of the lipoprotein receptor superfamily and are similarly conserved across phyla. Recent studies have indicated that differential Vtg expression, accumulation, and processing affect egg buoyancy and viability, yet the underlying mechanisms remain unknown (Finn, 2007c; 2007a; Finn, Kolarevic, Kongshaug, & Nilsen, 2009). Thus, given the diverse and significant physiological roles of the Vtg-Vtgr system, a better understanding of their molecular underpinnings will have direct relevance to not only animal health and egg quality but to human health as well, given the high degree of conservation among LLTPs and their receptors.

Structural anatomy of the Vtg-Vtgr system

Vtgs are grouped into complete and incomplete types. From the N-terminus, a complete type vertebrate Vtg consists of a signal peptide, a lipovitellin heavy chain (LvH), a phosvitin (Pv), a lipovitellin light chain (LvL), and a von Willebrand factor type D domain (Vwfd) that is cleaved into a beta component (β'-c) and a C-terminal coding region (CT) in teleosts, while an incomplete type Vtg lacks Pv and/or CT domains.

Acanthomorph fishes have at least three distinct forms of Vtg: VtgAa, VtgAb, and VtgC. VtgAa and VtgAb are complete types while VtgC is an incomplete type, lacking Pv, β'-c, and CT domains and any apparent receptor-binding site within its LvH (Reading et al., 2009; Reading and Sullivan, 2011). The Pv domains of complete type Vtgs are comprised

largely of serine residues that can bind metal ions and transport them into growing oocytes (Finn, 2007b; Ghosh & Thomas, 1995). Within oocytes, phosphates play the aforementioned role of enhancing Vtg plasma solubility, but have also been shown to contribute to the osmolyte pool responsible for governing oocyte hydration (Finn, 2007c). Trace minerals play a central role in embryogenesis, serving as catalytic or structural enzyme co-factors (Rowe & Eckhert, 1999).

In response to changes in photoperiod and temperature, complete type Vtgs are secreted into the bloodstream and bind ovarian Vtgrs as 400-600 kDa homodimers. A crystal structure of a complete type lamprey vitellogenin exists (Fig. 2) (Anderson, Levitt, & Banaszak, 1998; Finn, 2007c; Raag, Appelt, Xuong, & Banaszak, 1988). Depicted as a monomer, the large lipid binding pocket is clearly visible. The putative receptor binding peptide is located in an amphipathic α-helix nested within a β sheet and presents basic residues (lysine [K], arginine [R]) to acidic ligand binding residues (aspartic acid [D], glutamic acid [E]) (See Fig. 2) on Vtgrs (Li, Sadasivam, & Ding, 2003). Vtgrs belong to the low density lipoprotein receptor (LDLR) related proteins (LRPs) superfamily (Smolenaars et al., 2006; Van der Horst, Roosendaal, & Rodenburg, 2009).

We have characterized all three Vtg forms in white perch (*Morone americana*) and have begun to examine their interactions with Vtgrs on the surface of growing oocytes (Hiramatsu, Hara, et al., 2002a; Reading, Hiramatsu, & Sullivan, 2011). To date, we have fully characterized two ovarian Vtgrs, LR8 (Hiramatsu, Chapman, Lindzey, Haynes, & Sullivan, 2004a) and Lrp13 (Reading et al., 2014). Homologous to Vtgr in other egg-laying vertebrates and the very low-density lipoprotein receptor (VLDLR) in mammals, white perch LR8 is a 93 kDa type-I membrane protein bearing eight N-terminal ligand binding repeats. While apparent splice variants of transcripts that encode receptor proteins with or without an

O-linked sugar domain exist in other species, white perch LR8 has only been identified without an *O*-linked sugar domain (Finn, 2007a; Hiramatsu, Chapman, Lindzey, Haynes, & Sullivan, 2004b; Hiramatsu, Matsubara, et al., 2002b).

Recently, we used affinity purification coupled to tandem mass spectrometry (AP-MS/MS) (Fig. 3) and Western blotting (Fig. 4) to identify and characterize the molecular structure of an additional Vtgr from white perch ovarian tissue, Lrp13 (Reading et al., 2014). Possessing additional LDLR Class B YWxD (LDLb), epidermal growth factor-like precursor (EGF), and LDLR Class A ligand binding repeat (LDLa) motifs, this receptor appears to represent a novel class of vertebrate lipoprotein receptor (Fig. 5). An additional distinguishing feature of Lrp13 is that its cytoplasmic tail lacks any apparent endocytosis signal, raising the question of how this liganded-receptor complex is internalized, compartmentalized, and processed. Intriguingly, *in silico* analysis of the protein sequence indicates several potential phosphorylation sites and kinase binding sites on the intracellular domain of Lrp13 that may be related to its localization and trafficking, which may be different from that of the LR8 receptor.

Physiological significance of differential Vtg expression, uptake, and processing

It has been proposed that the evolution of multiple complete type Vtgs (e.g., VtgAa and VtgAb) enabled teleosts to successfully re-colonize marine environments (Finn, 2007a). Differential Vtg expression and proteolysis is correlated to the type of egg spawned, e.g., benthic or pelagic (Finn & Kristoffersen, 2007). In particular, VtgAa has been shown to undergo extensive proteolysis into free amino acids (FAA) within pelagic teleost oocytes (Finn, 2007c; Finn & Kristoffersen, 2007). These FAA serve as osmolytes and drive the influx of water via aquaporins. The resulting hydration also provides marine teleost embryos

with a source of water until osmoregulatory and drinking mechanisms are fully developed (Finn, 2007a; 2007c; Finn & Kristoffersen, 2007). However, it is likely that concordant neofunctionalization of Vtgrs was required as well, although receptor specificity and multiplicity have not been fully addressed in Vtg-Vtgr studies to date (Dieckmann, Dietrich, & Herz, 2010).

Structural and functional conservation among LLTPs and their receptors

A recent comparative genomic and phylogenetic study has placed Vtgs as a basal member of the LLTP superfamily (Hayward, Takahashi, Bendena, Tobe, & Hui, 2010). Vtg is ancestral to and shares considerable structural homology with apoB and apoE and LR8 shares a similar relationship with VLDLR and LDLR (Babin, Bogerd, Kooiman, Van Marrewijk, & Van der Horst, 1999; Babin, Deryckere, & Gannon, 1995; Baker, 1988; Durliat, Andre, & Babin, 2000; Mahley, Innerarity, Rall, & Weisgraber, 1984; Schneider, 2009; S. Takahashi et al., 2004; Y. Takahashi, Itoh, Oohashi, & Miyamoto, 2003; Wahli, 1988). This high degree of conservation allows us to use apolipoproteins and their receptors as models to inform and guide our investigations of the dynamic interactions between Vtgs and Vtgrs.

The receptor-binding motif of human apoE consists of a 25 amino acid peptide rich in basic (K and R) residues that recognizes VLDLR and LDLR (Innerarity, Friedlander, Rall, Weisgraber, & Mahley, 1983; Lalazar et al., 1988). Mutations of these key residues diminish or abolish receptor binding (Davis et al., 2007; Dyer & Curtiss, 1991; Innerarity et al., 1983; Lalazar et al., 1988; Mahley & Rall, 2000; Plack & Pritchard, 1968; Raffaï et al., 2000; Weisgraber, Innerarity, & Mahley, 1978; Zaiou et al., 2000). As in Vtgs, both the apoB and apoE receptor-binding domains form an amphipathic α-helix, with hydrophobic residues isolated and the basic amino acids clustered on opposite faces (Rall, Weisgraber, Innerarity,

& Mahley, 1982; Weisgraber & Mahley, 1996; Wilson, Wardell, Weisgraber, Mahley, & Agard, 1991). This secondary structure is required for bioactivity (Nikoulin & Curtiss, 1998; Zaiou et al., 2000). The native receptor-binding domain is stabilized by an adjacent amphipathic α-helix that does not interact directly with the receptor (Weisgraber & Mahley, 1996). This structural arrangement can be mimicked by synthetic peptides comprising the receptor-binding motif of apoB (11 amino acids) or apoE (15 amino acids) which have full binding activity for the LDLR (Cardin, Bowlin, & Krstenansky, 1988; Clay, Anantharamaiah, Mistry, Balasubramaniam, & Harmony, 1995; Dyer & Curtiss, 1991; Dyer, Smith, & Curtiss, 1991; Segrest et al., 1998). The corresponding synthetic monomer is unable to bind LDLR. The sequences flanking these native receptor-binding peptides are thought to be required for it to assume a proper 3D conformation. Therefore, the receptor-binding domains of apoB and apoE both utilize amphipathic structures to present tight clusters of homologous basic residues in a specific surface array as the basis for receptor recognition.

The receptor-binding sites of teleost complete type Vtgs resemble that of apoB. Chicken egg yolk very low-density lipoprotein (VLDL) was just as effective as native white perch Vtg in displacing ^{125}I-Vtg from perch ovarian membranes (Tao, Berlinsky, & Sullivan, 1996). Chickens incorporate apoB into VLDL but they do not produce apoE. The laying hen expresses a ~95 kDa receptor on the oocyte surface that binds Vtgs and VLDL and mediates their uptake by endocytosis (Hayashi, Ando, Stifani, & Schneider, 1989). This receptor also binds human apoE incorporated into lipid vesicles with an affinity similar to that which it exhibits for native complete type Vtgs (Steyrer, Barber, & Schneider, 1990). Collectively, these observations indicate that the receptor-binding motifs of complete type Vtgs are a simple, amphipathic, lysine- and arginine-rich polypeptide, like the receptor-binding motifs of apoB and apoE. One or both of the two disulfide bonds in the vitellogenin

receptor binding protein (VRBP), which are conserved in all chordates, seem to be essential for full receptor-binding activity. This expectation was confirmed when the VRBP was localized to an 85-residue stretch of the Lv domain of teleost Vtgs (Fig. 6) (Li, Sadasivam, & Ding, 2003).

The gene encoding LR8 belongs to a diverse superfamily that encompasses several lipoprotein receptors, including LDLR and VLDLR (H Bujo, 1994; Jingami & Yamamoto, 1995). The cysteine-rich ligand-binding domain of LDLR is comprised of seven ligand-binding repeats (LBRs), whereas this domain within VLDLR and LR8 consists of eight such LBRs. The high cysteine content of these receptors is responsible for the proper folding of this domain into a rigid structure that presents clusters of acidic ligand-binding residues located at the C-termini of each LBR. These negatively charged residues may interact with positively charged residues (K, R) in the receptor-binding sites of apoB, apoE, and Vtg (Stifani, Menn, Rodriguez, & Schneider, 1990). It remains unknown whether or not other negatively charged residues distributed among the LBRs play a role in ligand binding.

The Vtgr from chicken, *Xenopus*, and rainbow trout (*Oncorhynchus mykiss*) are homologs of the mammalian VLDLR (Cardinaux, Chapel, & Wahli, 1994; H Bujo, 1994; Prat, Coward, Sumpter, & Tyler, 1998). The deduced amino acid sequence of the white perch Vtgr, LR8, (~93 kDa; 844 residues) shares an average of ~74% identity with other known vertebrate Vtgrs or VLDLRs and has the same general domain structure, including an 8-repeat ligand-binding domain. With this degree of structural conservation, it is reasonable to expect that the general requirements for ligand binding of the Vtgr involve physical features common to Vtg and related apoproteins. Slight structural differences, however, must exist that confer ligand specificity among lipoprotein receptors. For example, in the rainbow trout, which expresses all lipoproteins found in mammals, the Vtgr is highly specific for Vtg (Babin,

2008; Babin et al., 1995; 1999; Babin & Vernier, 1989; Babin et al., 1997; Tao et al., 1996). Among lipoprotein receptors, ligand binding occurs through different LBRs. For example, LBR 5 is essential for binding of apoE, LBRs 2-7 cooperatively bind apoB, and repeats 1-3 are responsible for binding Vtg (Li, Sadasivam, & Ding, 2003a; Russell, Brown, & Goldstein, 1989). Vertebrate Vtgrs share an underlying ligand recognition mechanism as evidenced by the Vtgrs of chicken, *Xenopus*, and fish binding each other's complete type Vtgs (Stifani et al., 1988,1990ab; LeMenn and Nunez-Rodriguez, 1991; Tyler and Lancaster 1993; Berlinsky et al., 1995; Tao et al., 1996). Additionally, our laboratory has shown that complete type Vtgs from five species of fishes both closely and distantly related to white perch bound to the perch LR8 (Hiramatsu et al., 2002a). This mechanism extends to receptor-mediated uptake of Vtg by the oocytes. For example, killifish (*Fundulus sp*.) oocytes bind *Xenopus* complete type Vtgs *in vitro*, a process that can be inhibited by the addition of complete type Vtgs from either *Fundulus* or *Xenopus* (Kanungo, Petrino, & Wallace, 1990). Among all oviparous vertebrates, both the receptor-binding motif of complete type Vtgs and the ligand binding sites of Vtgrs have changed little during the course of evolution.

Vitellogenin polypeptide sequences (N=70) from an array of fishes were aligned by ClustalW to generate a dendrogram (Fig. 7). While approved gene names vary, it is clear that vitellogenins group by type and that the Aa and Ab types are present in Acanthomorph fishes which represent more derived teleosts.

Vtg lipidation

Despite evidence that LLTP lipidation dramatically affects receptor binding, type-specific Vtg cargoes have not yet been characterized in any vertebrate species (Aviram, Lund-Katz, Phillips, & Chait, 1988; Finn, 2007c). Lipid cargo size and composition have

been shown to have dramatic effects on LLTP size, shape, surface charge, and availability of crucial receptor binding residues on the particle surface (Aviram et al., 1988; McKeone, Patsch, & Pownall, 1993; Ren et al., 2010; Rowe & Eckhert, 1999).

Given the high degree of conservation among LLTPs and their receptors, apolipoproteins and their receptors can serve as models for the Vtg-Vtgr system. For instance, the crucial role of basic residues (lysine and arginine) in receptor binding of plasma lipoproteins was established three decades ago and subsequently confirmed in Vtgs (Finn, 2007a; 2007c; Finn et al., 2009; Innerarity et al., 1983; Li, Sadasivam, & Ding, 2003; Weisgraber et al., 1978). Additionally, core triglyceride content of LDL particles has been shown to affect the structure of apoB and its interactions with LDLR (Anderson et al., 1998; Aviram et al., 1988; Finn, 2007c; McKeone et al., 1993; Raag et al., 1988). Variable apolipoprotein lipidation results in changes in overall particle size, surface charge, and in the local conformation of basic lysine residues of apoB on the LDL surface that constitute the receptor binding domain (Aviram et al., 1988; Li, Sadasivam, & Ding, 2003; Nigon, Lesnik, Rouis, & Chapman, 1991). LDL particles of intermediate size bind LDLR with higher affinity than both smaller and larger particles (Nigon et al., 1991; Smolenaars et al., 2006; Van der Horst et al., 2009).

Comprising up to 20% of Vtg mass, lipids are the dominant Vtg cargo. The majority of these lipids are phospholipids that can be utilized as structural components during embryogenesis, while neutral lipids comprise a smaller portion of the cargo and can be used as energy substrates. Within Vtg molecules, the LvH domain comprises the majority of the lipid binding pocket. The amphipathic polyserine Pv domain has been implicated to play a role in enhancing Vtg plasma solubility and is predicted to contain a ‘lipid-lock’ homologous to that of apoB (Dashti, Gandhi, Liu, Lin, & Segrest, 2002; Hiramatsu, Hara, et al., 2002a;

Reading et al., 2011; Richardson et al., 2005). With the data currently available, Pv domain size does not appear to correlate with lipid cargo size, however, a relationship between Pv domain size and lipid type has been suggested (Finn, 2007a; Hiramatsu, Chapman, Lindzey, Haynes, & Sullivan, 2004a; Silversand & Haux, 1995).

Vtg lipid cargo is comprised of polar lipids (~70-80%), specifically phophatidylcholine (PC), along with neutral lipids (~20-30%), primarily triacylglycerides (TAG) (Finn, 2007a; Hiramatsu, Chapman, Lindzey, Haynes, & Sullivan, 2004b; Hiramatsu, Matsubara, et al., 2002b). Relatively short Pv domains appear to be associated with eggs that lack large oil globules and contain a high percentage of PC, while longer Pv domains are primarily found in fishes that spawn eggs with large oil droplets that contain a higher percentage of neutral lipids(Anderson et al., 1998; Finn, 2007a; Finn & Kristoffersen, 2007; Finn, Fyhn, Henderson, & Evjen, 1996; N & T, 2002; Norberg & Haux, 1985; Ohkubo, Sawaguchi, Hamatsu, & Matsubara, 2006; Sawaguchi, Ohkubo, Koya, & Matsubara, 2005; Silversand & Haux, 1995). While additional investigation is needed, given its highly amphipathic nature it has been suggested that Pv aids in shielding neutral lipids, maintaing Vtg plasma solubility, and that larger Pv domains might be required with larger neutral lipid cargoes (Finn, 2007a). Pv and LvH domain sizes are positively correlated and seem to support the notion that neutral lipids require both larger Pv shields and larger binding pockets (Finn, 2007a; Finn & Kristoffersen, 2007). The incomplete type VtgC, however, also transports lipids despite lacking a Pv domain, yet VtgC lipid cargoes comprise a smaller proportion of total protein mass (~13-14%) compared to complete type Vtgs (16-21%) (Finn, 2007a; 2007c; Finn & Kristoffersen, 2007; Sawaguchi, Ohkubo, Koya, & Matsubara, 2005; Silversand & Haux, 1995). How VtgC is taken up by oocytes remains unknown. While further investigation is

required, each Vtg type may be loaded with a distinct lipid cargo that in turn could underlie receptor specificity and subsequent differential processing.

Implications of proportional vitellogenin accumulation

The proportional Vtg ratios within post-vitellogenic oocytes has been measured for barfin flounder, striped bass, and white perch (Sawaguchi et al., 2008; Williams et al., 2014). Barfin flounder spawn pelagic eggs in saltwater that have a post-vitellogenic proportional ratio of LvA : LvB : LvC of about 9 : 15 : 1 (Sawaguchi et al., 2008). Striped bass spawn their eggs with a post-vitellogenic proportional ratio of 1.4 : 1.4 : 1 in fresh or brackish water (Williams et al., 2014). Interestingly, the proportional abundance of the adhesive eggs spawned by white perch in fresh or brackish water as measured by selected reaction monitoring tandem mass spectrometry is 5 : 35.7 : 1.

While the proportional Vtg ratios do not appear to correlate to egg buoyancy or salinity, they may relate to other aspects of early life history. Considering the data that are currently available, it appears that VtgC is the most variable form of vitellogenin within the post-vitellogenic oocyte, ranging from ~2.5% in perch to 26% in striped bass (Fig. 8) (Schilling et al., 2014; Williams et al., 2014). The ratio of complete type Vtgs, on the other hand, appears to fall within a much narrower range of 1 : 1 to 1 : 1.7 for VtgAa : VtgAb, respectively.

The three Perciform Vtgs are known to undergo different degrees of proteolysis within growing oocytes (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005). While VtgAa and, to a lesser extent, VtgAb are proteolyzed into free amino acids or small peptides, VtgC appears to remain largely intact through oocyte maturation and ovulation and remains available to developing embryos as a food source before they are able to feed on their own.

In mosqitofish, for example, the VtgC-derived yolk proteins are the last components of the egg yolk that remain to be consumed by yolk-sac fry (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005).

Depending upon temperature, fertilized white perch and striped bass larvae both hatch around 48 hrs post-fertilization (Eldridge, Whipple, Eng, Bowers, & Jarvis, 1981; Mansueti, 1964; North & Houde, 2003). When food is then restricted or deprived, white perch larvae survive for up to 15 days post-fertilization while striped bass larvae survive for up to 31 days post-fertilization (Fig. 9) (Eldridge et al., 1981; Mansueti, 1964). Additionally, the time to first feeding of these two species differs, with the white perch and the striped bass larvae beginning to feed at 3-5 days and 7-9 days post-fertilization, respectively. Therefore, although both of these closely related species share similar time frames during the earliest stages of development (i.e. from fertilization to hatch), the striped bass larvae appear to have an extended developmental window from hatching to first feeding and also these larvae have yolk stores that allow them to survive in the absence of food for twice as long as white perch after hatch. This disparity in early developmental stages post-hatch may relate to differences in VtgC yolk content of the white perch and striped bass eggs, which is 2.5-5% and 26%, respectively.

Summary and organization of dissertation

Despite intensive studies over the past two decades, the mechanisms underlying differential receptor-mediated Vtg uptake, compartmentalization, and processing in oocytes of oviparous vertebrates remain largely unexplained. In addition, the ovary proteome of white perch has remained similarly uncharacterized. In the following chapters, experiments aimed at further describing the Vtg-Vtgr system are described. The following chapters detail

experiments that utilized methods such as protein expression and purification, nanoLC-MS/MS, and AP-MS/MS to further characterize white perch liver, plasma, and ovary tissues.

CHAPTER 2: Machine learning support vector machines are highly effective at discerning the sub-cellular fraction (i.e., cytosolic or membrane) from which a protein originated based upon nanoLC-MS/MS data. The simple fractionation method utilized in this study effectively revealed an abundance of mitochondrial proteins.

CHAPTER 3: Male and female white perch have sex-specific plasma protein profiles both before and after induction with estradiol-17β (E_2) as analyzed by machine learning support vector machines. In addition, the relatively uncharacterized latrophilin and seven transmembrane domain-containing protein 1 (Eltd1), and kininogen 1 (Kng1) are shown to be E_2-responsive proteins in white perch.

CHAPTER 4: Selected reaction monitoring (SRM) absolute quantification tandem mass spectrometry indicates that 1) VtgAb is the predominant Vtg in white perch and the only Vtg detectable in the pre-vitellogenic liver, 2) of the three Vtgs, only VtgC can be detected in the pre-vitellogenic white perch ovary by SRM, and 3) immunohistochemistry also indicates that VtgC is present in the pre-vitellogenic white perch ovary and localizes exclusively to lipid droplets in vitellogenic oocytes.

All chapters are formatted for publication. **CHAPTER 2** was published in the *Journal of Proteome Research* in 2014 and is included herein with permission of the American

Chemical Society. As of March 3 2015, **CHAPTER 3** is under a second round of review at *PROTEOMICS*.

References

Anderson, T. A., Levitt, D. G., & Banaszak, L. J. (1998). The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure*, *6*(7), 895–909.

Aviram, M., Lund-Katz, S., Phillips, M. C., & Chait, A. (1988). The influence of the triglyceride content of low density lipoprotein on the interaction of apolipoprotein B-100 with cells. *J. Biol. Chem.*, *263*(32), 16842–16847.

Babin, P. J. (2008). Conservation of a vitellogenin gene cluster in oviparous vertebrates and identification of its traces in the platypus genome. *Gene*, *413*(1-2), 76–82. doi:10.1016/j.gene.2008.02.001

Babin, P. J., & Vernier, J. M. (1989). Plasma lipoproteins in fish. *J. Lipid Res.*, *30*(4), 467–489.

Babin, P. J., Bogerd, J., Kooiman, F. P., Van Marrewijk, W. J. A., & Van der Horst, D. J. (1999). Apolipophorin II/I, Apolipoprotein B, Vitellogenin, and Microsomal Triglyceride Transfer Protein Genes Are Derived from a Common Ancestor. *J. Mol. Evo.*, *49*(1), 150–160. doi:10.1007/PL00006528

Babin, P. J., Deryckere, F., & Gannon, F. (1995). Presence of an extended duplication in the putative low-density-lipoprotein receptor-binding domain of apolipoprotein B. Cloning and characterization of the domain in salmon. *Eur. J. Biochem.*, *230*(1), 45–51.

Babin, P. J., Thisse, C., Durliat, M., Andre, M., Akimenko, M. A., & Thisse, B. (1997). Both apolipoprotein E and A-I genes are present in a nonmammalian vertebrate and are highly expressed during embryonic development. *P. Natl. Acad. Sci. USA*, *94*(16), 8622–8627.

Baker, M. E. (1988). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem. J.*, *255*(3), 1057.

Boren, J., Lee, I., Zhu, W., Arnold, K., Taylor, S., & Innerarity, T. L. (1998). Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J. Clin. Invest.*, *101*(5), 1084–1093. doi:10.1172/JCI1847

Brandt, B. W., Zwaan, B. J., Beekman, M., Westendorp, R. G. J., & Slagboom, P. E. (2005). Shuttling between species for pathways of lifespan regulation: A central role for the vitellogenin gene family? *BioEssays*, *27*(3), 339–346. doi:10.1002/bies.20161

Cardin, A. D., Bowlin, T. L., & Krstenansky, J. L. (1988). Inhibition of lymphocyte proliferation by synthetic peptides homologous to human plasma apolipoproteins B and E. *Biochem. Bioph. Res Co.*, *154*(2), 741–745.

Cardinaux, J. R., Chapel, S., & Wahli, W. (1994). Complex organization of CTF/NF-I, C/EBP, and HNF3 binding sites within the promoter of the liver-specific vitellogenin gene. *J. Biol. Chem.*, *269*(52), 32947–32956.

Clay, M. A., Anantharamaiah, G. M., Mistry, M. J., Balasubramaniam, A., & Harmony, J. A. (1995). Localization of a domain in apolipoprotein E with both cytostatic and cytotoxic activity. *Biochemistry-US*, *34*(35), 11142–11151.

Dashti, N., Gandhi, M., Liu, X., Lin, X., & Segrest, J. P. (2002). The N-Terminal 1000 Residues of Apolipoprotein B Associate with Microsomal Triglyceride Transfer Protein to Create a Lipid Transfer Pocket Required for Lipoprotein Assembly. *Biochemistry-US*, *41*(22), 6978–6987. doi:10.1021/bi011757l

Davis, L. K., Hiramatsu, N., Hiramatsu, K., Reading, B. J., Matsubara, T., Hara, A., et al. (2007). Induction of three vitellogenins by 17beta-estradiol with concurrent inhibition of the growth hormone-insulin-like growth factor 1 axis in a euryhaline teleost, the tilapia (*Oreochromis mossambicus*). *Biol. Reprod.*, *77*(4), 614–625. doi:10.1095/biolreprod.107.060947

Dieckmann, M., Dietrich, M. F., & Herz, J. (2010). Lipoprotein receptors--an evolutionarily ancient multifunctional receptor family. *Biol. Chem.*, *391*(11), 1341–1363. doi:10.1515/BC.2010.129

Durliat, M., Andre, M., & Babin, P. J. (2000). Conserved protein motifs and structural organization of a fish gene homologous to mammalian apolipoprotein E. *E. J. Biochem.*, *267*(2), 549–559.

Dyer, C. A., & Curtiss, L. K. (1991). A synthetic peptide mimic of plasma apolipoprotein E

that binds the LDL receptor. *J. Biol. Chem.*, *266*(34), 22803–22806.

Dyer, C. A., Smith, R. S., & Curtiss, L. K. (1991). Only multimers of a synthetic peptide of human apolipoprotein E are biologically active. *J. Biol. Chem., 266*(23), 15009–15015.

Eldridge, M. B., Whipple, J. A., Eng, D., Bowers, M. J., & Jarvis, B. M. (1981). Effects of Food and Feeding Factors on Laboratory-Reared Striped Bass Larvae. *T. Am. Fish. Soc.*, *110*(1), 111–120. doi:10.1577/1548-8659(1981)110<111:EOFAFF>2.0.CO;2

Finn, R. N. (2007a). The maturational disassembly and differential proteolysis of paralogous vitellogenins in a marine pelagophil teleost: a conserved mechanism of oocyte hydration. *Biol. Reprod.*, *76*(6), 936–948. doi:10.1095/biolreprod.106.055772

Finn, R. N. (2007b). Vertebrate Yolk Complexes and the Functional Implications of Phosvitins and Other Subdomains in Vitellogenins. *Biol. Reprod.*, *76*(6), 926–935.

Finn, R. N. (2007c). Vertebrate yolk complexes and the functional implications of phosvitins and other subdomains in vitellogenins. *Biol. Reprod.*, *76*(6), 926–935. doi:10.1095/biolreprod.106.059766

Finn, R. N., & Kristoffersen, B. A. (2007). Vertebrate Vitellogenin Gene Duplication in Relation to the “3R Hypothesis”: Correlation to the Pelagic Egg and the Oceanic Radiation of Teleosts. *PloS One*, *2*(1), e169. doi:10.1371/journal.pone.0000169

Finn, R. N., Fyhn, H. J., Henderson, R. J., & Evjen, M. S. (1996). The sequence of catabolic substrate oxidation and enthalpy balance of developing embryos and yolksac larvae of turbot (*Scophthalmus maximus* L.). *Comp. Biochem. Physiol. A*, *115*(2), 133–151. doi:10.1016/0300-9629(96)00026-6

Finn, R. N., Kolarevic, J., Kongshaug, H., & Nilsen, F. (2009). Evolution and differential expression of a vertebrate vitellogenin gene cluster. *BMC Evol. Biol.*, *9*, 2. doi:10.1186/1471-2148-9-2

Ghosh, P., & Thomas, P. (1995). Binding of metals to red drum vitellogenin and incorporation into oocytes. *Mar. Environ. Res.*, *39*(1–4), 165–168. doi: 10.1016/0141-1136(94)00035-N

Bujo, H., Hermann, H., Kaderli, M, O., Jacobsen, L., Sugawara, S., Nimpf, J.,Yamamoto, T., Schneider, W, J. (1994). Chicken oocyte growth is mediated by an eight ligand binding repeat member of the LDL receptor family. *EMBO J.*, *13*(21), 5165.

Hayashi, K., Ando, S., Stifani, S., & Schneider, W. J. (1989). A novel sterol-regulated surface protein on chicken fibroblasts. *J. Lipid Res.*, *30*(9), 1421–1428.

Hayward, A., Takahashi, T., Bendena, W. G., Tobe, S. S., & Hui, J. H. L. (2010). Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans. *FEBS Lett.*, *584*(6), 1273–1278. doi:10.1016/j.febslet.2010.02.056

Hiramatsu, N., Chapman, R. W., Lindzey, J. K., Haynes, M. R., & Sullivan, C. V. (2004a). Molecular Characterization and Expression of Vitellogenin Receptor from White Perch (*Morone americana*). *Biol. Reprod.*, *70*(6), 1720–1730.

Hiramatsu, N., Chapman, R. W., Lindzey, J. K., Haynes, M. R., & Sullivan, C. V. (2004b). Molecular characterization and expression of vitellogenin receptor from white perch (*Morone americana*). *Biol. Reprod.*, *70*(6), 1720–1730. doi:10.1095/biolreprod.103.023655

Hiramatsu, N., Hara, A., Hiramatsu, K., Fukada, H., Weber, G. M., Denslow, N. D., & Sullivan, C. V. (2002a). Vitellogenin-Derived Yolk Proteins of White Perch, *Morone americana*: Purification, Characterization, and Vitellogenin-Receptor Binding1. *Biol. Reprod.*, *67*(2), 655–667.

Hiramatsu, N., Matsubara, T., Hara, A., Donato, D. M., Hiramatsu, K., Denslow, N. D., & Sullivan, C. V. (2002b). Identification, purification and classification of multiple forms of vitellogenin from white perch (*Morone americana*). *Fish Physiol. Biochem.*, *26*(4), 355–370. doi:10.1023/B:FISH.0000009266.58556.9a

Innerarity, T. L., Friedlander, E. J., Rall, S. C., Weisgraber, K. H., & Mahley, R. W. (1983). The receptor-binding domain of human apolipoprotein E. Binding of apolipoprotein E fragments. *J. Biol. Chem.*, *258*(20), 12341–12347.

Jingami, H., & Yamamoto, T. (1995). The VLDL receptor: wayward brother of the LDL

receptor. *Curr. Opin. Lipidol.*, *6*(2), 104–108.

Kanungo, J., Petrino, T. R., & Wallace, R. A. (1990). Oogenesis in Fundulus heteroclitus. VI. Establishment and verification of conditions for vitellogenin incorporation by oocytes in vitro. *J. Exp. Zool.*, *254*(3), 313–321. doi:10.1002/jez.1402540310

Lalazar, A., Weisgraber, K. H., Rall, S. C., Giladi, H., Innerarity, T. L., Levanon, A. Z., et al. (1988). Site-specific mutagenesis of human apolipoprotein E. Receptor binding activity of variants with single amino acid substitutions. *J. Biol. Chem.*, *263*(8), 3542–3545.

Li, A., Sadasivam, M., & Ding, J. L. (2003). Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. *J. Biol. Chem.*, *278*(5), 2799–2806. doi:10.1074/jbc.M205067200

Mahley, R. W., & Rall, S. C. (2000). Apolipoprotein E: far more than a lipid transport protein. *Annu. Rev. Genom. Hum. G.*, *1*, 507–537. doi:10.1146/annurev.genom.1.1.507

Mahley, R. W., Innerarity, T. L., Rall, S. C., & Weisgraber, K. H. (1984). Plasma lipoproteins: apolipoprotein structure and function. *J. Lipid Res.*, *25*(12), 1277–1294.

Mansueti, R. J. (1964). Eggs, Larvae, and Young of the White Perch, *Roccus americanus*, with Comments on Its Ecology in the Estuary. *Chesapeake Sci.*, *5*(1/2), 3.

doi:10.2307/1350789

McKeone, B. J., Patsch, J. R., & Pownall, H. J. (1993). Plasma triglycerides determine low density lipoprotein composition, physical properties, and cell-specific binding in cultured cells. *J. Clin. Invest.*, *91*(5), 1926–1933. doi:10.1172/JCI116411

Montorzi, M., Falchuk, K. H., & Vallee, B. L. (1995). Vitellogenin and Lipovitellin: Zinc Proteins of *Xenopus laevis* Oocytes. *Biochemistry-US*, *34*(34), 10851–10858. doi:10.1021/bi00034a018

Ohkubo, N., & Matsubara, T. (2002). Sequential utilization of free amino acids, yolk proteins and lipids in developing eggs and yolk-sac larvae of barfin flounder *Verasper moseri*. *Mar. Biol.*, *140*(1), 187–196. doi:10.1007/s002270100647

Nigon, F., Lesnik, P., Rouis, M., & Chapman, M. J. (1991). Discrete subspecies of human low density lipoproteins are heterogeneous in their interaction with the cellular LDL receptor. *J. Lipid Res.*, *32*(11), 1741–1753.

Nikoulin, I. R., & Curtiss, L. K. (1998). An apolipoprotein E synthetic peptide targets to lipoproteins in plasma and mediates both cellular lipoprotein interactions in vitro and acute clearance of cholesterol-rich lipoproteins in vivo. *J. Clin. Invest.*, *101*(1), 223–234. doi:10.1172/JCI1099

Norberg, B., & Haux, C. (1985). Induction, isolation and a characterization of the lipid

content of plasma vitellogenin from two *Salmo* species: rainbow trout (*Salmo gairdneri*) and sea trout (*Salmo trutta*). *Comp. Biochem. Physiol. B*, *81*(4), 869–876.

North, E. W., & Houde, E. D. (2003). Linking ETM physics zooplankton prey and fish early life histories to striped bass *Morone saxatilis* and white perch *M. americana* recruitment. *Mar. Ecol.-Prog. Ser.*, *260*, 219-236.

Ohkubo, N., Sawaguchi, S., Hamatsu, T., & Matsubara, T. (2006). Utilization of free amino acids, yolk proteins and lipids in developing eggs and yolk-sac larvae of walleye pollock *Theragra chalcogramma*. *Fisheries Sci.*, *72*(3), 620–630. doi:10.1111/j.1444-2906.2006.01192.x

Olofsson, S.-O., & Boren, J. (2005). Apolipoprotein B: a clinically important apolipoprotein which assembles atherogenic lipoproteins and promotes the development of atherosclerosis. *J. Intern. Med.*, *258*(5), 395–410. doi:10.1111/j.1365-2796.2005.01556.x

Plack, P. A., & Pritchard, D. J. (1968). Effect of oestradiol 3-benzoate on the concentrations of retinal and lipids in cod plasma. *Biochem. J.*, *106*(1), 257–262.

Prat, F., Coward, K., Sumpter, J. P., & Tyler, C. R. (1998). Molecular characterization and expression of two ovarian lipoprotein receptors in the rainbow trout, *Oncorhynchus mykiss*. *Biol. Reprod.*, *58*(5), 1146–1153.

Raag, R., Appelt, K., Xuong, N. H., & Banaszak, L. (1988). Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 A resolution. *J. M. Biol.*, *200*(3), 553–569.

Raffaï, R., Weisgraber, K. H., MacKenzie, R., Rupp, B., Rassart, E., Hirama, T., et al. (2000). Binding of an antibody mimetic of the human low density lipoprotein receptor to apolipoprotein E is governed through electrostatic forces. Studies using site-directed mutagenesis and molecular modeling. *J. Biol. Chem.*, *275*(10), 7109–7116.

Rall, S. C., Weisgraber, K. H., Innerarity, T. L., & Mahley, R. W. (1982). Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects. *P. Natl. Acad. Sci. USA*, *79*(15), 4696–4700.

Reading, B. J., Hiramatsu, N., & Sullivan, C. V. (2011). Disparate binding of three types of vitellogenin to multiple forms of vitellogenin receptor in white perch. *Biol. Reprod.*, *84*(2), 392–399. doi:10.1095/biolreprod.110.087981

Reading, B. J., Hiramatsu, N., Schilling, J., Molloy, K. T., Glassbrook, N., Mizuta, H., et al. (2014). Lrp13 is a novel vertebrate lipoprotein receptor that binds vitellogenins in teleost fishes. *J. Lipid Res.*, *55*(11), 2287–2295. doi:10.1194/jlr.M050286

Ren, G., Rudenko, G., Ludtke, S. J., Deisenhofer, J., Chiu, W., & Pownall, H. J. (2010). Model of human low-density lipoprotein and bound receptor based on CryoEM. *P. Natl. Acad. Sci. USA*, *107*(3), 1059–1064. doi:10.1073/pnas.0908004107

Richards, M. P. (1997). Trace mineral metabolism in the avian embryo. *Poultry Sci.*, *76*(1), 152–164.

Richardson, P. E., Manchekar, M., Dashti, N., Jones, M. K., Beigneux, A., Young, S. G., et al. (2005). Assembly of lipoprotein particles containing apolipoprotein-B: structural model for the nascent lipoprotein particle. *Biophys. J.*, *88*(4), 2789–2800. doi:10.1529/biophysj.104.046235

Rowe, R. I., & Eckhert, C. D. (1999). Boron is required for zebrafish embryogenesis. *J. Exp. Biol.*, *202*(12), 1649–1654.

Russell, D. W., Brown, M. S., & Goldstein, J. L. (1989). Different combinations of cysteine-rich repeats mediate binding of low density lipoprotein receptor to two different proteins. *J. Biol. Chem.*, *264*(36), 21682–21688.

Sawaguchi, S., Ohkubo, N., Amano, H., Hiramatsu, N., Hara, A., Sullivan, C. V., & Matsubara, T. (2008). Controlled accumulation of multiple vitellogenins into oocytes during vitellogenesis in the barfin flounder, *Verasper moseri*. *Cybium Int. J. Ichthyol.*, *32*(suppl 2), 262.

Sawaguchi, S., Ohkubo, N., Koya, Y., & Matsubara, T. (2005). Incorporation and utilization of multiple forms of vitellogenin and their derivative yolk proteins during vitellogenesis and embryonic development in the mosquitofish, *Gambusia affinis*. *Zool. Sci.*, *22*(6), 701–710.

Schilling, J., Nepomuceno, A., Schaff, J. E., Muddiman, D. C., Daniels, H. V., & Reading, B. J. (2014). Compartment Proteomics Analysis of White Perch (*Morone americana*) Ovary Using Support Vector Machines. *J. Proteome Res.*, *13*, 1515-1526. doi:10.1021/pr401067g

Schneider, W. J. (2009). Receptor-mediated mechanisms in ovarian follicle and oocyte development. *Gen. Comp. Endocr.*, *163*(1–2), 18–23. doi: 10.1016/j.ygcen.2008.11.032

Segrest, J. P., Jones, M. K., Mishra, V. K., Pierotti, V., Young, S. H., Boren, J., et al. (1998). Apolipoprotein B-100: conservation of lipid-associating amphipathic secondary structural motifs in nine species of vertebrates. *J. Lipid Res.*, *39*(1), 85–102.

Silversand, C., & Haux, C. (1995). Fatty acid composition of vitellogenin from four teleost species. *J. Comp. Physiol. B*, *164*(8), 593–599. doi:10.1007/BF00389799

Smolenaars, M. M. W., Madsen, O., Rodenburg, K. W., & Van der Horst, D. J. (2006). Molecular diversity and evolution of the large lipid transfer protein superfamily. *J. Lipid Res.*, *48*(3), 489–502. doi:10.1194/jlr.R600028-JLR200

Steyrer, E., Barber, D. L., & Schneider, W. J. (1990). Evolution of lipoprotein receptors. The chicken oocyte receptor for very low density lipoprotein and vitellogenin binds the mammalian ligand apolipoprotein E. *J. Biol. Chem.*, *265*(32), 19575–19581.

Stifani, S., Menn, F. L., Rodriguez, J. N., & Schneider, W. J. (1990). Regulation of

oogenesis: The piscine receptor for vitellogenin. *BBA - Lipid Lipid Met.*, *1045*(3), 271–279. doi:10.1016/0005-2760(90)90130-P

Takahashi, S., Sakai, J., Fujino, T., Hattori, H., Zenimaru, Y., Suzuki, J., et al. (2004). The very low-density lipoprotein (VLDL) receptor: characterization and functions as a peripheral lipoprotein receptor. *J. Atheroscler. Thromb.*, *11*(4), 200–208.

Takahashi, Y., Itoh, F., Oohashi, T., & Miyamoto, T. (2003). Distribution of apolipoprotein E among lipoprotein fractions in the lactating cow. *Comp. Biochem. Phys. B*, *136*(4), 905–912. doi:10.1016/j.cbpc.2003.09.004

Tao, Y., Berlinsky, D. L., & Sullivan, C. V. (1996). Characterization of a vitellogenin receptor in white perch (*Morone americana*). *Biol. Reprod.*, *55*(3), 646–656.

Taskinen, M.-R. (2005). Type 2 diabetes as a lipid disorder. *Curr. Mol. Med.*, *5*(3), 297–308.

Van der Horst, D. J., Roosendaal, S. D., & Rodenburg, K. W. (2009). Circulatory lipid transport: lipoprotein assembly and function from an evolutionary perspective. *Mol. Cell. Biochem.*, *326*(1-2), 105–119. doi:10.1007/s11010-008-0011-3

Wahli, W. (1988). Evolution and expression of vitellogenin genes. *Trends Genet.*, *4*(8), 227–232.

Weisgraber, K. H., & Mahley, R. W. (1996). Human apolipoprotein E: the Alzheimer's disease connection. *FASEB J.*, *10*(13), 1485–1494.

Weisgraber, K. H., Innerarity, T. L., & Mahley, R. W. (1978). Role of lysine residues of plasma lipoproteins in high affinity binding to cell surface receptors on human fibroblasts. *J. Biol. Chem.*, *253*(24), 9053–9062.

Williams, V. N., Reading, B. J., Amano, H., Hiramatsu, N., Schilling, J., Salger, S. A., et al. (2014). Proportional accumulation of yolk proteins derived from multiple vitellogenins is precisely regulated during vitellogenesis in striped bass (*Morone saxatilis*). *J. Exp. Zool.*, *321*(6), 301–315. doi:10.1002/jez.1859

Wilson, C., Wardell, M. R., Weisgraber, K. H., Mahley, R. W., & Agard, D. A. (1991). Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. *Science*, *252*(5014), 1817–1822.

Zaiou, M., Arnold, K. S., Newhouse, Y. M., Innerarity, T. L., Weisgraber, K. H., Segall, M. L., et al. (2000). Apolipoprotein E;-low density lipoprotein receptor interaction. Influences of basic residue and amphipathic alpha-helix organization in the ligand. *J. Lipid Res.*, *41*(7), 1087–1095.

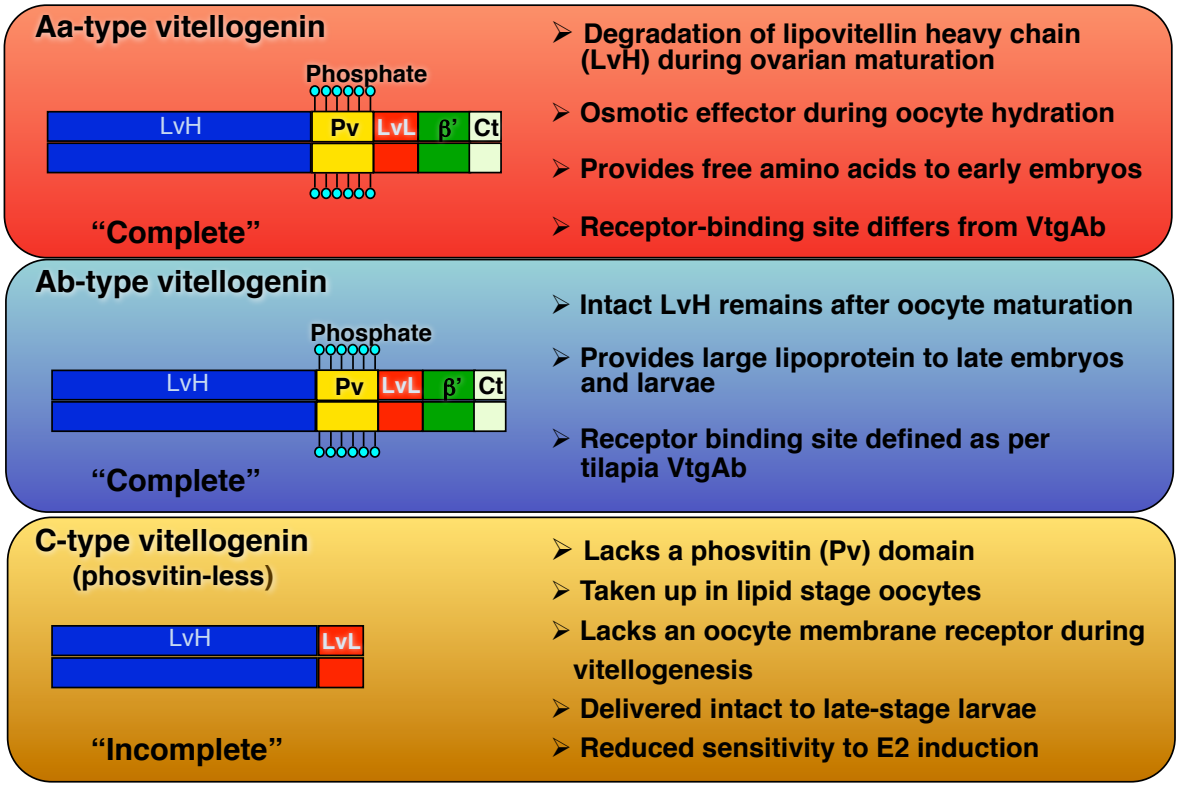


Figure 1. The primary domains of the three forms of vitellogenins are depicted. Yolk proteins derived from Aa-type vitellogenin are cleaved into free amino acids during final oocyte maturation. Ab-type vitellogenin receives partial or no proteolysis on lipovitellin heavy chain and are utilized during late embryonic growth. Adapted from Hiramatsu et al., 2002 (Biol. Reproduction, Fish Physiol Biochem); Reading et al., 2009 (Mar Biotech).

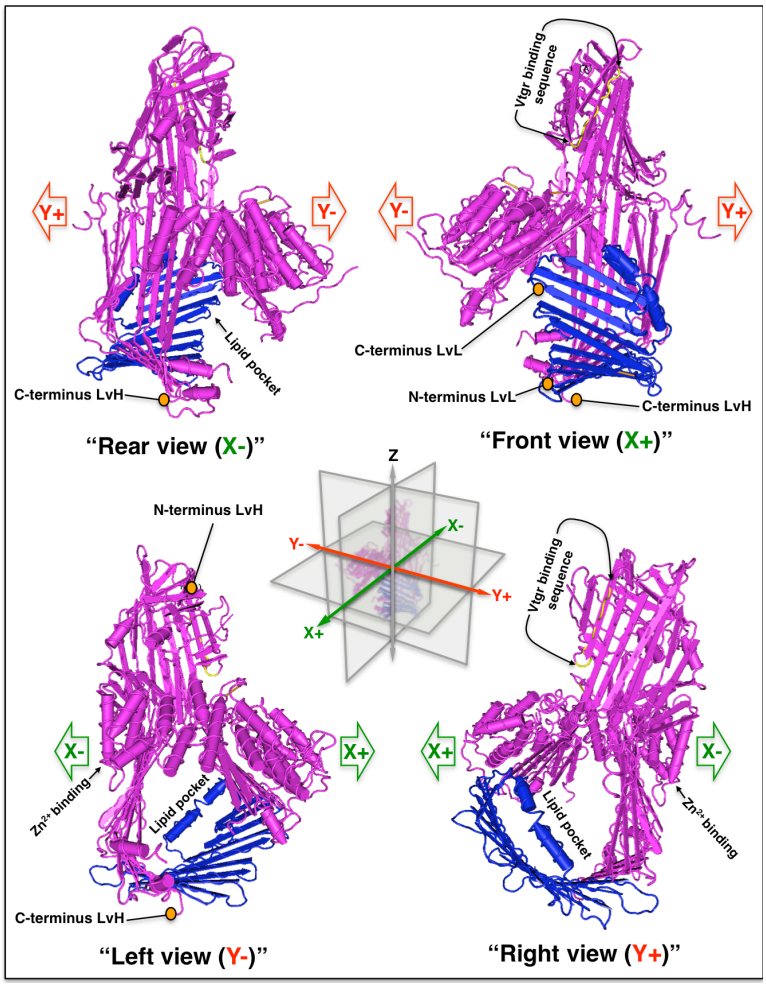


Figure 2. 3D representation of a partial lamprey (*Ichthyomyzon unicuspis*) complete type vitellogenin. The lipovitellin heavy domain (LvH) is colored purple while the lipovitellin light domain (LvL) is blue. The LR8 binding sequence is in yellow (Li et al., 2003). The large lipid binding pocket and putative Zn^{2+} binding site are also indicated (pdb: 1LSH). From Anderson, Levitt, & Banaszak, 1998.

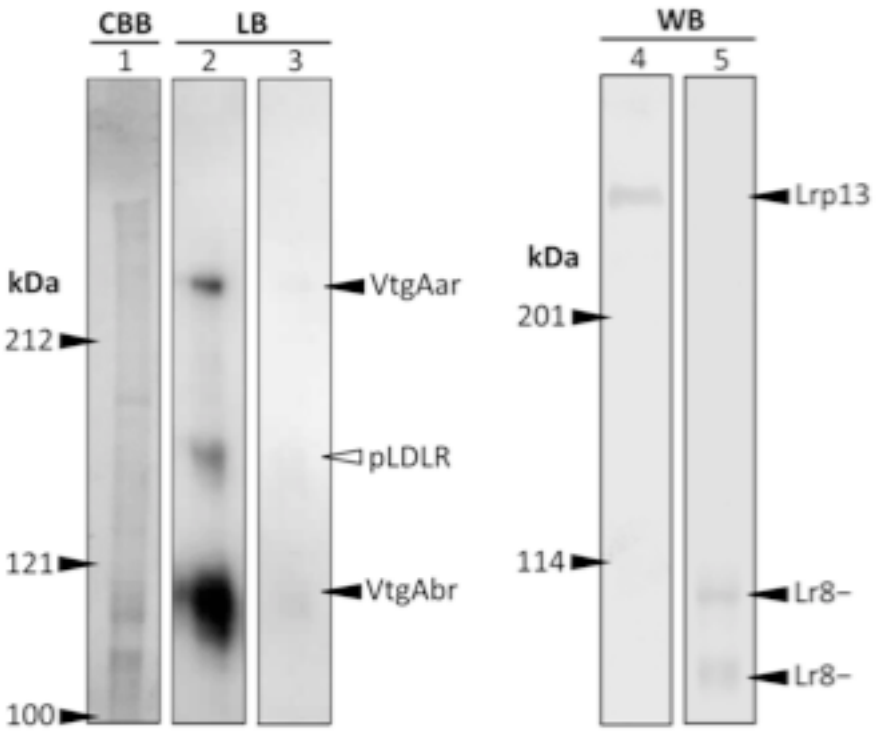


Figure 3. (Left) Nonreducing 7.5% acrylamide gel stained with Coomassie brilliant blue (CBB) and ligand blot (LB) of white perch ovary membrane proteins. The LB in lane 2 was prepared using 0.25 µg/ml of DIG-labeled VtgAa/b and the LB in lane 3 was performed in the presence of a 200-fold excess molar ratio of unlabeled VtgAa/b. (Right) Nonreducing 5% acrylamide gel Western blot (WB) of white perch ovary membrane proteins. The WB in lane 4 was prepared with α-WpLrp13 and the WB in lane 5 was performed with α-WpLr8−. Numbers to the left of gels or blots indicate the sizes of molecular weight markers (kDa). Positions of VtgAar, VtgAbr, pLDLR, Lr8−, and Lrp13 are indicated. From Reading et al., 2014 (J Lipid Res).

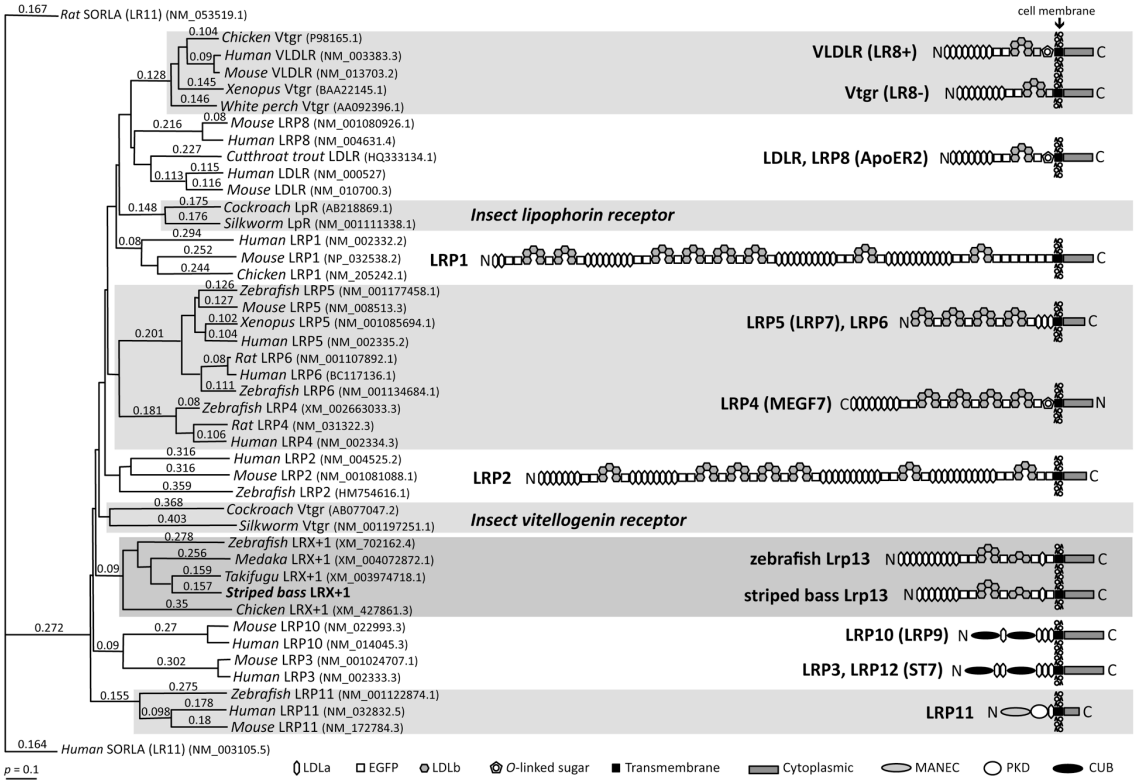


Figure 4. (Left) ClustalW dendrogram showing relationships between low-density lipoprotein receptor family polypeptide sequences. GenBank accession numbers are provided. Numbers above each branch are p-distances. (Right) Models representing linear domain structures of low-density lipoprotein receptor family members as defined at the bottom. From Reading et al., 2014 (J Lipid Res).

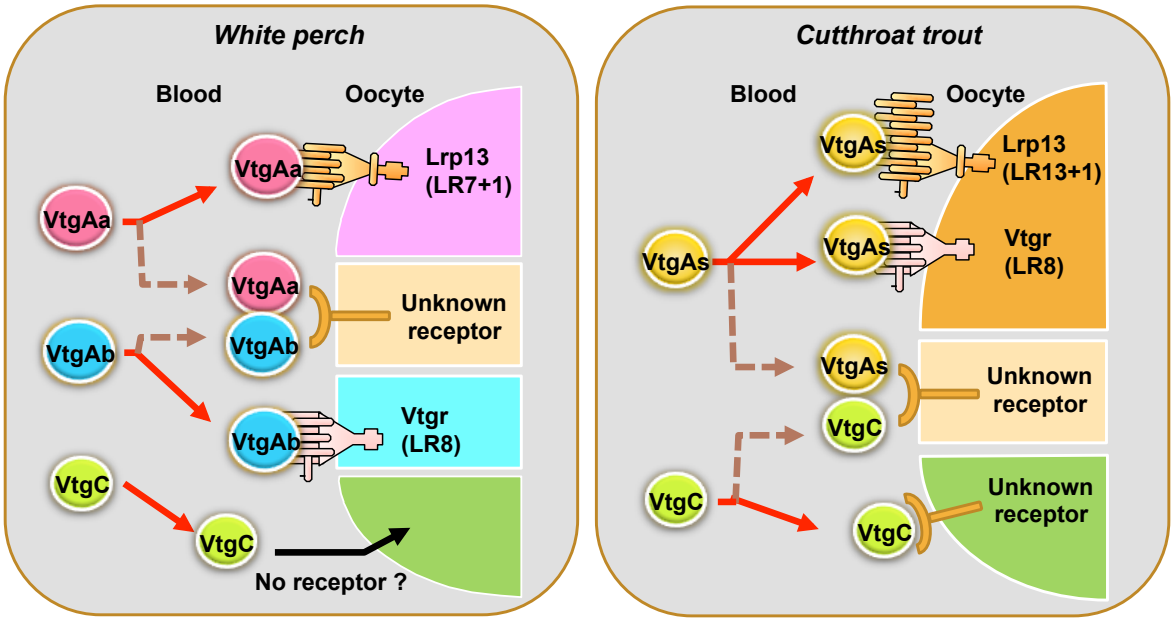


Figure 5. Multiple vitellogenins (Vtg) and their receptors involved in Vtg-derived yolk formation of white perch and cutthroat trout. The A-type Vtgs (VtgAa, VtgAb and VtgAs) bind the 'classical' LR8-type Vtg receptor (Vtgr) and/or low-density lipoprotein receptor related protein 13 (Lrp13). Receptor proteins for C-type Vtg are detected in ligand blots of trout ovarian membrane, but not in white perch. Receptor proteins that universally bind multiple Vtg subtypes are also detected in the ovarian membrane preparations of both species. LR8: lipoprotein receptor (LR) with 8 ligand binding (LB) repeats; LR7+1: LR with 7+1 LB repeats; LR13+1: LR with 13+1 LB repeats. From Hiramatsu et al., 2015 (Gen Comp Endocr).

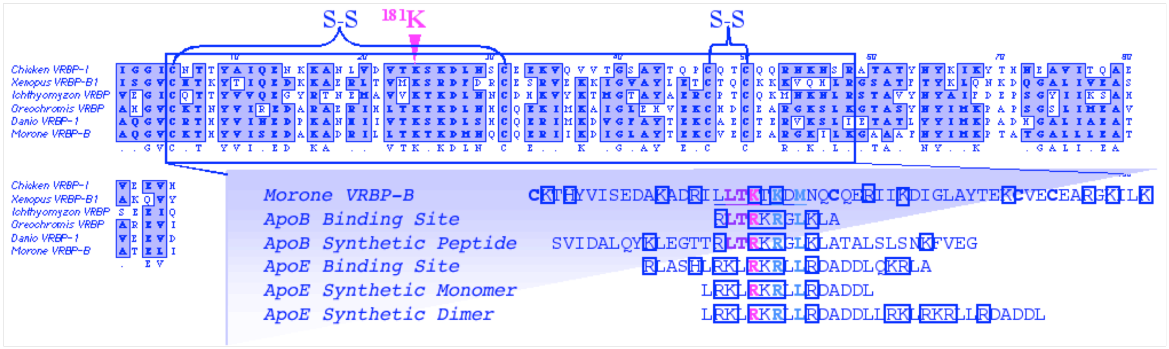


Figure 6. The vitellogenin receptor binding domain of complete type *Morone* VtgAb features that are similarly conserved in *Gallus*, *Xenopus*, *Oreochromis*, *Danio*, and human apoE and apoB synthetic peptides (Dyer et al., 1995). Beyond overall sequence similarity, the sequences above contain highly conserved cysteines that form disulfide bonds crucial for presentation of the positive residues (lysine [K] and arginine [R]) at positions 181 and 183. Identified as essential for receptor binding, ^{181}K is noted in pink (Li et al., 2003).

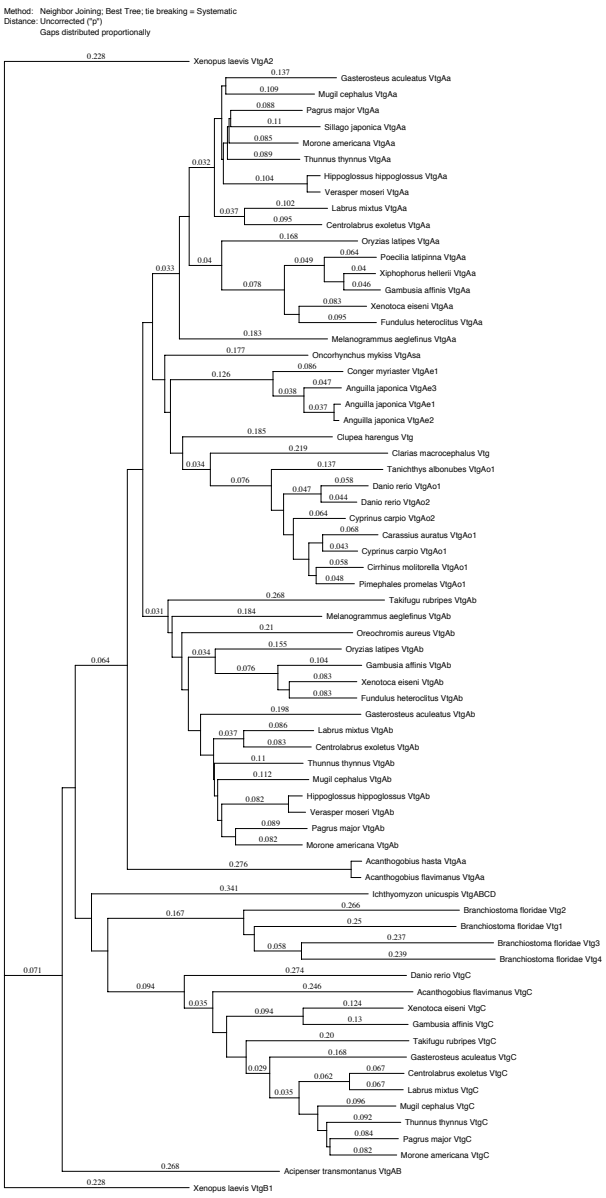


Figure 7. Vitellogenin polypeptide sequences (N=70) from an array of fishes were aligned by ClustalW to generate a dendrogram. Vitellogenins group by type and the Aa and Ab types are present in Acanthomorph fishes which represent more derived teleosts.

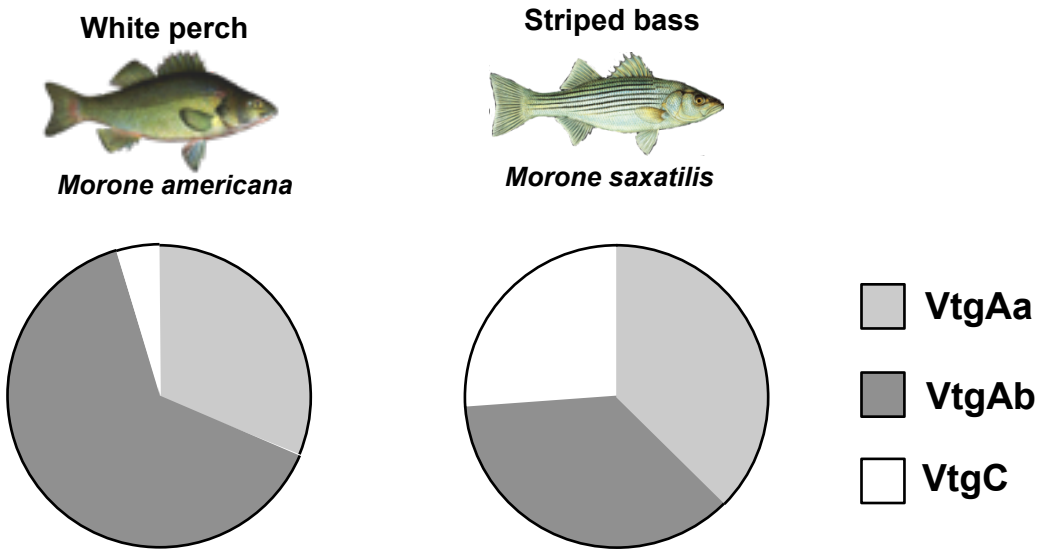


Figure 8. In white perch, yolk proteins derived from VtgC are minor components of the total egg yolk (< 5%), whereas in striped bass they are major components of the egg yolk (~ 25%). [Williams et al., 2014 (J Exp Zool A); Schilling et al., 2014 (J Proteome Res)].

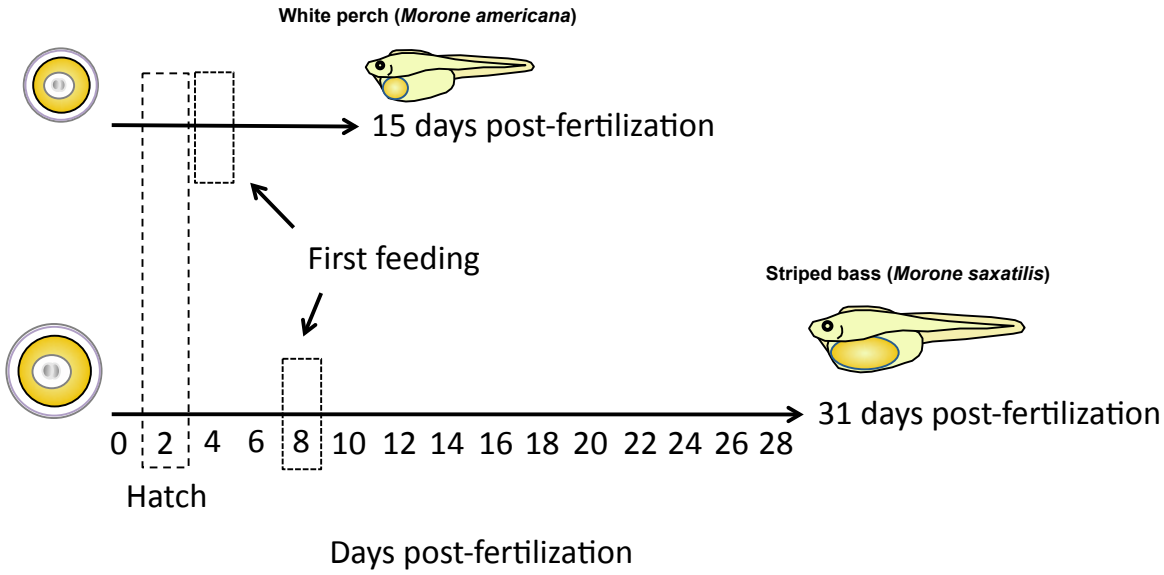


Figure 9. Average survival duration of food-restricted white perch and striped bass larvae. Dashed boxes indicate approximate time of hatching (~2 days) and onset of first feeding (~4 days in white perch, ~8 days in striped bass). [Mansuetti, 1964; Eldridge, et al., 1981; North & Houde, 2003].

CHAPTER 2

Compartment Proteomics Analysis of White Perch (*Morone americana*) Ovary Using Support Vector Machines

Justin Schilling [1], Angelito Nepomuceno [2], Jennifer E. Schaff [3], David C. Muddiman [2], Harry V. Daniels [1] and Benjamin J. Reading [1,]*

[1] Department of Applied Ecology, College of Agriculture and Life Sciences, North Carolina State University, 127 David Clark Labs, Raleigh, North Carolina, United States

[2] W. M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina, United States

[3] Genomic Sciences Laboratory, North Carolina State University, Raleigh, North Carolina, United States

Reprinted with permission from Schilling, J., Nepomuceno, A., Schaff, J. E., Muddiman, D. C., Daniels, H. V., & Reading, B. J. (2014). Compartment proteomics analysis of white perch (*Morone americana*) ovary using support vector machines. Journal of Proteome Research, 13, 1515−1526. doi:10.1021/pr401067g.

KEYWORDS: compartment proteomics; support vector machines; modulated modularity clustering; ovary; oocytes; vitellogenin; transcriptome; mitochondria; alternatively spliced variants

ABSTRACT: Compartment proteomics enable broad characterization of target tissues. We employed a simple fractionation method and filter-aided sample preparation (FASP) to

characterize the cytosolic and membrane fractions of white perch ovary tissues by semiquantitative tandem mass spectrometry using label-free quantitation based on normalized spectral counts. FASP depletes both low-molecular-weight and high-molecular-weight substances that could interfere with protein digestion and subsequent peptide separation and detection. Membrane proteins are notoriously difficult to characterize due to their amphipathic nature and association with lipids. The simple fractionation we employed effectively revealed an abundance of proteins from mitochondria and other membrane-bounded organelles. We further demonstrate that support vector machines (SVMs) offer categorical classification of proteomics data superior to that of parametric statistical methods such as analysis of variance (ANOVA). Specifically, SVMs were able to perfectly (100% correct) classify samples as either membrane or cytosolic fraction during cross-validation based on the expression of 242 proteins with the highest ANOVA *p*-values (i.e. those that were not significant for enrichment in either fraction). The white perch ovary cytosolic and membrane proteomes and transcriptome presented in this study can support future investigations into oogenesis and early embryogenesis of white perch and other members of the genus *Morone*.

INTRODUCTION

White perch (*Morone americana*) are native to lakes, rivers, and estuaries along the Mid-Atlantic and Northeastern seaboards of North America. Given their relatively broad niche distribution, white perch are a key indicator species for ecological monitoring.[1-3] White perch are temperate basses (genus *Morone*), a group of fishes that includes species of economic importance to commercial and recreational fisheries (striped bass, *M. saxatilis*) and aquaculture (hybrid striped bass, *M. saxatilis* X *M. chrysops*). Female reproductive

dysfunction (i.e. production of inviable eggs, embryos, and larvae) remains a major obstacle to the expansion of finfish aquaculture and influences recruitment in fisheries and subsequent management practices.

While transcriptomics and proteomics offer promise for greatly advancing the understanding of reproductive molecular biology, these methods rely on extant homologous databases, which remain incomplete for many non-model species including white perch and the other members of the genus *Morone*. Recent studies have provided the first ovary transcriptome sequences and proteomic analyses available for the striped bass.[4,5] The white perch is an important research model for oogenesis of fishes of genus *Morone*[6-12] yet its ovary transcriptome and proteome are not described. Additionally, Acanthomorph teleosts, such as the temperate basses, express three distinct forms of vitellogenin (VtgAa, VtgAb, and VtgC). These egg yolk precursors are produced by the liver and released into the circulatory system where they are taken up specifically by growing oocytes. Understanding the disparate accumulation and processing of vitellogenins by the oocytes is relevant to egg quality, since these egg yolk components not only provide nutrition to embryos and larvae at specific developmental stages, but contribute to oocyte hydration and egg buoyancy as well. Therefore, the objectives of the present study were to generate an ovary transcriptome and cytosolic and membrane ovary proteomes for white perch to serve as resources for future investigations of reproduction and egg quality in temperate basses. We also measured the proportional abundance of the three vitellogenins in the white perch ovary at the late vitellogenic stage, just prior to ovulation.

Cellular compartment fractionation can improve coverage in proteomics studies, enabling detection of moderate and low abundance proteins.[13,14] We employed a simple and cost effective cellular compartment fractionation in order to broadly characterize the white perch

ovary cytosolic and membrane proteomes representative of ovarian follicles on the cusp of ovulation. Furthermore, we employed support vector machines (SVMs) to analyze these data. The SVMs are supervised machine learning non-probabilistic binary linear classifiers that can efficiently perform non-linear classification as well using the kernel function[15] and are superior to conventional parametric analytical methods for categorical assignment of high-dimensionality data frequently encountered in 'omics' studies.[16,17] We used SVMs to classify samples as cytosolic or membrane fractions based on protein expression data. These classifiers performed perfectly even when data of those proteins whose expression was not significantly enriched in either fraction by one-way analysis of variance (ANOVA) was used as input.

EXPERIMENTAL

Sample Collection and Preparation

Adult female white perch were reared under controlled photo-thermal conditions at the North Carolina State University Aquaculture Facility (Raleigh, NC).[18,19] Females (N = 3) were anesthetized with Finquel MS-222 (Argent Chemical Laboratories, Redmond, WA), and whole ovary tissues were collected by dissection at a single time point in March 2013 (body weight 114.95 $\pm$ 23.72 g; total length 195.67 $\pm$ 12.10 mm; ovary weight 10.02 $\pm$ 1.91 g). The white perch is a multiple-clutch, group-synchronous spawner in which ovarian follicles at several stages of development are present in the adult ovary.[11,19] At this time in the annual reproductive cycle, the dominant clutch of follicles in the white perch ovary is late vitellogenic. A sample of each ovary was fixed in Bouin's solution for histology at the North Carolina State University College of Veterinary Medicine Histology Laboratory (Raleigh, NC). Fractionation of the ovarian follicles was carried out using methods adapted from

previous studies.[9,10] Approximately 2 g of each ovary were pooled and homogenized by approximately ten strokes of Dounce homogenization in 24 mL (1 g/4 mL) of Tris-buffered saline (20 mM Tris-HCL pH 8.0, 150 mM NaCl, and 2 mM $CaCl_2$) (all chemicals were purchased from Sigma Aldrich, St. Louis, MO unless otherwise stated) and centrifuged at 13,000 × *g* for 15 minutes and 4°C. The resulting supernatant was retained as the ovary cytosolic fraction. The pellet comprising the ovary membrane fraction was resuspended in Tris-buffered saline and centrifuged at 13,000 × *g* for 15 minutes and 4°C and the supernatant was discarded. This procedure was repeated five additional times. The membrane pellet was then resuspended in ice-cold Tris-buffered saline and homogenized using a Brinkmann™Polytron™ PT10/35 tissue homogenizer on setting 5 for two consecutive 30 second bursts on ice. The ovary membrane fraction was then centrifuged a final time at 500 × *g* for 5 minutes at 4°C, after which the pellet was sequentially aspirated through 18, 22 and finally 25 gauge needles until completely resuspended in ice-cold Tris-buffered saline. Ovary cytosolic and membrane fractions were diluted with ice-cold Tris-buffered saline to final protein concentrations of 2 mg/mL determined by Bradford Assay and frozen at -80°C until use.

White Perch Ovary Transcriptome

Insert sequences from a UniZap II Lambda cDNA library of vitellogenic white perch ovary[11] were selectively amplified from the multiple cloning site by PCR using High Fidelity Platinum PCR Supermix (Life Technologies, Grand Island, NY) and an SK forward primer (5'-CGC TCT AGA ACT AGT GGA TC-3') and an oligo $(dT)_{12-18}$ reverse primer. The PCRs were performed according to Reading *et al*.[8] with the exception that the following thermal parameters were used: 94°C 2 minutes and then 35 cycles of 94°C for 30 seconds, 56°C for

30 seconds, and 68°C for 5 minutes. Five μg of PCR products was submitted to the NCSU Genomic Sciences Laboratory (Raleigh, NC) for standard 454 pyrosequencing library preparation. The sample was sequenced on one half-plate in the GS FLX 454 (Roche, Branford, CT). Vector sequence beginning with the SK primer and including the *Xho* cloning site was removed from the 5' end of all short read sequences, which were then assembled using a version of CAP3.[20] The parameters of the CAP3 assembly were: base quality cutoff for clipping = 12, overlap length cutoff = 30, overlap percent identity cutoff = 75, overlap similarity score cutoff = 500, clipping range = 250, and minimum number of good reads at clipping position = 2. These sequences were subjected to BLAST (blastx)[21] of the NCBI database and annotated according to the Gene Ontology Consortium[22] using Blast2GO 2048 M version 12.2.0.[23-25] Parameters for blastx were: Expect value 1×10^{-3} and HSP Length Cutoff 33. Parameters for the gene ontology (GO) annotations were: E-value-hit-filter 1×10^{-6}, Annotation Cutoff 55, GO Weight 5, and HSP-Hit Coverage Cutoff 0. Combined GO graphs for the annotated sequences (1,261 total) were created using percentages of 2nd level GO terms for Biological Process, 3rd level GO terms for Molecular Function, and 4th level GO terms for Cellular Component. Represented GO classes were restricted to those with 50 or more entities (sequence cutoff = 50.0); Sequence Filter = 50, Score alpha = 0.6, Node Score Filter = 10.

Filter-Aided Sample Preparation and Digestion

FASP was utilized with minor changes.[26] The starting amounts for the cytosolic and membrane fractions were 100 mg of total protein. Disulfide bonds were reduced by adding 3 mL of 50 mM dithiothreitol (DTT) to 30 mL of sample followed by incubation at 56°C for 30 minutes. Samples were then mixed with 200 mL of 8 M urea in 0.1 M Tris-HCl pH 8.5. The

solution was then transferred onto a Vivacon 500 30 kDa MW cutoff filter (Sartorius Stedim Biotech, Goettingen, Germany) and centrifuged at 14,000 × *g* for 15 minutes at 21°C. The urea wash step was repeated once more and the flow through was discarded. Iodoacetamide was prepared at a concentration of 50 mM in 8 M urea, and 100 mL was added onto the filter. The samples were then incubated for 20 minutes in the dark at room temperature. Samples were then centrifuged at 14,000 × *g* for 10 minutes at 21°C. Wash steps then proceeded with 100 mL of 8 M urea three times and 100 mL of 0.05 ammonium bicarbonate in H_2O three times. After each series of urea and ammonium bicarbonate washes, the flow though was discarded. Each wash was completed through centrifugation for 10 minutes at 14,000 × *g*. Trypsin was prepared with 50 mM ammonium bicarbonate and added to samples at an enzyme to protein ratio of 1:100. The filter was then placed onto a new centrifuge tube. Solutions were then incubated at 37°C overnight. Tryptic peptides were filtered through by centrifugation at 14,000 × *g* for 10 minutes at 21°C.

nanoReversed Phase Chromatography and Tandem Mass Spectrometry

Protein concentrations of the digests were obtained using a Nanodrop at A_{280} (Thermo Scientific, Wilmington, DE). Samples were reconstituted to a protein concentration of 0.2 mg/mL using mobile phase A (98/2/0.2% water/acetonitrile/formic acid) and a total of 5 mL was injected onto the trap for desalting. Separation of peptides was performed using a Thermo Scientific EASY nLC II (Thermo Scientific, San Jose, CA) in line with a cHiPLC nanoflex system (AB Sciex, Framingham, MA). A vented column configuration,[27] a ChromXP C18-CL 3 µm trap column, and a ChromXP C18-CL 75 mm × 15 cm analytical column were used for these experiments. The initial condition of 2% mobile phase B (2/98/0.2% water/acetonitrile/formic acid) was increased to 35% over 201 minutes, then steeply ramped

to 95% mobile phase B over 10 minutes and maintained at 95% mobile phase B for 10 minutes to wash the column. The column was then equilibrated at 2% mobile phase B for 12 minutes. Each sample was analyzed in triplicate.

The nanoLC configuration above was coupled to a quadrupole orbitrap mass spectrometer[28] (Q-Exactive, Thermo Scientific, San Jose, CA). The Q-Exactive was operated using optimized global proteomics parameters described by Randall *et. al*.[29] Eluting peptides were ionized by applying 2.0 kV to a union between the outlet of the LC system and the emitter tip. The MS inlet capillary was set to 275°C. MS transients were acquired with 70 k_{FWHM} resolving power at m/z = 200. The automatic gain control (AGC) target for MS acquisitions was set to 1×10^6 with a maximum ion injection time of 30 milliseconds. The scan range was set to 400 to 1600 m/z. Data dependent acquisition was set for 12 MS/MS spectra and the dynamic exclusion was set to 30 seconds. The MS/MS resolving power was set to 17.5 k_{FWHM} at m/z=200. The AGC was set to 2×10^5 with a maximum ion injection time of 250 milliseconds.

Mass Spectrometry Protein Identifications

Chromatogram files (.RAW) were processed into a peak list format (.MGF) using Proteome Discoverer (Thermo Scientific, San Jose, CA). The resulting .MGF files were then searched using MASCOT[30] (Matrix Science, Boston, MA) against the striped bass ovary transcriptome (GenBank: SRX007394)[4] translated in all six open reading frames with OrfPredictor[31] and combined into one FASTA file that also contained the protein sequences for the white perch VtgAa, VtgAb, and VtgC (GenBank Accession DQ020120.1, DQ020121.1 and DQ020122.1, respectively) and the white perch vitellogenin receptor (LR8, GenBank Accession AY173045.1) for MASCOT batch search. These four white perch

sequences were included, as the *vtgAa*, *vtgAb*, and *vtgC* are not expressed in the ovary and the full-length vitellogenin receptor ortholog is not represented in the striped bass ovary transcriptome.[4,5] Additionally, the white perch ovary transcriptome was similarly translated and combined into one FASTA file for MASCOT batch search. Sequences for human keratins and porcine trypsin sequences were added to both databases. The protein sequences in both databases also were inverted for use in identifying possible false positive peptides. The parameters for searching the data in MASCOT were as follows: Carbamidomethyl (C) was set as a fixed modification and Carbamyl (K, N-term), Deamidation (NQ), Oxidation (M), Phospho (STY) were then set to variable modifications. The precursor ions search tolerance were 5 ppm, and the fragment ion tolerance was set to ± 0.02 Da. Statistical filtering using a 1% false discovery rate for identifying proteins was performed using ProteoIQ.[32-34]

Semiquantitative analysis was accomplished using spectral counts. Spectral counts are confidently identified tryptic peptides of each protein. Normalized spectral counts (NSpC) were calculated using the total spectral counts within the replicates of each sample and followed with the maximum spectral counts obtained from each ovary membrane and cytosolic replicate analysis. These normalized spectral counts (NSpC) for each of the three technical replicates per biological sample were exported from ProteoIQ and transformed to account for zero values [$\log_{10}(y + 1)$, where y = NSpC].[35]

Striped bass ovary contigs corresponding to the white perch peptides identified by MASCOT were further subjected to BLAST and annotated using Blast2GO as described above. Parameters for blastx were: Expect value 1×10^{-3} and HSP Length Cutoff 33. Parameters for the GO annotations were: E-value-hit-filter 1×10^{-6}, Annotation Cutoff 55, GO Weight 5, and HSP-Hit Coverage Cutoff 0. Combined GO graphs for the annotated

sequences (882 total) were created using percentages of 2nd level GO terms for Biological Process, 3rd level GO terms for Molecular Function, and 6th level GO terms for Cellular Component. Represented GO classes were restricted to those with 50 or more entities (sequence cutoff = 50.0); Sequence Filter = 50, Score alpha = 0.6, Node Score Filter = 10.

Data Analysis

Comparisons of the enrichment of proteins in the white perch ovary cytosolic and membrane fractions were evaluated by one-way analysis of variance (ANOVA, $\alpha = 0.05$) of log10(y + 1) transformed NSpC values using the GLM procedure in the SAS/STAT® software, Version 9.3 of the SAS System for Windows (Copyright © 2012 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA). Following ANOVA, we employed two independent methods of α correction: 1) Standard Bonferroni and 2) Benjamini and Hochberg false discovery rate. Only peptides with $\geq$ 0.5 average NSpC ($NSpC_{avg}$) across all six technical replicates were confidently considered to be enriched in the cytosolic or membrane fraction.[36]

Residual values from the ANOVA for proteins significantly enriched in the membrane or cytosolic fractions ($p < 0.001$) were input for modulated modularity clustering (MMC) performed using Pearson correlation coefficient.[37] This stringent p-value cutoff was implemented to avoid inclusion of false positives in the clustering and also to provide a reasonable catalog of proteins for further analysis by MMC.[36] Similar p-value cutoffs are typically implemented in large-scale gene expression studies.[38] Relevance association networks were generated from the MMC modules using Cytoscape version 2.8.2 (www.cytoscape.org). Interactions between proteins within MMC modules were determined

by correlations exceeding $|r| \geq 0.99$. The DAVID Functional Classification Tool[39] was used to group proteins based on functional similarity within MMC modules. Default parameters for DAVID were used and approved gene abbreviations for all white perch proteins were manually collected from the NCBI or GeneCards.[40]

We performed K-means clustering as an unsupervised learning tool to map protein expression [$\log_{10}(y + 1)$ transformed NSpC values] to either membrane or cytosolic cellular fractions using WEKA version 3.6.7 (http://www.cs.waikato.ac.nz/ml/weka/). The K-means algorithm clusters *n* objects into *k* partitions based on attributes, in this case protein expression. We then evaluated the precision of clustering into 2 clusters using WEKA sequential minimal optimization algorithm SVMs classifier.[5,15,17] We employed two cross-validation strategies to estimate classifier performance: 1) A percentage split whereby 66% of the data were randomly selected and used to train the SVMs and the remaining 33% of the data were input as a cross-validation and 2) A three-fold stratified hold-out with $n = 3$ folds where one fold was used for cross-validation and $n - 1$ folds of the randomly re-ordered dataset were used for training. The SVMs were used to classify samples as either cytosolic or membrane using the complete protein expression data set as well as only those proteins that were not significantly enriched in either fraction by ANOVA (*p*-values > 0.05). Both classes (cytosolic and membrane fractions) were properly represented in the SVMs training and cross-validation data sets. The performance of the SVMs was evaluated as a percent of correct classification during the cross-validation. As a negative control, the NSpC values for the whole data set were randomly reordered and entered into SVMs as described above.

RESULTS

Ovary Histology

The most advanced clutch of follicles in each ovary was of the late vitellogenic stage (oocyte diameter = 538 ± 24 mm) (Figure 1). The frequencies of follicles by stage in the white perch ovary were: 44% primary growth (stage I), 7% perinucleolar (stage II), 8% lipidic (stage III), 8% cortical granule (stage IV), 20% vitellogenic (stage V), 12% late vitellogenic (stage VI), and < 1% atretic (stage VIII). This distribution is typical of the late vitellogenic stage white perch ovary.[19]

White Perch Ovary Transcriptome

A total of 18,031 short reads were assembled into 362 contigs with 1,368 singletons remaining (GenBank: GAQS00000000). The contigs and singletons have average lengths of 339 bp and 223 bp, respectively. The complete list of white perch ovary transcript sequences, in FASTA format, is provided as Supplementary File S1 (Supporting Information). A total of 1,261 sequences were annotated with GO terms (73% of the total 1,730 sequences) and the number of unknown unique sequences was 30. The GO annotation breakdown of all annotated sequences is shown in Figure 2: Cellular Component (4^{th} level), Molecular Function (3^{rd} level), and Biological Process (2^{nd} level).

Tandem Mass Spectrometry

The complete data from the nanoLC-MS/MS and ProteoIQ are provided as Supplementary Files S2 and S3 (Supporting Information). A total of 310 proteins were identified when the nanoLC-MS/MS data were searched against the white perch ovary transcriptome (Supplementary File S3). A total of 882 unique proteins were unambiguously

identified using the striped bass ovary transcriptome as the reference database (Supplementary File S2). Of these, 28 (~3.1%) proteins fell below the $NSpC_{avg}$ confidence cutoff of 0.5. When the proteins identified using the white perch ovary transcriptome were compared to the proteins identified using the striped bass ovary transcriptome, 227 orthologs were found (~73% of the 310 proteins identified) (Figure 3). The GO annotation breakdown of all sequences identified using the striped bass ovary transcriptome is shown in Figure 4: Cellular Component (6^{th} level), Molecular Function (3^{rd} level), and Biological Process (2^{nd} level).

Using the Bonferroni correction, none of the proteins were significantly enriched in either fraction (adjusted p-value $< 5.66 \times 10^{-5}$). Using the Benjamini and Hochberg procedure, 618 proteins (~70% of 882 identified proteins) were significantly enriched in either the cytosolic or membrane fraction by ANOVA ($p \leq 0.0347$) using the striped bass ovary transcriptome as the reference database (Table 1). Of these 618 proteins, 262 were enriched in the cytosolic fraction and 356 were enriched in the membrane fraction. Figure 5 illustrates 114 white perch ovary proteins identified exclusively in the cytosolic fraction, 169 proteins found exclusively in the membrane fraction, and 599 proteins found in both fractions. The 25 proteins most significantly enriched in the cytosolic or membrane fraction are listed in Table 2. The remaining 264 proteins did not vary significantly ($p > 0.0347$) between cytosolic and membrane fractions by ANOVA (Table 3).

Of the 882 proteins identified, approximately 2.7% (24 of 882) were expressed in at least two open reading frames (Table 4). These putative alternatively spliced variants are noted with an additional letter (a, b, or c) following their respective contig numbers.

The three forms of white perch vitellogenin were significantly enriched in the cytosolic fraction (ANOVA $p < 0.0001$) (Figure 6). The VtgAb was the most dominant of all proteins

identified in this study. The ratio of VtgAa:VtgAb:VtgC based on the total NSpC values from cytoplasmic and membrane fractions is 7.7:16:1.

Support Vector Machines Classification of Ovary Cytosolic and Membrane Fractions

The SVMs were able to classify the samples as either cytosolic or membrane fractions perfectly (100% correct classification) during cross-validation using both the 66% percentage split and the three-fold stratified hold-out methods. The SVMs performed the same when expression data of the 882 proteins identified using the striped bass ovary transcriptome reference database and the 310 proteins identified using the white perch ovary transcriptome reference database were used as inputs. Additionally, SVMs were equally precise at classification when supplied with only 242 proteins (~25% of all identified proteins) with the highest ANOVA *p*-values (0.0643-0.9979, i.e. those proteins that were not significantly enriched for either cellular fraction). When the data were randomized, the SVMs were unable to predict the cell fraction and this is expected (i.e. the correct assignment during cross-validation = 50% or equivalent to that of random chance when classifying data into two categories).

Modulated Modularity Clustering

Proteins significantly enriched for either cytosolic or membrane fraction (ANOVA $p < 0.001$, 277 proteins) were organized on the basis of co-variable expression into 16 modules using MMC (Figure 7A). Relevance association networks within modules are depicted with edges between protein nodes determined by correlations exceeding a threshold value ($|r| \geq 0.99$) (Figure 7B).

DAVID Analysis

A total of 16 modules were generated by MMC, 14 of which contained more than 6 proteins. We used DAVID to assess for these 14 MMC modules the degree to which GO biological processes and pathways are overrepresented (Table 5). Eight of these 14 modules showed significant GO enrichment. Seven modules (2, 7, 8, 9, 13, 14, and 15) showed significant enrichment for mitochondrial proteins involved in mitochondrial protein synthesis, electron transport chain, and proton-motive force. Module 10 is the largest module containing 43 proteins with enrichment for proteasome, cell cycle, protein synthesis, protein folding, and metal ion binding.

DISCUSSION

The present study provides the first proteomic characterization of white perch ovary tissues, utilizing cellular fractionation to maximize coverage of the cytosolic and membrane cell fractions. Proteome resources are available for several commercially important finfishes including channel catfish (*Ictalurus punctatus*),[41] Atlantic salmon (*Salmo salar*),[42,43] rainbow trout (*Oncorhynchus mykiss*),[44-46] Senegalese sole (*Solea senegalensis*),[47] yellow perch (*Perca flavescens*),[48] gilthead seabream (*Sparus aurata*),[49,50] European seabass (*Dicentrarchus labrax*),[51] zebrafish (*Danio rerio*),[50.52-54] fathead minnow (*Pimephales promelas*)[55] and striped bass (*Morone saxatilis*).[5] These resources, however, do not yet exist for white perch, despite their importance as a bioindicator species and in the study of teleost fish reproduction.[1,2,3,8-10,52] Many of the proteins identified in these studies mirror those we found in striped bass, including metabolic enzymes, chaperones, and regulators of protein synthesis (Figures 2 and 4), which is typical of the fish ovary.[5]

The proportion of white perch ovary proteins identified using the striped bass ovary transcriptome as the reference database is 7.86% (882 proteins/11,208 database transcripts), while the proportion of proteins identified using the white perch ovary transcriptome is 17.91% (310 proteins/1,730 database transcripts). Although the homologous white perch ovary transcriptome yielded the higher number of protein identifications per transcript sequence, we used the protein data identified with the heterologous striped bass ovary transcriptome for downstream analyses due to the greater number of overall protein identifications.

Of the 882 proteins identified, 283 (32%) were exclusive to either the cytosolic or membrane fractions (Figure 5). Additionally, subunits of the 26S-proteasome, a multi-subunit cytosolic macromolecular complex, were only significantly enriched in the cytosolic fraction. Cellular fractionation also enabled detection of an abundance of membrane-bound mitochondrial proteins in the membrane fraction, none of which were enriched in the cytosolic fraction. This indicates that the fractionation methods employed were able to effectively separate membrane-associated and cytosolic proteins.

Seven of eight MMC modules had significant DAVID GO enrichment for the mitochondrion, five of which are shown in Table 5. DAVID analysis of Module 10 revealed four annotation clusters corresponding to proteasome, cell cycle, protein synthesis, protein folding, and metal ion binding. Modules 8 and 9 also contain clusters enriched for components of the 26S-proteasome, supporting its regulation on multiple levels.[5]

Beyond ATP production, regulation of the fertilization Ca^{2+} oscillation, and mediating proapoptotic cascades,[56] the precise roles of mitochondria in oocytes and embryos are largely unknown in general and in teleost fishes in particular. The difficulty in accessing the mitochondrial compartment has been a major obstacle in advancing the understanding of

these additional roles.[56-59] Mitochondria are stockpiled during oogenesis, reaching peak abundance in fully mature oocytes that remains unchanged through early embryonic development.[60] Previtellogenic *Xenopus laevis* oocytes are characterized by rapidly increasing numbers of mitochondria and their ovulated eggs have remarkably been shown to contain ~10^7 mitochondria each.[57-59] The fractionation method we employed was able to enrich the detection of mitochondrial proteins and was likely aided by the relatively high abundance of mitochondria present in late vitellogenic white perch oocytes. Given their role in activation of the caspase pathway and apoptosis,[62] mitochondria are interesting candidates for further investigation into their role in egg quality of striped bass, particularly in relation to the phenomena of ovarian atresia and egg overripening.[63]

Alternative gene transcript splicing may underlie species-specific differences in phenotype.[64] The putative spliced variant proteins listed in Table 4 encompass a diverse array of biological functions including mitochondrial beta-oxidation and tricarboxylic acid cycle (*eci1* and *aco2*, respectively), deubiquitination (*uchl3*), nucleocytoplasmic transport (*ipo5*), molecular chaperone (*cct2*, *cct7,* and *hsp90b1*), and cell cycle regulation (*vcp*). Alternatively spliced variants have previously been reported for *eci1*, *cct2*, *cct7*, and *uchl3*. Of the 5 putative spliced variants included in the MMC modules, 3 were also included in the relevance networks (Figure 7B). Specifically, *aco2* (contig01840b) was included in network 9, and unnamed contig09506a and *cct5* (contig09515b) were included in network 10. Two of these putative spliced variants, *aco2* and *cct5*, are located at or near the central network hubs and have perfect connectivity to all nodes within their respective relevance networks. The remaining putative spliced variants are members of highly intra-connected relevance networks that contain many edges. Proteins located at the center of relevance association networks have many interaction partners and are often associated with signaling,

development, and disease and generally may be ascribed a regulatory function.[65] Further investigation into alternatively spliced variants may uncover distinct regulatory functions for these proteins in fish oocytes and developing embryos.

Support vector machines have been used to accurately classify gene and protein expression patterns in response to endocrine disruption and during different ovarian growth stages in fishes.[5, 66-69] The utility of SVMs for classification of proteomics data that are not significant by linear statistics is evident in the present study. While we unambiguously identified 618 proteins (~70% of the total proteins) by mass spectrometry that were significantly enriched in either the cytosolic or membrane fraction by one-way ANOVA ($p \leq 0.0347$), the remaining 264 proteins did not significantly vary in abundance between cytosolic and membrane fractions (Table 1). When the expression data of 242 proteins with ANOVA *p*-values = 0.0643-0.9979 were input, SVMs were able to classify the samples as cytosolic or membrane fraction with perfect (100%) precision during cross-validation. Therefore, SVMs are an effective analytical method when it comes to classification of proteomics data based upon cell fraction, even when expression data of proteins that were not significantly enriched for either fraction were used as the input.

The three white perch vitellogenins were shown to comprise 31.2% VtgAa, 64.8% VtgAb, and 4.0% VtgC of the total vitellogenin-derived egg yolk in late vitellogenic oocytes (Supplementary File S2; Figure 6). This ratio (7.7:16:1, respectively) more closely resembles that of barfin flounder (*Verasper moseri*) than that of the closely related striped bass (1.43:1.4:1, or 37.38% VtgAa, 36.55% VtgAb, and 26.07% VtgC).[70,71] These findings further support the concept of species-specific variation in egg yolk composition of fishes with multiple vitellogenin systems.[72] Such a stark difference in yolk composition between the white perch and the closely related striped bass may relate to distinct patterns of oocyte

hydration that could underlie differences in reproductive strategies. Specifically, striped bass are anadromous, entering brackish and fresh water from the sea to spawn their neutrally buoyant, pelagic eggs. While marine populations do exist, white perch are primarily a freshwater species that spawns adhesive, semipelagic eggs in estuaries, rivers, lakes, and marshes.[73] The proportional yolk composition in oocytes of various fishes to date indicates that VtgC is the most variable vitellogenin-derived egg yolk component between species. For instance, in barfin flounder, grey mullet (*Mugil cephalus*) and white perch, VtgC derived yolk proteins are minor and comprise ~5% of yolk in late vitellogenic oocytes,[70,74] whereas in mosquitofish (*Gambusia affinis*)[75] and striped bass[70] it comprises a significantly larger proportion of the yolk (20-30%). Slight variation in composition of VtgAa and VtgAb are observed between species, however these are typically between 1:1 and 1:2 (VtgAa:VtgAb) with the exception of the goldsinny wrasse (*Ctenolabrus rupestris*).[76] This indicates that VtgC is likely an important factor in egg yolk composition and differences in its accumulation by oocytes of diverse fish species may relate to specific early life histories and reproductive strategies. Further investigation is required to more fully understand this phenomenon.

CONCLUSIONS

Membrane proteins perform crucial cellular functions and are difficult to characterize due to their amphipathic nature and association with membrane lipids.[77] Mature fish oocytes are complex cells with a highly-yolked cytoplasm, making characterization of their membrane proteins even more challenging. The relatively simple and inexpensive compartment proteomics methods we employed helped to enrich detection of membrane-associated proteins, particularly those related to the mitochondrion and other membrane-bounded

organelles. Additionally, we show that SVMs are a superior analytical method for classification of proteomics data based upon cell fraction.

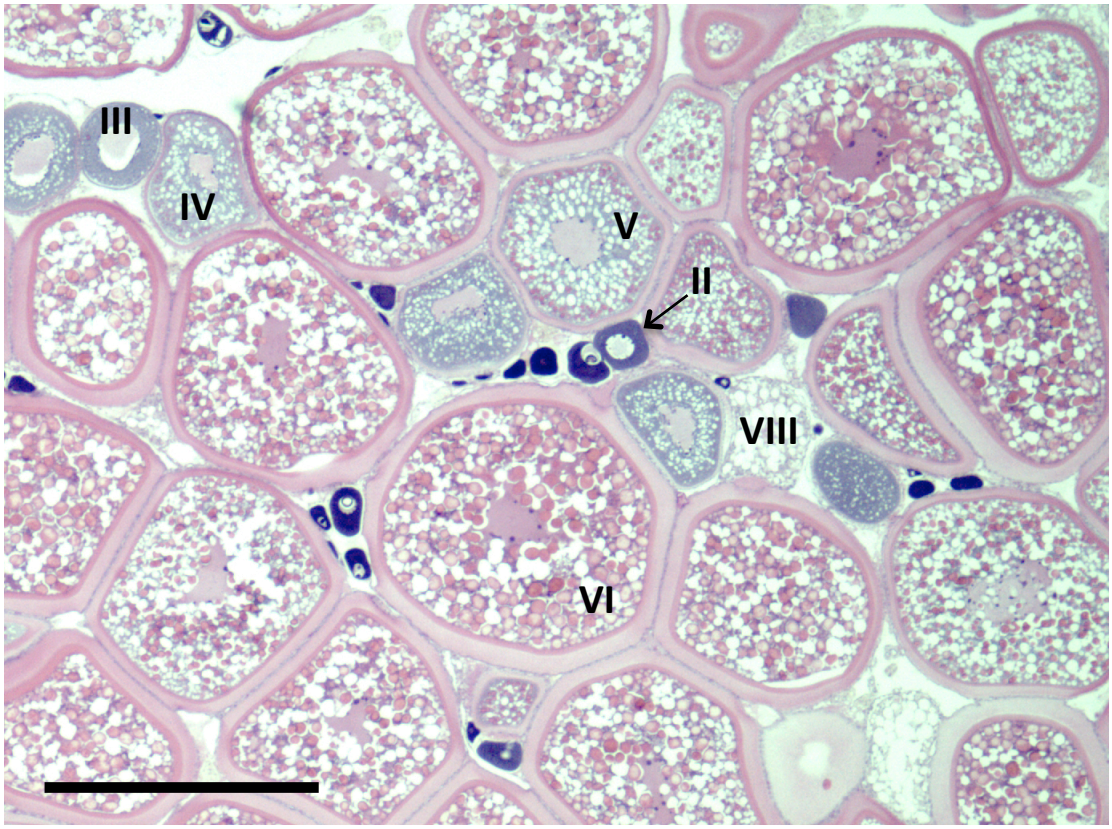


Figure 1. Hematoxylin and eosin staining of a representative white perch ovary collected in March 2013 depicting all stages of ovarian growth including perinucleolar (stage II), lipidic (stage III), cortical granule (stage IV), vitellogenic (stage V), late vitellogenic (stage VI), and atretic (stage VIII) follicles. The primary growth oocytes are unlabeled but appear darkly stained with basophilic dye. Bar = 500 mm.

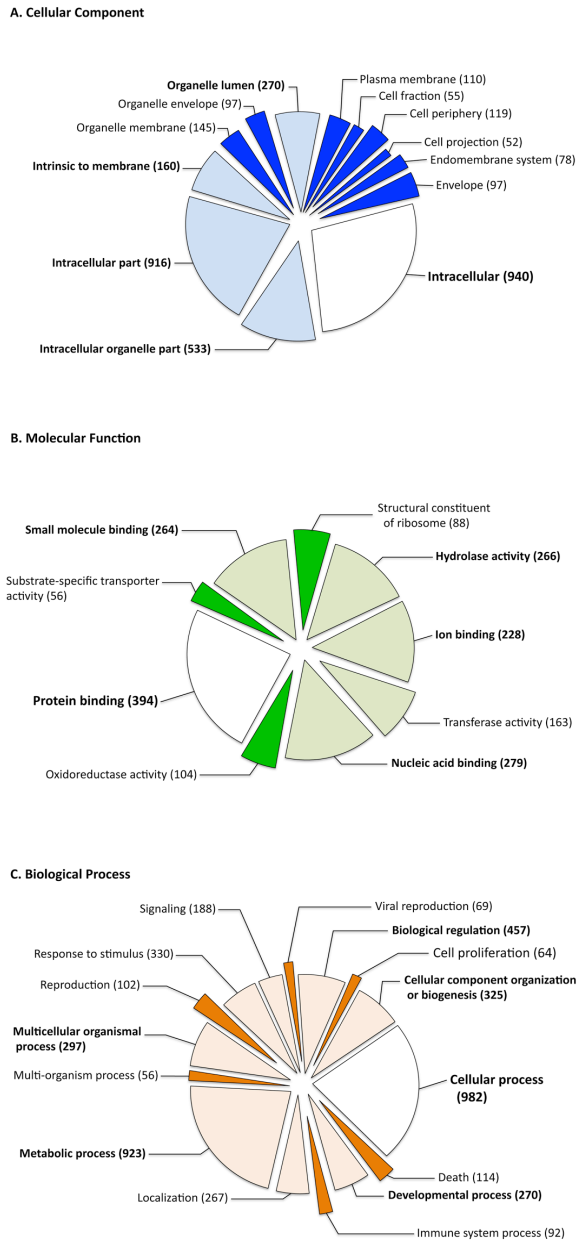


Figure 2. Gene ontology graph of (A) Cellular Component (4th level GO terms), (B) Molecular Function (3rd level GO terms), and (C) Biological Process (2nd level GO terms) of annotated genes in the white perch ovary transcriptome. The number of GOs in each class is shown and sections that contained 50-100 entries are represented by dark color, 100 and up by light color, and the predominant class is indicated by white.

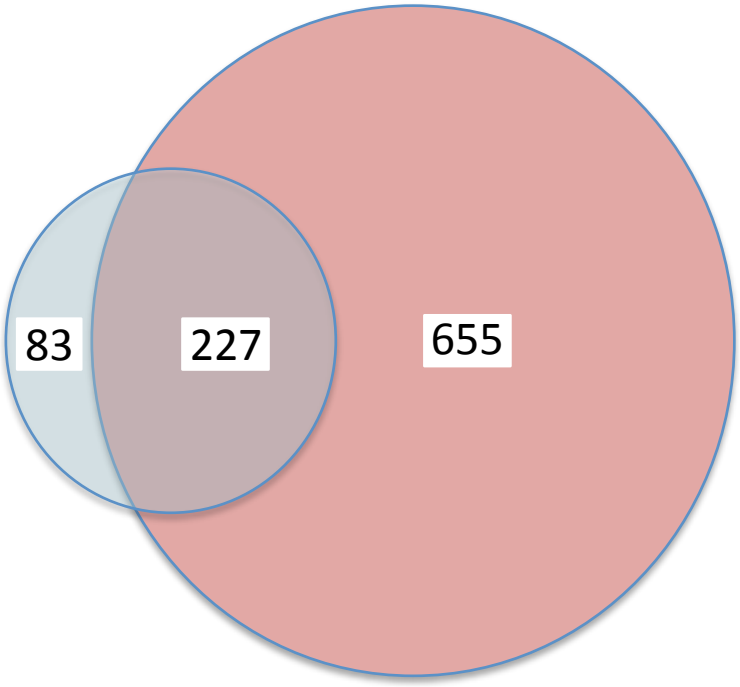


Figure 3. Venn diagram depicting white perch ovary proteins uniquely and commonly detected using the white perch ovary transcriptome (blue) and the striped bass ovary transcriptome (red) as reference databases.

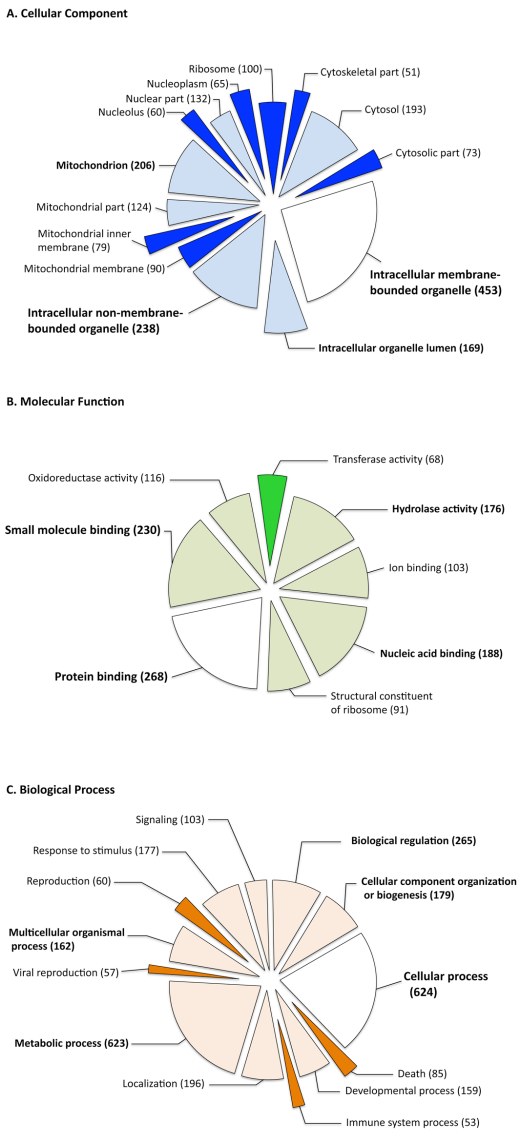


Figure 4. Gene ontology graph of **(A)** Cellular Component (6th level GO terms), (**B**) Molecular Function (3rd level GO terms), and (**C**) Biological Process (2nd level GO terms) of annotated genes in the white perch ovary proteome using the striped bass ovary transcriptome as the reference database. The number of GOs in each class is shown and sections that contained 50-100 entries are represented by dark color, 100 and up by light color, and the predominant class is indicated by white.

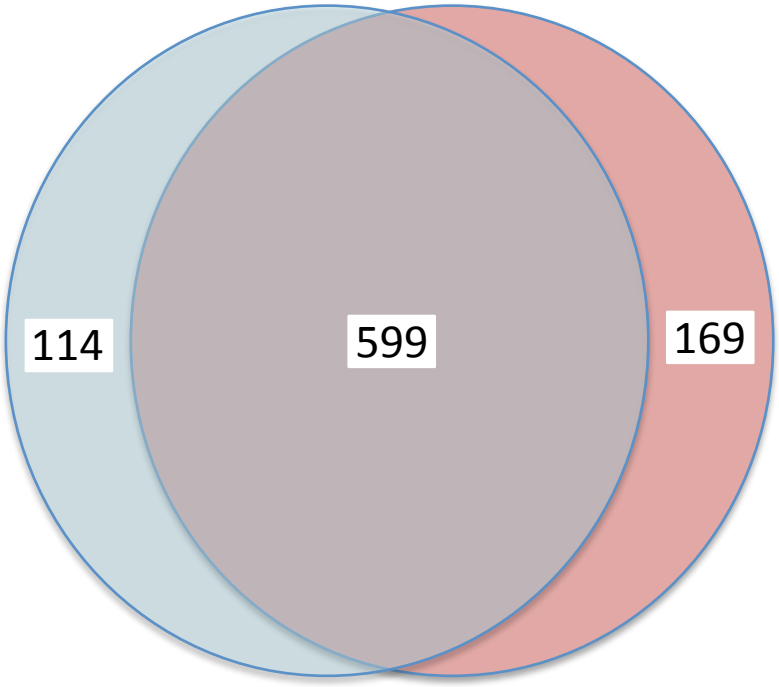


Figure 5. Venn diagram depicting the 114 white perch ovary proteins identified only in the cytosolic fraction (blue), the 169 proteins found only in the membrane fraction (red), and the 599 proteins found in both fractions using the striped bass ovary transcriptome as the reference database.

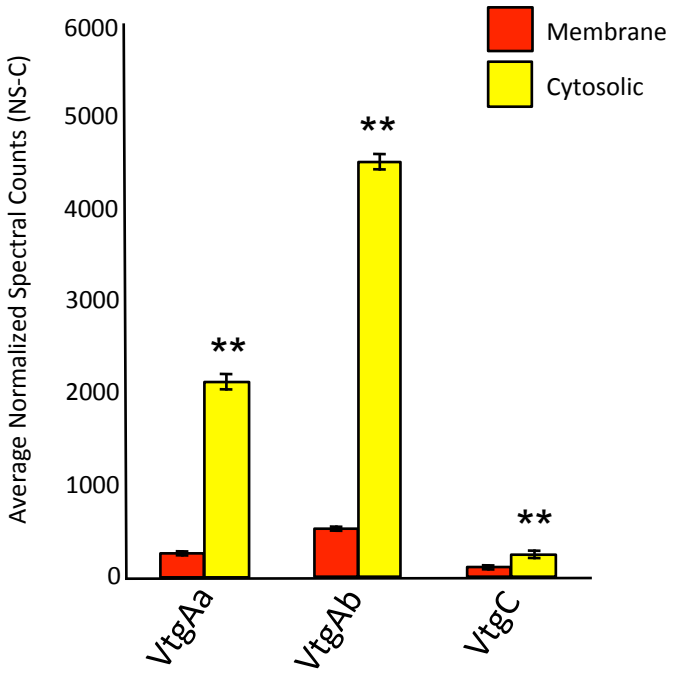


Figure 6. Average normalized spectral count (NSpC) values of VtgAa, VtgAb, and VtgC, in white perch late vitellogenic ovary cytosolic and membrane fractions (**ANOVA p-value < 0.0001).

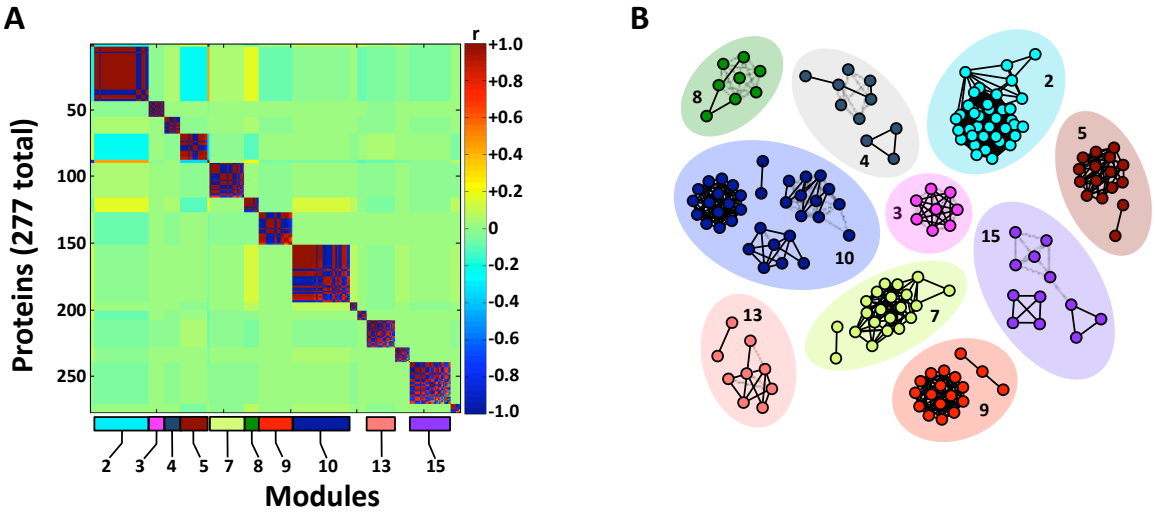


Figure 7. (**A**) Modulated modularity clustering (MMC) heat map of 277 correlated white perch ovary cystosolic and membrane proteins with ANOVA *p*-values < 0.001 (16 modules). The proteins were identified by tandem mass spectrometry using the striped bass ovary transcriptome as the reference database. (**B**) Relevance networks of ovary proteins correlated within MMC modules ($|r| \geq 0.99$). Color-coding of networks corresponds to the horizontal bar shown under the heat map in panel A. Only those relevance networks with 6 or more co-variable proteins are depicted.

Table 1. ANOVA *p*-value range of 882 proteins identified in white perch ovary cytosolic and membrane fractions using the striped bass ovary transcriptome as the reference database. Employing the Benjamini and Hochberg procedure resulted in a significance cutoff of $p \leq 0.0347$.

***p*-value range**	**number of proteins**
< 0.0001	152
0.0001-0.001	125
0.001-0.0347	341
> 0.0347	264
Total	**882**

Table 2. Approved gene name, striped bass contig number, cytosolic and membrane normalized [$\log_{10}(1+ \text{NSpC}_{avg})$] values, and ANOVA *p*-values for white perch ovary proteins identified in either the cytosolic or membrane fraction using the striped bass ovary transcriptome as the reference database.

gene name	contig	cytosolic	membrane	*p*-value
psmc1	00084	1.29	0.00	<0.0001
nasp	10035	1.13	0.00	<0.0001
atp6v0a1	01914	1.03	0.00	<0.0001
eif5a	09973	0.93	0.00	<0.0001
eif3g	09716	0.91	0.00	<0.0001
st13	10088	0.89	0.00	<0.0001
atox1	10100	0.85	0.00	<0.0001
dapl1	10937	0.83	0.00	<0.0001
wdr77	09418	0.78	0.00	<0.0001
glo1	09544	0.76	0.00	<0.0001
tmsb10	10743	0.76	0.00	<0.0001
uchl3	09596	0.68	0.00	<0.0001
gpi	02320	0.68	0.00	<0.0001
psmd8	00653	0.67	0.00	<0.0001
eif2s3	00484	0.67	0.00	<0.0001
nasp	05921	0.64	0.00	<0.0001
psmd7	09369	0.64	0.00	<0.0001
psmd11	09512	0.70	0.00	0.0001
uba1	08566	0.83	0.00	0.0002
arhgdia	07299	0.71	0.00	0.0002
adsl	00964	0.67	0.00	0.0003

Table 2 (continued)

psmc4	03016	0.67	0.00	0.0003
psmb7	01014	0.68	0.00	0.0004
set	10700	0.61	0.00	0.0009
cct7	03096c	0.68	0.00	0.0021
ndufa9	01307	0.00	1.21	<0.0001
pdha1	02360	0.00	1.16	<0.0001
zp2	09768	0.00	1.13	<0.0001
ndufab1	00326	0.00	1.06	<0.0001
cox5a	01120	0.00	1.05	<0.0001
cyc1	10402	0.00	1.05	<0.0001
a	00014	0.00	1.01	<0.0001
h2afv	00769	0.00	0.99	<0.0001
ndufb10	11204	0.00	0.95	<0.0001
cox6b1	00746	0.00	0.95	<0.0001
clpp	10012	0.00	0.93	<0.0001
mrpl11	10899	0.00	0.93	<0.0001
pdhb	00913	0.00	0.93	<0.0001
atp5i	11097	0.00	0.93	<0.0001
ndufb4	01462	0.00	0.91	<0.0001
csnk2a1	00757	0.00	0.91	<0.0001
mrpl46	10483	0.00	0.89	<0.0001
mrpl19	00670	0.00	0.89	<0.0001
uqcrc1	04683	0.00	0.86	<0.0001
lsm14b	09506	0.00	0.86	<0.0001
ndufa6	02334	0.00	0.86	<0.0001
coq6	00590	0.00	0.84	<0.0001
syncrip	08534	0.00	0.84	<0.0001
snrpa1	09352	0.00	0.84	<0.0001

[a] indicates that no approved gene name exists.

Table 3. Approved gene name, striped bass contig number, cytosolic and membrane normalized [$\log_{10}(1+ NSpC_{avg})$] values, and ANOVA *p*-values for white perch ovary proteins identified in both cytosolic and membrane fractions using the striped bass ovary transcriptome as the reference database.

gene name	contig	cytosolic	membrane	*p*-value
pfn2	03911	0.91	0.91	0.9979
rab1a	09027	0.53	0.55	0.9714
cfl2	02581	0.23	0.26	0.9691
e1f4e	02291	0.76	0.78	0.9603
chmp4	09068	0.12	0.15	0.9433
rap1a	05887	0.48	0.43	0.9303
sod1	09354	0.48	0.43	0.9234
rtn4	00963	0.53	0.43	0.9146
habp4	09906	0.30	0.27	0.9146
rps20	11198	0.98	0.97	0.9136
larp7	04517	0.30	0.27	0.9114
arf1	02467	0.31	0.27	0.9072
akr1a1	00347	0.37	0.35	0.8985
tuba4a	10875	1.59	1.59	0.8965
ifi30	10433	0.48	0.49	0.8958
tuba1b	10905	1.08	1.06	0.8923
cct7	03096a	0.31	0.36	0.8876
nhp2	07687	0.37	0.43	0.8864
psmc3	06251	0.30	0.35	0.8788
myh9	03682	0.53	0.49	0.8685
mrps21	11032	0.23	0.26	0.8676
a	01051b	0.85	0.86	0.863
kpnb1	09333	0.23	0.26	0.8619
rars	10188	0.23	0.15	0.8539
ppp1cc	01459	0.22	0.27	0.8477

[a] indicates that no approved gene name exists.

Table 4. Approved gene name, striped bass contig number, cytosolic and membrane normalized [$\log_{10}(1+ NSpC_{avg})$] values, and ANOVA *p*-values for white perch ovary proteins identified in multiple open reading frames using the striped bass ovary transcriptome as the reference database.

gene name	contig	cytosolic	membrane	*p*-value
vcp	10277a	1.33	1.12	0.0072
vcp	10277b	0.85	0.68	0.1543
vcp	10277c	1.14	0.68	0.0023
uchl3	09596a	0.64	0.15	0.0134
uchl3	09596b	0.67	0.00	0.0021
a	01051a	1.15	0.75	0.0011
a	01051b	0.85	0.86	0.8455
a	09506a	0.48	1.22	0.0005
a	09506b	0.00	0.86	<.0001
ipo5	02029a	1.25	0.75	0.0003
ipo5	02029b	0.83	0.26	0.0095
hsp90b1	00400a	0.31	0.84	0.0004
hsp90b1	00400b	1.50	1.86	0.0301
eci1	01183a	0.78	1.16	0.0145
eci1	01183b	0.12	0.91	0.0014
cct7	03096b	0.76	0.43	0.0155
cct7	03096c	0.67	0.00	0.0021
cct5	09515a	0.97	0.35	0.0015
cct5	09515b	1.47	0.60	0.0007
cct2	00164a	1.31	0.43	0.0066
cct2	00164b	0.85	0.55	0.0744
aco2	01840a	0.48	1.12	0.0045
aco2	01840b	0.64	1.33	<.0001

[a] indicates that no approved gene name exists.

Table 5. Enrichment of white perch ovary proteins by Gene Ontology (GO) class within modulated modularity clustering (MMC) modules 2, 7, 8, 9, and 10 using DAVID. The proteins were identified by tandem mass spectrometry using the striped bass ovary transcriptome as the reference database.

MMC module	GO Class and protein members (contig)	*P*-value	enrichment score
2	**Mitochondria**	**1.0×10^{-1}**	**0.98**
	hspa9 (01684), *ilf2* (00470), *pitrm1* (02629), *ndufs8* (10963), *cyb5a* (10385)		
7	**Mitoribosome**	**3.1×10^{-4}**	**1.99**
	mrpl14 (03533), *mrpl19* (00670), *mrpl23* (06544), *mrps34* (00091)		
8	**Mitochondria**	**7.4×10^{-5}**	**2.75**
	atpj5 (10488), *atp5c1* (01361), *uqcrh* (01584)		
9	**Mitochondria**	**1.0×10^{-3}**	**2.73**
	atpj5 (10488), *atp5l* (00098), *uqcrq* (09375), *slc25a6* (01147)		
10	**Proteasome**	**5.4×10^{-2}**	**0.89**
	psmb6 (10749), *psmb7* (01014), *thop1* (00710)		
	Ribosome	**2.3×10^{-1}**	**0.49**
	cdk1 (11033), *cct5* (9515), *eif2s3* (00484), *wars* (05574)		
	Proteasome	**9.2×10^{-1}**	**0.03**
	impdh2 (00960), *eno1* (04955), *thop1* (00710)		

The corresponding striped bass contig number for each approved gene name is indicated in parentheses.

ASSOCIATED CONTENT

Supporting Information. Supplementary File S1: The white perch ovary transcriptome. Supplementary File S2: A table of 882 unique proteins unambiguously identified in the late vitellogenic white perch ovary when the nanoLC-MS/MS and ProteoIQ data were searched against the striped bass ovary transcriptome as the reference database. Supplementary File S3: A table of 310 unique proteins unambiguously identified in the late vitellogenic white perch ovary when the nanoLC-MS/MS and ProteoIQ data were searched against the white perch ovary transcriptome as the reference database. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: (919) 515-3380. Fax: (919) 515-5327. E-Mail: bjreadin@ncsu.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Brad Ring and John Davis (North Carolina State University Aquaculture Facility) for care and maintenance of the white perch.

ABBREVIATIONS

ANOVA, analysis of variance; SVMs, support vector machines; MMC, modulated modularity clustering; FASP, filter aided sample preparation; DTT, dithiothreitol; LC, liquid chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; nanoLC, nanoscale liquid chromatography; nanoLC-MS/MS, nanoscale liquid chromatography coupled to tandem mass spectrometry; AGC, automatic gain control; GO, gene ontology; Vtg, vitellogenin

REFERENCES

1. Barnthouse, L. W.; Glaser, D.; DeSantis, L. Polychlorinated biphenyls and Hudson River white perch: Implications for population-level ecological risk assessment and risk management. *Integr. Environ. Assess. Manag.* **2009**, *5*, 435.

2. King, R. S.; Beaman, J. R.; Whigham, D. F.; Hines, A. H.; Baker, M. E.; Weller, D. E. Watershed land use is strongly linked to PCBs in white perch in Chesapeake Bay subestuaries. *Environ. Sci. Technol.* **2013**, *38*, 6546–52.

3. Hiramatsu, N.; Matsubara, T.; Fujita, T.; Sullivan, C. V.; Hara, A. Multiple piscine vitellogenins: biomarkers of fish exposure to estrogenic endocrine disruptors in aquatic environments. *Marine Biol.* **2006**, *149*, 35–47.

4. Reading, B. J.; Chapman, R. W.; Schaff, J. E.; Scholl, E. H.; Opperman, C. H.; Sullivan, C. V. An ovary transcriptome for all maturational stages of the striped bass (*Morone saxatilis*), a highly advanced perciform fish. *BMC Res Notes* **2012**, *5* (111), 1-12, DOI: http://dx.doi.org/10.1186/1756-0500-5-111.

5. Reading, B. J.; Williams, V. N.; Chapman, R. W.; Williams, T. I.; Sullivan, C. V. Dynamics of the striped bass (*Morone saxatilis*) ovary proteome reveal a complex network of the translasome. *J. Proteome Res.* **2013**, *12*, 1691-99.

6. Tao, Y.; Berlinsky, D. L.; Sullivan, C. V. Characterization of a vitellogenin receptor in white perch (*Morone americana*). *Biol. Reprod.* **1996**, *55*, 646–56.

7. Hiramatsu, N.; Luo, W.; Reading, B. J.; Sullivan, C. V. Multiple ovarian lipoprotein receptors in teleosts. *Fish Physiol. Biochem.* **2013**, *39,* 29–32, DOI: http://dx.doi.org/10.1007/s10695-012-9612-6.

8. Reading, B. J.; Hiramatsu, N.; Sawaguchi, S.; Matsubara, T.; Hara, A.; Lively, M. O.; Sullivan, C. V. Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. *Mar. Biotechnol.* **2009**, *11*, 169–87.

9. Reading, B. J.; Hiramatsu, N.; Sullivan, C. V. Disparate binding of three types of vitellogenin to multiple forms of vitellogenin receptor in white perch. *Biol. Reprod.* **2011**, *84*, 392–99.

10. Hiramatsu, N.; Matsubara, T.; Hara, A.; Donato, D. M.; Hiramatsu, K.; Denslow, N. D.; Sullivan, C. V. Identification, purification and classification of multiple forms of vitellogenin from white perch (*Morone americana*). *Fish Physiol. Biochem.* **2002**, *26*, 355–70.

11. Hiramatsu, N.; Chapman, R. W.; Lindzey, J. K.; Haynes, M. R.; Sullivan, C. V. Molecular characterization and expression of vitellogenin receptor from white perch (*Morone americana*). *Biol. Reprod.* **2004**, *70*, 1720–30.

12. Hiramatsu, N.; Luo, W.; Reading, B. J.; Sullivan, C. V.; Mizuta, H.; Ryu, Y. W.; Nishimiya, O.; Todo, T.; Hara, A. Multiple ovarian lipoprotein receptors in teleosts. *Fish Physiol. Biochem.* **2012,** *39*, 29–32.

13. Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nature* **2010,** *7,* 681-85.

14. Rockstroh, M.; Müller, S.; Jende, C.; Kerzhner, A.; Bergen, Von, M.; Tomm, J. M. Cell fractionation - an important tool for compartment proteomics. *JIOMICS* **2011**, *1*, 135–43.

15. Witten, I. H.; Frank, E.; Hall, M. A. *Data Mining: Practical Machine Learning Tools and Techniques,* 3rd Edition; Elsevier: Burlington, MA, USA, **2011**.

16. Chapman, R. W.; Mancia, A.; Beal, M.; Veloso, A.; Rathburn, C.; Blair, A.; Sanger, D.; Holland, A.F.; Warr, G.W.; Didonato, G. A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica*, in the South Atlantic bight. *Mol. Ecol.* **2009**, *18*, 2415–25.

17. Johansson, P.; Ringnér, M. Classification of Genomic and Proteomic Data Using Support Vector Machines. *Fundamentals of Data Mining in Genomics and Proteomics*; Springer: Boston, MA, USA, **2007**, 187–202.

18. King, W., V; Berlinsky, D. L.; Sullivan, C. V. Involvement of gonadal steroids in final oocyte maturation of white perch (*Morone americana*) and white bass (*M. chrysops*): *in vivo* and *in vitro* studies. *Fish Physiol. Biochem.* **1995,** *14*, 489–500.

19. Jackson, L. F.; Sullivan, C. V. Reproduction of white perch: The annual gametogenic cycle. *T. Am. Fish. Soc.* **2005**, *124*, 563–77.

20. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999,** *9*, 868-77, DOI: http://dx.doi.org/10.1101/gr.9.9.868.

21. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–10.

22. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29.

23. Götz, S.; García-Gómez, J. M.; Terol, J.; Williams, T. D.; Nagaraj, S. H.; Nueda, M. J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–35.

24. Conesa, A.; Götz, S. Blast2GO: A Comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, 1–12, DOI: http://dx.doi.org/10.1155/2008/619832.

25. Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–6.

26. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–62.

27. Andrews, G. L.; Shuford, C. M.; Burnett, J. C.; Hawkridge, A. M.; Muddiman, D. C. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chromatogr. B* **2009**, *877*, 948–54.

28. Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q

Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell Proteomics* **2011**, *10*, 1-11, DOI: http://dx.doi.org/10.1074/mcp.M111.011015.

29. Randall, S. M.; Cardasis, H. L.; Muddiman, D. C. Factorial experimental designs elucidate significant variables affecting data acquisition on a quadrupole orbitrap mass spectrometer. *J. Am. Soc. Mass Spectr.* **2013**, *24*, 1501–12.

30. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–67.

31. Min, X. J.; Butler, G.; Storms, R.; Tsang, A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucl. Acids Res.* **2005**, *33*, W677–80, DOI: http://dx.doi.org/10.1093/nar/gki394.

32. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383–92.

33. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–58.

34. Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol. Cell. Proteomics* **2005**, *4*, 762-72.

35. Wu, X.; Chau-Wen, T.; Edwards, N. HMMatch: Peptide identification by spectral matching of tandem mass spectra using hidden Markov models. *J. Comput. Biol.* **2007**, *14* (8), 1025-43, DOI: http://dx.doi.org/10.1089/cmb.2007.0071.

36. Gokce, E., Shuford, C. M., Franck, W. L., Dean, R. A., Muddiman, D. C. Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows. *J. Am. Soc. Mass Spectrom*. **2011**, *22*, 2199-208.

37. Stone, E. A.; Ayroles, J. F. Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet*. **2009**, *5*, e1000479, DOI: http://dx.doi.org/10.1371/journal.pgen.1000479.

38. Chan, Y. C.; Raengpradub, S.; Boor, K. J.; Wiedmann, M. Microarray-Based Characterization of the *Listeria monocytogenes* Cold Regulon in Log- and Stationary-Phase Cells. *Appl. Environ. Microbiol.* **2007**, *73* (20), 6484, DOI: http://dx.doi.org/10.1128/AEM.00897-07.

39. Dennis, G., Jr; Sherman, B. T.; Hosack, D. A.; Yang, J. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **2003**, *4* (9), R60.1-11.

40. Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* **1998**, *14*, 656–64.

41. Booth, N. J.; Bilodeau-Bourgeois, A. L. Proteomic analysis of head kidney tissue from high and low susceptibility families of channel catfish following challenge with *Edwardsiella ictaluri*. *Fish Shellfish Immun.* **2009**, *26*, 193–6.

42. Provan, F.; Bjørnstad, A.; Pampanin, D. M.; Lyng, E.; Fontanillas, R.; Andersen, O. K.; Koppe, W.; Bamber, S. Mass spectrometric profiling – a diagnostic tool in fish? *Mar. Environ. Res.* **2006**, *62* (Suppl), S105–8.

43. Liu, X.; Afonso, L.; Altman, E.; Johnson, S.; Brown, L.; Li, J. O-acetylation of sialic acids in N-glycans of Atlantic salmon (*Salmo salar*) serum is altered by handling stress. *PROTEOMICS* **2008**, *8*, 2849–57.

44. Martin, S. A. M.; Vilhelmsson, O.; Médale, F.; Watt, P.; Kaushik, S.; Houlihan, D. F. Proteomic sensitivity to dietary manipulations in rainbow trout. *BBA Proteins Proteom.* **2003**, *1651*, 17–29.

45. Rime, H.; Guitton, N.; Pineau, C.; Bonnet, E.; Bobe, J.; Jalabert, B. Post-ovulatory ageing and egg quality: a proteomic analysis of rainbow trout coelomic fluid. *Reprod. Biol. Endocrinol.* **2004**, *2*, 26, DOI: http://dx.doi.org/10.1186/1477-7827-2-26.

46. Wulff, T.; Hoffmann, E. K.; Roepstorff, P.; Jessen, F. Comparison of two anoxia models in rainbow trout cells by a 2-DE and MS/MS-based proteome approach. *Proteomics* **2008**, *8*, 2035–44.

47. Forné, I.; Agulleiro, M. J.; Asensio, E.; Abián, J.; Cerdà, J. 2-D DIGE analysis of Senegalese sole (*Solea senegalensis*) testis proteome in wild-caught and hormone-treated F1 fish. *Proteomics* **2009**, *9*, 2171–81.

48. Reddish, J. M.; St-Pierre, N.; Nichols, A.; Green-Church, K.; Wick, M. Proteomic analysis of proteins associated with body mass and length in yellow perch, *Perca flavescens*. *Proteomics* **2008**, *8*, 2333–43.

49. Zilli, L.; Schiavone, R.; Storelli, C.; Vilella, S. Molecular mechanisms determining sperm motility initiation in two sparids (*Sparus aurata* and *Lithognathus mormyrus*). *Biol. Reprod.* **2008**, *79*, 356–66.

50. Ziv, T.; Gattegno, T.; Chapovetsky, V.; Wolf, H.; Barnea, E.; Lubzens, E.; Admon, A. Comparative proteomics of the developing fish (zebrafish and gilthead seabream) oocytes. *Comp. Biochem. Physiol. D* **2008**, *3*, 12–35.

51. Zilli, L.; Schiavone, R.; Zonno, V.; Rossano, R.; Storelli, C.; Vilella, S. Effect of cryopreservation on sea bass sperm proteins. *Biol. Reprod.* **2005**, *72*, 1262–67.

52. Knoll-Gellida, A.; André, M.; Gattegno, T.; Forgue, J.; Admon, A.; Babin, P. J. Molecular phenotype of zebrafish ovarian follicle by serial analysis of gene expression and proteomic profiling, and comparison with the transcriptomes of other animals. *BMC Genomics* **2006**, *7,* 46, DOI: http://dx.doi.org/10.1186/1471-2164-7-46.

53. Groh, K. J.; Nesatyy, V. J.; Segner, H.; Eggen, R. I. L.; Suter, M. J.-F. Global proteomics analysis of testis and ovary in adult zebrafish (*Danio rerio*). *Fish Physiol. Biochem*. **2011**, *37*, 619–47, DOI: http://dx.doi.org/10.1007/s10695-010-9464-x.

54. Groh, K. J.; Schönenberger, R.; Eggen, R. I. L.; Segner, H.; Suter, M. J.-F. Analysis of protein expression in zebrafish during gonad differentiation by targeted proteomics. *Gen. Comp. Endocr*. **2013**, *193*, 210-20.

55. Martyniuk, C. J.; Alvarez, S. Proteome analysis of the fathead minnow (*Pimephales promelas*) reproductive testes. *J. Proteomics* **2013**, *79*, 28–42, DOI: http://dx.doi.org/10.1016/j.jprot.2012.11.023.

56. Elmore, S. Apoptosis: A Review of Programmed Cell Death. *Toxicol. Pathol.* **2007,** *35*, 495-516.

57. Dumollard, R.; Duchen, M.; Carroll, J. The role of mitochondrial function in the oocyte and embryo. *Curr. Top. Dev. Biol.* **2007,** *77*, 21–49.

58. Wang, G.; Yan, S. Mitochondrial DNA content and mitochondrial gene transcriptional activities in the early development of loach and goldfish. *Int. J. Dev. Biol.* **1992**, *36*, 477–82.

59. Marinos, E.; Billett, F. S. Mitochondrial number, cytochrome oxidase and succinic dehydrogenase activity in *Xenopus laevis* oocytes. *J. of Embryol. Exp. Morph.* **1981**, *62*, 395-409.

60. Marinos, E. The number of mitochondria in *Xenopus laevis* ovulated oocytes. *Cell Differ.* **1985**, *16*, 139–43.

61. Billett, F. S.; Adam, E. The structure of the mitochondrial cloud of *Xenopus laevis* oocytes. *J. of Embryol. and Exp. Moprh.* **1976**, *33*, 697-710.

62. Greenwood, J.; Gautier, J. From oogenesis through gastrulation: developmental regulation of apoptosis. *Semin. Cell Dev. Biol.* **2005**, *16*, 215–24.

63. Kerby, J. H. Striped bass and striped bass hybrids. In *Culture of nonsalmonid freshwater fishes*. Stickney, R. R., Ed; CRC Press: Boca Raton, FL, USA, **1986**; pp 127-47.

64. Barbosa-Morais, N. L.; Irimia, M.; Pan, Q.; Xiong, H. Y.; Gueroussov, S.; Lee, L. J.; Slobodeniuc, V.; Kutter, C.; Watt, S.; Colak, R.; Kim, T.; Misquitta-Ali, C. M.; Wilson, M. D.; Kim, P. M.; Odom, D. T.; Frey, B. J.; Blencowe, B. J. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **2012**, *338*, 1587–93.

65. Buljan, M.; Chalancon, G.; Eustermann, S.; Wagner, G.P.; Fuxreiter, M.; Bateman, A.; Madan Babu, M. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **2012**, *46*, 871–83.

66. Wang, R. L.; Bencic, D.; Biales, A.; Lattier, D.; Kostich, M.; Villenueve, D.; Ankley, G. T.; Lazorchak, J.; Toth, G. DNA Microarray-based ecotoxological biomarker discovery in a small fish model species. *Environ. Toxicol. Chem*. **2008**, *27* (3), 664–75.

67. Wang, R. L.; Bencic, D.; Biales, A.; Flick, R.; Laz orchak, J.; Villeneuve, D.; Ankley, G. T. Discovery and validation of gene classifiers for endocrine-disrupting chemicals in zebrafish (danio rerio). *BMC Genomics* **2012**, *13*, 358, DOI: http://dx.doi.org/10.1186/1471-2164-13-358.

69. Ornostay, A.; Cowie, A M.; Hindle, M.; Baker, C. J. O.; Martyniuk, C. J. Classifying chemical mode of action using gene networks and machine learning: A case study with the herbicide linuron. *Comp. Biochem. Physiol. D* **2013**, 8 (4), 263–74.

70. Sawaguchi, S.; Ohkubo, N.; Amano, H.; Hiramatsu, N.; Hara, A.; Sullivan, C. V.; Matsubara, T. Controlled accumulation of multiple vitellogenins into oocytes during vitellogenesis in the barfin flounder, *Verasper moseri*. *Cybium Int. J. Ichthyol*. **2008**, *32* (2), 262.

71. Williams, V. N.; Reading, B. J.; Amano, H.; Hiramatsu, N.; Schilling, J.; Salger, S. A.; Islam Williams , T.; Gross, K.; Sullivan, C. V. Proportional Accumulation of Yolk Proteins Derived from Multiple Vitellogenins is Precisely Regulated During Vitellogenesis in Striped bass (*Morone saxatilis*) *J. Exp. Zool. Part A* In review, September **2013**.

72. Williams, V. N.; Reading, B. J.; Hiramatsu, N.; Amano, H.; Glassbrook, N.; Hara, A.; Sullivan, C. V. Multiple vitellogenins and product yolk proteins in striped bass, *Morone saxatilis*: molecular characterization and processing during oocyte growth and maturation. *Fish Physiol. Biochem.* **2013,** DOI: http://dx.doi.org/10.1007/s10695-013-9852-0.

73. Hassler, T. J. Species profiles: Life histories and environmental requirements of coastal fishes and invertebrates (Pacific Southwest) -- striped bass. In *U.S. Fish and Wildlife Service Biological Report*. *82* (11.82). U.S. Army Corps of Engineers, TR EL-82-4, **1988**; pp 1-29.

74. Amano, H.; Fujita, T.; Hiramatsu, N.; Kagawa, H.; Matsubara, T.; Sullivan, C. V.; Hara, A. Multiple vitellogenin-derived yolk proteins in gray mullet (*Mugil cephalus*): disparate proteolytic patterns associated with ovarian follicle maturation. *Mol. Reprod. Dev.* **2008**, *75*, 1307–17.

75. Sawaguchi, S.; Ohkubo, N.; Koya, Y.; Matsubara, T. Incorporation and utilization of multiple forms of vitellogenin and their derivative yolk proteins during vitellogenesis and embryonic development in the mosquitofish, *Gambusia affinis*. *Zool. Sci*. **2005**, *22*, 701–10.

76. Kolarevic, J.; Nerland, A.; Nilsen, F.; Finn, R. N. Goldsinny wrasse (*Ctenolabrus rupestris*) is an extreme *vtgAa*-type pelagophil teleost. *Mol. Reprod. Dev.* **2008**, *75*, 1011-20.

77. Wu, C. C.; Yates, J. R. The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* **2003**, *21*, 262–67.

CHAPTER 3

Machine Learning Reveals Sex-Specific 17β-Estradiol-Responsive Expression Patterns in White Perch (*Morone americana*) Plasma Proteins

Justin Schilling[1], Angelito I. Nepomuceno[2], Antonio Planchart[3,4], Jeffrey A. Yoder[4,5,6], Robert M. Kelly[7], David C. Muddiman[2,6], Harry V. Daniels[1], Naoshi Hiramatsu[8], and Benjamin J. Reading[1]

[1]Department of Applied Ecology, North Carolina State University, Raleigh, North Carolina 27695, United States

[2]W. M. Keck FTMS Laboratory for Human Health Research, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States

[3]Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina 27695, United States

[4]Center for Human Health and the Environment, North Carolina State University, Raleigh, North Carolina 27695, United States

[5]Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina 27607, United States

[6]Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina 27607, United States

[7]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695, United States

[8]Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Hokkaido 041-8611, Japan

Correspondence: Benjamin J. Reading, 127 David Clark Laboratories, Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695-7617, USA

Telephone: 919-515-3830

Fax: 919-515-5327

E-mail: bjreadin@ncsu.edu

Abbreviations: EDCs, endocrine disrupting compounds; **E_2**, 17β-estradiol; **EE2**, 17α-ethynylestradiol; **IC**, initial control; **ANNs**, artificial neural networks; **SVMs**, support vector machines; **MMC**, modulated modularity clustering; **AUROC**, area under the receiver operating characteristic curve; **CV**, 10-fold cross-validations; **BFW**, Bonferroni familywise error; **BH**, Benjamini-Hochberg false discovery rate; **ORF**, open reading frame; **GO**, Gene Ontology; **EGF**, epidermal growth factor; **AGC**, automatic gain control; **NCE**, normalized collision energy

Keywords:

endocrine disrupting compounds/machine learning/semiquantitative proteomics/support vector machines/plasma

Total number of words: 8638

In second round of review at *PROTEOMICS* as of March 3, 2015.

Justin Schilling (JS) was the primary author of this manuscript, which was a collaborative effort that also involved Angelito I. Nepomuceno (AIN), Antonio Planchart (AP), Jeffrey A. Yoder (JAY), Robert M. Kelly (RMK), David C. Muddiman (DCM), Harry V. Daniels (HVD), Naoshi Hiramatsu (NH) and Benjamin J. Reading (BJR). AIN assisted JS with all mass spectrometry experiments. JS performed all additional experiments and composed Figures 1, 2, and 3.

Abstract

With growing abundance and awareness of endocrine disrupting compounds (EDCs) in the environment, there is a need for accurate and reliable detection of EDC exposure. Our objective in the present study was to observe differences within and between the global plasma proteomes of sexually mature male and female white perch (*Morone americana*) before (Initial Control, IC) and after 17β-estradiol (E_2) induction. Semiquantitative nanoLC-MS/MS data were analyzed by machine learning support vector machines (SVMs) and by two-way ANOVA. By ANOVA, the expression levels of 44, 77, and 57 proteins varied significantly by gender, treatment, and the interaction of gender and treatment, respectively. SVMs perfectly classified male and female perch IC and E_2-induced plasma samples using the protein expression data. E_2-induced male and female perch plasma proteomes contained significantly higher levels of the yolk precursors vitellogenin Aa and Ab (VtgAa, VtgAb), as well as latrophilin and seven transmembrane domain-containing protein 1 (Eltd1) and kininogen 1 (Kng1). This is the first report that Eltd1 and Kng1 may be E_2-responsive proteins in fishes and therefore may be useful indicators of estrogen induction.

Word count: 178

Introduction

There is growing concern over the increasing prevalence of natural and anthropogenic hormone mimics and antagonists present in waterways around the world. Known as environmental endocrine-disrupting compounds (EDCs), these pollutants can alter gene expression and adversely affect sexual development and reproductive function in a wide range of fishes and other aquatic vertebrate species [1]. EDCs can mimic endogenous hormones and affect normal patterns of hormone synthesis and metabolism as well as hormone receptor expression and function [2, 3]. The endogenous female hormones estradiol, estrone, and estriol play central roles in regulating vertebrate reproductive processes. Of these, 17β-estradiol (E_2) is the most biologically active form of estrogen synthesized naturally in vertebrate ovaries and its synthetic derivative, 17α-ethynylestradiol (EE_2), is the primary form used in oral contraceptives and hormone replacement therapies for humans [4-7]. After their excretion by humans, estrogenic steroids often persist after sewage treatment and are detectable in municipal wastewater effluents at levels that could pose health risks to aquatic vertebrates and to humans as well [8-10].

Aquatic vertebrates are good models for EDC exposure because industrial and municipal effluents that drain into waterways they inhabit often contain EDCs [11]. White perch (*Morone americana*) are native to fresh and brackish waters throughout Northeastern North America and thus are a good model for studying the effects of exposure to EDCs and other environmental contaminants in fishes found in a wide variety of habitats [12-14]. Current strategies for monitoring EDC exposure in fishes rely heavily on recommendations put forward by the Organization for Economic Cooperation and Development (OECD; Paris, France) in 2002 [15, 16]. In the past, such studies/strategies have relied primarily on identification and use of biomarkers of EDC exposure, such as the presence of the female

yolk precursors, vitellogenins, in the blood of immature fishes or sexually mature males exposed to estrogen mimicking EDCs [17-19]. Other female specific plasma proteins, such as choriogenins, also have been successfully used in this regard [20]. These particular biomarker approaches offer simple and effective solutions for the detection of estrogenic EDC exposure in populations of aquatic vertebrates. However, they have limitations. For example, it is difficult to discern the exposure impact of sexually mature female fishes to estrogen mimicking EDCs based on the presence of vitellogenins in the blood as females typically produce these yolk precursors at high levels during ovarian growth [18, 21-23]. Additionally, proper biomarker characterization is required prior to any assay implementation. In the case of fish vitellogenins, this is often complex as there are several forms and they exhibit different sensitivities to estrogen induction [19, 24, 25]. Therefore, proper use and interpretation of vitellogenins as biomarkers of estrogen or estrogenic EDC exposure require thorough characterization of the yolk protein precursors in a particular fish or aquatic species.

Due to recent computational advances in data mining, the power of machine learning has been realized for use in the analysis of transcriptomics and proteomics data. While classical linear statistical analysis methods require a sample size N = n + 1 where n is the number of proteins measured to fully assess the univariates [26], machine learning works by identifying patterns in complex datasets and it may be used to classify biological samples by treatment based on the expression patterns of many hundreds or thousands of genes or proteins. We consider such machine learning evaluations to provide transcriptomic or proteomic “fingerprints”. As examples, artificial neural networks (ANNs) are a type of machine learning that have been used to evaluate environmental perturbations of oysters [26, 27] and to predict egg quality in striped bass [28] using transcriptomics data. Recently, another type of

machine learning called support vector machines (SVMs) has been used to evaluate EDC effects measured using microarray gene expression analysis of zebrafish (*Danio rerio*) [29]. We also have used SVMs to accurately classify the proteomes in different growth stages of striped bass (*Morone saxatilis*) ovaries [30] and to differentiate between white perch ovary cell fractions [22]. Thus, proteomics methods paired with pattern recognition by machine learning potentially offer superior alternatives for evaluation of hormone or EDC responses compared to prior biomarker-based approaches.

Our objective in the present study was to observe differences within and between the global plasma proteomes of mature male and female white perch before and after E_2 induction. As such, we performed three technical replicate nanoLC-MS/MS analyses on each of four groups of biological sample pools: 1) (Initial Control, IC) male plasma, 2) IC female plasma, 3) E_2-induced male plasma, and 4) E_2-induced female plasma. Here, we report that using SVMs we were able to identify: 1) sex-specific differences between the plasma proteomes of mature male and female white perch, 2) differences between the plasma proteomes of control white perch and fish induced with E_2, and 3) sex-specific differences between the plasma proteomes of mature male and female white perch induced with E_2.

Materials and methods

Sample collection and preparation

Three year old adult white perch (body weight 127.63 ± 33.02 g; total length 207.43 ± 17.3 mm; all values listed as average ± standard deviation) were reared in recirculating aquaculture systems under controlled photothermal conditions at the North Carolina State University Aquaculture Facility (Raleigh, NC) [31, 32]. Water temperature was maintained between 10-12 ^{o}C during the course of the study and for 30 days prior to induction with E_2.

All animals were handled according to approved institutional animal care and use committee (IACUC) protocols. Sexually mature female and male fish (N = 3 each) were anesthetized with Finquel MS-222 (Argent Chemical Laboratories, Redmond, WA), and initial control (IC) blood was collected from the caudal vein by heparinized needle and syringe. Blood was transferred into 12 x 75 mm tubes containing 25 µL of aprotinin solution (40 TIU/mL aprotinin in 0.9% NaCl) (all chemicals were purchased from Sigma Aldrich, St. Louis, MO unless otherwise stated). Samples were centrifuged and the plasma was collected and stored at -80 °C until use. Sexually mature male (N = 15) and female (N = 13) white perch were injected with 0.1 mL of 5 mg/mL E_2 dissolved in propylene glycol, per 100 g body weight. Additional E_2 injections were administered at seven and fourteen days after the initial injection based on established protocols [18, 21, 33]. This induction procedure was based on the finding that estrogen administration to male oviparous vertebrates produces lasting changes in vitellogenin expression patterns, termed a cellular “memory effect”, whereby re-stimulation with estrogen after an initial dosage drives far greater expression levels of vitellogenins [34]. Final bleeds were taken 16 days following the initial injection. E_2-induced blood samples were centrifuged and the plasma was collected and stored at -80 ^{o}C until use.

Filter-aided sample preparation and nanoLC-MS/MS

Plasma samples were thawed and diluted with Tris-buffered saline (20 mM Tris-HCL pH 8.0, 150 mM NaCl, and 2 mM $CaCl_2$) to final protein concentrations of 5 mg/mL as determined by Bradford assay [35, 36]. Four biological sample pools were prepared from the plasmas as follows [37]: (1) IC male plasma (N = 3 fish), (2) IC female plasma (N = 3 fish), (3) E_2-induced male plasma (N = 15 fish), and (4) E_2-induced female plasma (N = 13 fish). Each of

these four pooled plasma samples was separately prepared for tandem mass spectrometry using a modified filter-aided sample preparation (FASP) protocol [22, 38]. DTT was used to reduce disulfide bonds by incubating samples with 3 µL of 50 mM DTT per 30 µL (100 µg) of sample. Samples were incubated with DTT for 30 min at 56 °C. Proteins were denatured by adding 200 µL of 8 M urea in 0.1 M Tris-HCl pH 8.5 to each sample. Each sample was then transferred onto a Vivacon 500 30 kDa MW cutoff filter (Sartorius Stedim Biotech, Goettingen, Germany) and centrifuged at 14,000 × *g* for 15 min at 21 °C. To ensure denaturation, an additional 200 µL of 8 M urea in 0.1 M Tris-HCl pH 8.5 was added to each sample before centrifuging again as above. Each sample was on-filter alkylated by adding 100 µL of 50 mM iodoacetamide prepared in 8 M urea atop each filter. Samples were incubated in the dark at room temperature for 20 min and then centrifuged at 14,000 × *g* for 10 min at 21 °C. Each filter was washed three times with 100 µL of 8 M urea by centrifugation for 10 min at 14,000 × *g* followed by three washes with 100 µL of 0.1M Tris pH 7.5. Each sample was then placed in a new centrifuge tube and modified trypsin freshly prepared in 0.1M Tris pH 7.5 was added to each sample at an enzyme to protein ratio of 1:100. Following overnight digestion at 37 °C, tryptic peptides were collected by centrifugation at 14,000 × *g* for 10 min at 21 °C.

All nanoReversed phase chromatography and MS/MS methods were identical to those described previously [22]. Peptides were separated using a Thermo Scientific EASY nLC II system (Thermo Scientific, San Jose, CA) in line with a cHiPLC nanoflex system (AB Sciex, Framingham, MA). A ChromXP C18-CL 3 µm trap column and a ChromXP C18-CL 75 µm × 15 cm analytical column were used for nanoLC peptide separation in a vented column configuration [39] coupled to a quadrupole orbitrap mass spectrometer (Q-Exactive, Thermo Scientific, San Jose, CA) [40]. Each biological sample pool was analyzed in triplicate.

All spectra were collected in profile mode. Mass resolving power of the precursor peptide ions were acquired at $70k_{FWHM}$ at *m*/*z* = 200. The automatic gain control (AGC) target for MS acquisitions was set to 1E6 with a maximum ion injection time of 30 ms. The scan range was set from 400 to 1600 *m*/*z*. Tandem MS spectra were acquired with $17.5k_{FWHM}$ resolving power at *m*/*z* = 200. The AGC target for MS/MS acquisition were set to 2E4 with a maximum injection time of 120 msec. Data dependent acquisition was performed on the top 12 candidates with dynamic exclusion time set to 60 seconds. An underfill ratio of 1.0% was used as a threshold for the candidate peptides selected for MS/MS with an isolation width of 2.0 *m*/*z*. A normalized collision energy (NCE) of 27 was used for fragmentation for all peptides. Microscans were set to 1 msec for both MS and MS/MS. Charge state screening was enabled. Unassigned and 1+ charge states were rejected from MS/MS isolation and activation. The ambient ion used for lock mass was *m*/*z* = 445.12003.

Protein identifications and semi-quantitative spectral counting

Chromatogram files (.RAW) were processed into a peak list format (.MGF) using Proteome Discoverer (Thermo Scientific, San Jose, CA) and batch searched using MASCOT [41] (Matrix Science, Boston, MA) against 21,829 open reading frames (ORFs) predicted from the striped bass genome sequence assembly (Reading et al., *unpublished data*) translated in all six ORFs using AUGUSTUS [42]. Polypeptide sequences for human keratins and porcine trypsin were included in the database. A target-reverse protein database was made to determine false discovery rate. MASCOT data search parameters consisted of carbamidomethyl (C) as a fixed modification and carbamyl (K, N-term), deamidation (NQ), oxidation (M), and phospho (STY) as variable modifications. The MASCOT precursor ion

search tolerance was 5 ppm and the fragment ion tolerance was ±0.02 Da. ProteoIQ was used to filter protein identifications using a 1% false discovery rate [43-45]. Semiquantitative proteomics analysis was carried out using normalized tryptic peptides of each protein (normalized spectral counts). Normalized spectral counts (NSpC) were calculated using the total spectral counts within the replicates of each sample and followed with the maximum spectral counts obtained from each IC and E_2-induced plasma replicate analysis for each of the genders. These NSpC for each of the three technical replicates per biological sample pool were exported from ProteoIQ and transformed to account for zero values [$\log_{10}(y + 1)$, where y = NSpC] [46]. Proteins with ≥ 0.25 average NSpC ($NSpC_{avg}$) across all twelve technical replicates (three technical replicates per treatment) were analyzed by two-way ANOVA [47]. Predicted striped bass proteins corresponding to the white perch peptides identified by MASCOT were analyzed by BLAST at the National Center for Biotechnology Information (NCBI) and annotated.

Western blotting

We raised polyclonal antisera against the three forms of white perch vitellogenin using two unique synthetic peptides from the lipovitellin heavy chain per target protein (GeneTel Laboratories LLC, Madison, WI). In order to provide an orthogonal confirmation of vitellogenin expression, we performed Western blotting of male white perch IC and E_2-induced plasmas using these antibodies [21, 23, 48].

Data analysis

Comparisons of protein expression levels between samples were evaluated by two-way ANOVA ($\alpha = 0.05$) of $\log_{10}(y + 1)$ transformed NSpC values using the general linear model

(GLM) procedure in the SAS/STAT software (version 9.3, SAS Institute Inc., Cary, NC, USA). The two-way ANOVA was used for the comparisons of: 1) treatment (IC versus E_2-induced plasmas), 2) gender (IC and E_2-induced male plasma versus IC and E_2-induced female plasma, and 3) the interaction of treatment and gender (IC male and female by E_2-induced male and female). We employed two independent methods of α correction following two-way ANOVA: 1) standard Bonferroni correction and 2) Benjamini-Hochberg method [22]. ANOVA residual values were input for modulated modularity clustering (MMC) analysis performed using Pearson correlation coefficient [49]. MMC graphs were generated using residual values from 1) the entire dataset (all 4 groups) and for the following specific pair-wise comparisons of groups: 2) IC male and IC female, 3) IC male and E_2-induced male, 4) IC female and E_2-induced female, and 5) E_2-induced male and E_2-induced female.

For each of the MMC graphs, the DAVID Functional Classification Tool [50] was used to identify Gene Ontology enrichment of proteins within all modules. Default parameters for DAVID were used, and approved gene abbreviations for all white perch proteins were manually collected from the NCBI or GeneCards using approved gene names for zebrafish (*Danio rerio*) [51].

We used WEKA (version 3.6.7, http://www.cs. waikato.ac.nz/ml/weka/) to perform the sequential minimal optimization (SMO) support vector machines (SVMs) machine learning classifier algorithm [52]. The SVMs were used to classify the $\log_{10}(y + 1)$ transformed NSpC semiquantitative protein expression data as IC male plasma, E_2-induced male plasma, IC female plasma, or E_2-induced female plasma using the complete perch plasma protein expression dataset and no preprocessing step. Additionally, a *k*-means clustering algorithm was included as a preprocessing step prior to performing SVMs in the analyses of the complete protein expression dataset. Data were clustered in $k = 3$ clusters (where each of

the clusters represented either IC male plasma, IC female plasma, or E_2-induced plasmas of both genders) and $k = 2$ clusters (where each of the clusters represented E_2-induced male plasma or E_2-induced female plasma). We employed two cross-validation strategies for the comparisons to estimate classifier performance as follows: 1) a percentage split whereby 66% of the data were randomly selected and used for training and the remaining 33% of the data were input as a cross-validation and 2) a stratified hold-out with *k*-folds where one fold was used for cross-validation and $k - 1$ folds of the randomly reordered dataset were used for training. A 10-fold stratified hold-out was used for SVMs cross-validation when the samples were classified as IC male plasma, E_2-induced male plasma, IC female plasma, or E_2-induced female plasma using the complete protein expression dataset and when the *k*-means algorithm clustered the dataset into 3 clusters (IC male plasma, IC female plasma, or E_2-induced plasmas of both genders). A *k*-fold stratified hold-out cross-validation was not performed for comparison of E_2-induced male and E_2-induced female plasmas that were clustered into only 2 groups by *k*-means preprocessing. In all cross-validations, the SVMs were naive to the hold-out data and percent correct assignment was used to evaluate classifier performance.

All data classes were properly represented in the machine learning training and cross-validation datasets. The performance of the SVMs was evaluated as a percent of correct classification during the cross-validations. Additionally, the Kappa statistic and area under the receiver operating characteristic curve (AUROC) of each model also were reported. As a negative control of machine learning, the NSpC values for the whole dataset were randomly reordered 10-times and each was entered into SVMs as described above, with or without *k*-means preprocessing, as appropriate.

Results

Tandem mass spectrometry

A total of 104 unique proteins were detected by tandem mass spectrometry in the white perch plasma proteomes (Supplemental Table 1). A list of the complete nucleotide sequences for open reading frames encoding these proteins, and found in the striped bass genome sequence are included as Supplemental Table 2. Of these proteins, 94 were detected with $NSpC_{avg} \geq 0.25$. All 58 proteins detected in IC male plasma were also detected in IC female plasma, however, 39 proteins were found in IC female plasma but not IC male plasma (Figure 1 A). A comparison of proteins detected in IC male and E_2-induced male plasmas revealed 26 proteins exclusive to IC male plasma, 15 exclusive to E_2-induced male plasma, and 35 found in both (Figure 1 B). Similarly, comparing proteins detected in IC female and E_2-induced female plasmas detected 48 proteins exclusive to IC female plasma, 6 exclusive to E_2-induced female plasma, and 49 found in both (Figure 1 C). Lastly, comparing proteins detected in E_2-induced male and E_2-induced female plasmas detected 2 proteins exclusive to IC male plasma, 7 exclusive to E_2-induced female plasma, and 48 found in both (Figure 1 D).

Western blotting

Western blotting confirmed that none of the three forms of vitellogenin were expressed in IC male white perch plasma (Figure 2). After induction with E_2, however, expression of all three forms of white perch vitellogenin (VtgAa, VtgAb, and VtgC) was evident by Western blotting (Figure 2).

Inferential statistics

Three comparisons of white perch plasma protein expression were made by two-way ANOVA using proteins with $NSpC_{avg} \geq 0.25$: 1) Treatment (IC male and female versus E_2-induced male and female), 2) Gender (IC and E_2-induced male versus IC and E_2-induced female), and 3) Treatment by Gender (IC male and female by E_2-induced male and female) (Table 2). When protein expression levels of IC male and female versus E_2-induced male and female plasma were compared by two-way ANOVA after a correction, 57 of the 94 (60.6%, $p \leq 0.000531$, Bonferroni) and 77 of the 94 (81.9%, $p \leq 0.034$, Benjamini-Hochberg) proteins detected were found to vary significantly by gender. When protein expression levels between IC and E_2-induced male versus IC and E_2-induced female plasma were compared as above, 26 of the 68 (38.2%, $p \leq 0.$ 0.000735, Bonferroni) and 44 of the 68 (64.7%, $p \leq 0.029$, Benjamini-Hochberg) proteins detected were found to vary significantly by gender. When protein expression levels between IC male and female by E_2-induced male and female plasma were compared as above, 32 of the 93 (34.4%, $p \leq 0.0003$, Bonferroni) and 57 of the 93 (61.2%, $p \leq 0.0263$, Benjamini-Hochberg) proteins detected were found to vary significantly by treatment and gender. Differentially expressed proteins are listed in Supplemental Materials (Supplemental Table 3).

MMC and DAVID analysis

The entire white perch plasma protein expression dataset also was analyzed by two-way ANOVA and the resulting residual values for each protein were input for MMC. A table of these residual values is included in Supplemental Materials (Supplemental Table 4). ANOVA residual values for predicted striped bass proteins corresponding to white perch plasma proteins detected by mass spectrometry were organized on the basis of covariable

expression using MMC. MMC graphs were generated representing proteins from the entire mass spectrometry dataset (Figure 3, A), IC male and IC female (Figure 3, B), IC male and E_2-induced male (Figure 3, C), IC female and E_2-induced female (Figure 3, D), and E_2-induced male and E_2-induced female (Figure 3, E). The output from each MMC analysis is included in Supplemental Materials (Supplemental Table 5).

When all residual values from two-way ANOVA were input into MMC, a total of 15 modules were generated (Figure 3, A). Analysis by DAVID revealed three proteins (F2, Tfa, Eno2) in module 14 with significant GO enrichment for metal ion binding. The MMC graph of two-way ANOVA residual values from IC male and IC female contains 27 modules, the largest of which (module 15) contains 11 proteins (Figure 3, B). Module 15 was significantly enriched for GO terms associated with glycolysis (Aldoaa, Eno2, Gapdhs). The MMC graph of IC male and E_2 male residuals contains 6 modules, the largest of which (module 2) contains 23 proteins, four of which (Cp, Eno2, Rock2b, zgc:163025) are involved in metal ion binding (Figure 2, C). Module 5 was significantly enriched for GO terms associated with glycolysis (Aldoaa, Eno2, Tpi1b) (Figure 3, C). MMC generated 8 modules when residuals from IC female and E_2 female were input (Figure 3, D). Modules 2 and 3 are the largest, each containing 16 proteins, yet neither module shows significant GO-term enrichment. Modules 5 and 7, however, show significant GO-term enrichment for proteins related to SERPIN (serine protease inhibitor) activity (Agt, Serpina1, Serpind1) and blood coagulation (F2, Fgg, Plg), respectively. When E_2 male and E_2 female residuals were input into MMC, a total of 10 modules were generated, one of which (module 9) shows significant GO-term enrichment for proteins related to extracellular space, signal peptide and secretion (Apoa2, F2, Serpina1) (Figure 3, E).

Supplemental Figure 1 depicts glycolysis/gluconeogenesis KEGG Pathway analysis of the entire white perch plasma protein dataset as measured by mass spectrometry. Detected white perch plasma proteins are indicated by a red box.

Machine learning

When $\log_{10}(y + 1)$ transformed NSpC values from the complete white perch plasma proteome dataset were input directly into SVMs with no preprocessing step, all four groups (IC male, E_2-induced male, IC female, and E_2-induced female) were classified with perfect accuracy (100% correct classifier assignment) using 10-fold or 66%-split cross-validations (Table 3). The Kappa statistics and AUROCs were both optimal at 1.00. When the same dataset was preprocessed by *k*-means clustering, the algorithm was unable to correctly group the entire dataset into 4 clusters. Instead, the *k*-means algorithm clustered the entire white perch plasma protein dataset into 3 clusters that included IC male plasma, IC female plasma, or E_2-induced plasmas of both sexes. This indicates that the *k*-means algorithm was not sensitive enough to differentiate the between the E_2-induced sex-specific responses in protein expression. Following *k*-means clustering, the SVMs were again able to classify all data with perfect accuracy using both cross-validation strategies (Table 3). When only the E_2-induced male and E_2-induced female protein expression data were preprocessed by *k*-means into 2 clusters and then fed into SVMs, the two groups were classified with perfect accuracy using the 66%-split cross-validation (100% correct assignment) (Table 3). With just 2 clusters, the 10-fold cross validation was not possible for this comparison.

Ten randomized datasets were generated for each of the IC and E_2-induced comparisons to serve as negative controls for machine learning and the resulting SVM classifier performance is reported in Table 4. The correct classification by SVMs of the randomized

datasets ranged from 5.0 to 40.0%, which was slightly less than that predicted from random assignment into the respective 4, (25.0%), 3 (33.3%), and 2 groups (50.0%). Kappa statistics ranged from -0.156 to 0.022 and AUROC ranged from 0.210 to 0.707. A Kappa statistic of less than 0.200 is considered poor performance[53]. An AUROC of 0.500 is comparable to random assignment [52]. A negative Kappa statistic and/or an AUROC of less than 0.500 indicate that machine learning classifier performance is worse than that predicted by random assignment.

Discussion

The SVMs were able to distinguish the plasma proteomes of IC white perch from that of E_2-induced fish and this approach should be amenable to assessing the effects of E_2 on white perch and possibly other aquatic vertebrates. These findings are similar to a previous report where SVMs were used to classify microarray gene expression data from zebrafish exposed to different types of EDCs [54]. However, these authors report inconsistent classifier performance, which they attribute to varying data complexity related to small biological sample size. These authors state that the classifiers used less than 100 input attributes (i.e., expressed genes). Here, the SVMs classifier performance was 100% correct when the expression of 104 unique proteins was considered, further supporting the contention that a considerable number of inputs are required for optimal classification. It also has been shown that the cross-validation performance of ANNs when classifying microarray gene expression data decreases dramatically when the input attributes are restricted from 250 to 100 genes [28] as sample size and attribute number are both important considerations for successful machine learning performance.

In the future, similar studies could be conducted whereby white perch or other sentinel species of fishes are exposed to a panel of different estrogenic EDCs and/or mixtures tested to see whether machine learning can classify the treatments or exposures. For instance, different doses of EDCs may elicit different plasma proteome responses and combinations of EDCs may have synergistic effects on the responses of plasma proteomes. These findings set the stage for using machine learning to classify exposure of aquatic animals to complex mixtures of EDCs. Since wastewater, for instance, may contain a mixture of different EDCs, it becomes difficult to implement the practical identification of particular EDC(s) to which an organism has been exposed. Machine learning may offer a possible solution to this problem as pattern recognition could be used to identify the particular gene or protein expression signatures diagnostic of exposure to one or more particular EDCs. Further research will be required to validate this possibility. While we only evaluated the effects of E_2 on white perch, importantly SVMs did detect subtle sex-specific responses to E_2 in the plasma proteomes, which to our knowledge has not been demonstrated at this scale in any other fish. Such resolution of machine learning suggests that it may be possible to decipher gene or protein expression differences between organisms treated with different EDCs or different concentrations of the same EDC in future studies.

SVMs classification of sexually mature male and female white perch plasma proteomes proves that these methods may be employed to identify the gender of fishes from blood and/or plasma samples using a semiquantitative proteomics approach. This also opens the possibility that these or similar methods may be used for gender identification of sexually immature fish as well. Hidden Markov models have been used to identify age- and gender-related differences in plasma metabolomics data of human children [55]. In practice, gender identification of sexually immature fish usually requires lethal sampling and gonad histology.

Therefore, the machine learning and proteomics procedures described here would be beneficial for the study of stock structures of imperiled fishes, where lethal sampling is not a desirable approach.

Blood plasma contains an extremely large number of proteins across an even larger dynamic range of abundances. As such, plasma has been regarded as an especially difficult sample for proteomics analysis [56] and here we detected just 94 distinct proteins. While the white perch blood samples used in this study were centrifuged as described in order to isolate the plasma and later subjected to FASP, no additional fractionation methods were utilized prior to analysis by mass spectrometry. Following E_2 induction, the increase in abundance of a few proteins, particularly VtgAb, likely impaired detection of less abundant proteins and thus limited the depth of protein detection. In spite of this, it is notable that SVMs were still able to perfectly distinguish between male and female plasma samples, both before and after E_2 induction. The spike in abundance of a few predominant plasma proteins following E_2 induction likely affected the outcomes of the plasma protein comparisons as analyzed by ANOVA. Specifically, many of the proteins that were detected at low levels in IC male and IC female plasma were not detected following E_2 induction and were found to vary significantly by two-way ANOVA following Bonferroni and/or Benjamini-Hochberg correction (Supplemental Tables 1 and 2).

When the entire white perch plasma protein dataset is examined, many proteins directly involved in or closely associated with glycolysis are generally apparent (e.g., Aldoaa, Eno1a, Eno2, Eno3, Gapdh, Gapdhs, Pgam2, Tpi1b). Consistent with the findings in the present study, estrogens have been shown to have strong anabolic effects in vertebrates [57]. A key enzyme in muscle glycogenolysis, Pygma (Contig440-0.0) expression levels varied significantly between IC and E_2-induced perch plasma perch plasma (p = 0.034).

Additionally, other factors generally representative of metabolic state were apparent in the plasma proteomes. Expression levels of Creatine kinase, muscle b (Ckmb, Contig15929-0.2), a metabolic enzyme that plays a key role in energy homeostasis [58], varied significantly between IC and E_2-induced plasma by two-way ANOVA ($p = 0.0002$) and by the interaction of treatment and gender ($p = 0.0002$) (Supplemental Table 2).

Expression levels of Eltd1 (Contig838-0.0), an epidermal growth factor (EGF) seven-transmembrane receptor, varied significantly between male and female plasma ($p < 0.0001$) and by the interaction of treatment and gender ($p = 0.0026$) (Supplemental Table 3). We believe this study to be the first to suggest that Eltd1 may be an E_2-responsive protein in a fish. Interestingly, the *Eltd1* locus is adjacent to one of the vitellogenin gene loci (*vtg1*) in the genome sequences of platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), and teleost fishes [59]. The biological relevance of this association is unclear at this time, however vitellogenins are the predominant egg yolk precursors in oviparous vertebrates and their hepatic synthesis is primarily responsive to E_2 [59-66]. While males possess vitellogenin genes, expression will occur only if induced by an estrogenic compound. Accordingly, the two main forms of vitellogenin in white perch, VtgAa and VtgAb, were not detected in IC male plasma by MS/MS (Supplemental Table 1). While the third form and incomplete type vitellogenin, VtgC, was detected by Western blotting in male plasma following E_2 induction (Figure 2), VtgC was not detected in either male or female plasma by nanoLC-MS/MS as it is known to be expressed at significantly lower levels than VtgAa and VtgAb in white perch [22]. Following induction with E_2, however, VtgAa and VtgAb levels increased markedly in male plasma ($p < 0.001$) (Supplemental Table 1) and VtgAb was the most abundant protein detected following induction with E_2 in both male and female plasmas (Supplemental Table 1). Two putative VtgAb proteins were detected using the AUGUSTUS

predicted striped bass ORFs as a reference database (Contig26928-0.0 and Contig26928-0.1) in IC female plasma and E_2-induced male and female plasma (Supplemental Tables 1 and 2). While multiple *Vtg* genes have been found in *Danio rerio* [66], subsequent curation of the striped bass genome consolidated both of the VtgAb contigs above to the same ORF (Reading et al., *unpublished data*). Detection of VtgAb in the plasma of IC female perch is consistent with the stage of their reproductive cycle at the time of their sampling (during vitellogenesis) based on photothermal conditioning [32].

Members of three proteinase inhibitor families were detected in perch plasma by MS/MS: 1) Serpina1 and Serpind1 (Serpina1, Serpind1, serpin family); 2) Fetuin B (Fetub, fetuin family), and 3) Kininogen 1 (Kng1, Type 3 cystatin family). Serpins are primarily found in blood plasma where they help to maintain homeostasis by regulating proteolytic enzymes[64]. Two putative Serpina1 proteins and one putative Serpind1 protein were detected by tandem mass spectrometry in white perch plasma of both genders (Contig12216-0.0, Contig1392-0.0, and Contig10856-0.0, respectively; see Supplemental Table 1). Expression of the two detected Serpina1 proteins varied significantly between IC plasma and E_2-induced plasma ($p < 0.0001$ for Contig1392-0.0 and Contig12216-0.0) (Supplemental Table 3). Expression levels of the Serpind1 protein varied significantly between IC plasma and E_2-induced plasma ($p < 0.0001$); however, Serpina1 (Contig1392-0.0) and Serpind1 were only detected in IC male and IC female plasmas (Supplemental Table 1). Serpina1 (Contig12216-0.0) was detected in all plasma samples, however, as with both putative Serpina1 proteins, the number of peptides detected decreased in male and female plasmas following E_2-induction. In addition to bovine uterine SERPINA14, a number of serpins have been shown to be estrogen-responsive in the laying hen (*Gallus gallus*),

specifically SERPINB3, SERPINB11, and SERPINB12, which play a key role in the secretion and/or deposition of egg white proteins [67-70].

Fetuin-B (Fetub) is liver-derived cysteine protease inhibitor with roles in a diverse range of processes including fatty acid metabolism, osteogenesis and bone resorption, regulation of the insulin and hepatocyte growth factor receptors, response to systemic inflammation, and prevention of premature hardening of the zona pellucida surrounding mammalian oocytes [71]. Two putative forms of Fetub (Contig13043-0.0 and Contig13357-0.1) were detected. Contig13357-0.1 was not detected in IC male plasma but was detected in E_2-induced male plasma as well as in both IC and E_2-induced female plasma and its protein levels decreased after E_2 induction. When analyzed by two-way ANOVA for the interaction between treatment and gender, Contig13357-0.1 varied significantly ($p = 0.0017$) (Supplemental Table 3). It has been previously shown that E_2 induction in tilapia (*Oreochromis mossambicus*) suppresses the growth hormone-insulin-like growth factor 1 axis [24] and therefore it would be interesting to further investigate the putative role of Fetub as a potential E_2-induced regulator of growth factors in this process.

Kininogen 1 (Kng1) is a cysteine proteinase inhibitor that displays a wide range of functions including blood coagulation and bradykinin transport [72]. In this study, Kng1 (Contig9596-0.0) was detected at significantly higher levels in both male ($p = 0.0013$) and female plasma following E_2 induction ($p < 0.0001$) compared to IC male and IC female plasma, respectively (Supplemental Tables 1 and 3). While kininogens have previously been isolated and characterized from a number of fish species[72-76] and shown to be estrogen responsive in the rat (*Rattus rattus*) [77], we believe this study to be the first to suggest that Kng1 may be an E_2-responsive protein in a fish.

Concluding remarks

Following induction with E_2, both male and female white perch plasma proteomes contained significantly higher levels of VtgAa, VtgAb, Eltd1, and Kng1 than IC plasmas. It appears that E_2 induction generally resulted in an increase in proteins directly or indirectly involved in glycolysis in the plasmas of both male and female perch. SVMs were able to perfectly classify male and female plasma samples as either IC or E_2-induced based on protein expression data as detected by mass spectrometry. Additional studies of the effects of EDCs on vertebrates may benefit from employing SVMs and/or other forms of machine learning in data analysis as these methods appear robust.

Author contributions

J. S. was primarily responsible for the design and execution of the experiments described herein as well as data analysis and preparation of this manuscript. A. I. N. provided assistance with sample preparation for and performed tandem mass spectrometry. A. P. assisted with experimental design and data analysis. J. A. Y. assisted with experimental design and data analysis. R. M. K. helped to conceive of the experimental design. D. C. M. assisted with proteomics measurements and the preparation of this manuscript. H. V. D. assisted with the preparation of this manuscript. N. H. provided assistance with writing of this manuscript. B. J. R. provided assistance with interpretation and analysis of data as well as preparation of this manuscript.

Funding sources

This research was funded in part by the North Carolina Agricultural Research Service Daniels Project number 06819.

Acknowledgements

We thank Brad Ring and John Davis (North Carolina State University Aquaculture Facility) for care and maintenance of the white perch.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] Tyler, C. R., Jobling, S., Sumpter, J. P., Endocrine disruption in wildlife: a critical review of the evidence. *Crit. Rev. Toxicol.* 1998, *28*, 319–361.

[2] Vos, J. G., Dybing, E., Greim, H. A., Ladefoged, O., Lambré, C. et al., Health Effects of Endocrine-Disrupting Chemicals on Wildlife, with Special Reference to the European Situation. *Crit. Rev. Toxicol.* 2000, *30*, 71–133.

[3] Schoeters, G., Hond, Den, E., Dhooge, W., van Larebeke, N., Leijs, M., Endocrine disruptors and abnormalities of pubertal development. *Basic Clin. Pharmacol. Toxicol.* 2008, *102*, 168–175.

[4] Kuster, M., de Alda, M. J. L., Barceló, D., Analysis and distribution of estrogens and progestogens in sewage sludge, soils and sediments. *TrAC.* 2004, *23,* 790-798.

[5] Williams, R. J., Johnson, A. C., Smith, J. J. L., Kanda, R., Steroid estrogens profiles along river stretches arising from sewage treatment works discharges. *Environ. Sci. Technol.* 2003, *37*, 1744–1750.

[6] Pojana, G., Bonfà, A., Busetti, F., Collarin, A., Marcomini, A., Estrogenic potential of the Venice, Italy, lagoon waters. *Environ. Toxicol. Chem.* 2004, *23*, 1874–1880.

[7] Muller, M., Rabenoelina, F., Balaguer, P., Patureau, D., Lemenach, K. et al., Chemical and biological analysis of endocrine-disrupting hormones and estrogenic activity in an advanced sewage treatment plant. *Environ. Toxicol. Chem.* 2008, *27*, 1649–1658.

[8] Manickum, T., John, W., Occurrence, fate and environmental risk assessment of endocrine disrupting compounds at the wastewater treatment works in Pietermaritzburg (South Africa). *Sci. Total Environ.* 2014, *468-469*, 584–597.

[9] Esteban, S., Gorga, M., Petrovic, M., González-Alonso, S., Barceló, D. et al., Analysis and occurrence of endocrine-disrupting compounds and estrogenic activity in the surface waters of Central Spain. *Sci. Total Environ.* 2014, *466-467*, 939–951.

[10] Luo, Y., Guo, W., Ngo, H. H., Nghiem, L. D., Hai, F. I. et al., A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci. Total Environ.* 2014, *473-474*, 619–641.

[11] Ying, G.-G., Kookana, R. S., Ru, Y.-J., Occurrence and fate of hormone steroids in the environment. *Environ. Int.* 2002, *28*, 545–551.

[12] Barnthouse, L. W., Glaser, D., DeSantis, L., Polychlorinated Biphenyls and Hudson River White Perch: Implications for Population-Level Ecological Risk Assessment and Risk Management. *Integr. Environ. Assess. Manag.* 2009, *5*, 435-444.

[13] Kamman, N. C., Burgess, N. M., Driscoll, C. T., Simonin, H. A., Mercury in Freshwater Fish of Northeast North America – A Geographic Perspective Based on Fish Tissue Monitoring Databases. *Ecotoxicology* 2005, *14*, 163–180.

[14] Kavanagh, R. J., Balch, G. C., Liparissis, Y., Niimi, Sherry, J. et al., Endocrine disruption and altered gonadal development in white perch (*Morone americana*) from the lower Great Lakes region. *Environ. Health Persp.* 2004, *112*, 898-902.

[15] Hecker, M., Hollert, H., Cooper, R., Vinggaard, A. M., Akahori, Y. et al., The OECD validation program of the H295R steroidogenesis assay: Phase 3. Final inter-laboratory validation study. *Environ. Sci. Pollut. Res. Int.* 2011, *18*, 503–515.

[16] Gelbke, H. P., Kayser, M., Poole, A., OECD test strategies and methods for endocrine disruptors. *Toxicology* 2004, *205*, 17–25.

[17] Kishida, M., Specker, J. L., Vitellogenin in the surface mucus of tilapia (*Oreochromis mossambicus*): Possibility for uptake by the free-swimming embryos. *J. Exp. Zool.* 1994, *268*, 259–268.

[18] Reading, B. J., Hiramatsu, N., Sullivan, C. V., Disparate binding of three types of vitellogenin to multiple forms of vitellogenin receptor in white perch. *Biol. Reprod.* 2011, *84*, 392–399.

[19] Hiramatsu, N., Matsubara, T., Fujita, T., Sullivan, C. V., Hara, A., Multiple piscine vitellogenins: biomarkers of fish exposure to estrogenic endocrine disruptors in aquatic environments. *Mar. Biol.* 2006, *149*, 35–47.

[20] Hong, L., Fujita, T., Wada, T., Amano, Hiramatsu, N. et al., Choriogenin and vitellogenin in red lip mullet (*Chelon haematocheilus*): purification, characterization, and evaluation as potential biomarkers for detecting estrogenic activity. *Comp. Biochem. Phys. C* 2009, *149*, 9–17.

[21] Reading, B. J., Hiramatsu, N., Sawaguchi, S., Matsubara, T., Hara, A. et al., Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. *Mar. Biotechnol.* 2009, *11*, 169–187.

[22] Schilling, J., Nepomuceno, A., Schaff, J. E., Muddiman, Daniels, H. V. et al., Compartment Proteomics Analysis of White Perch (*Morone americana*) Ovary Using Support Vector Machines. *J. Proteome Res.* 2014, *13*, 1515-1526.

[23] Williams, V. N., Reading, B. J., Amano, H., Hiramatsu, N., Schilling, J. et al., Proportional accumulation of yolk proteins derived from multiple vitellogenins is precisely regulated during vitellogenesis in striped bass (*Morone saxatilis*). *J. Exp. Zool.* 2014, *321*, 301–315.

[24] Davis, L. K., Hiramatsu, N., Hiramatsu, K., Reading, B. J., Matsubara, T. et al., Induction of three vitellogenins by 17beta-estradiol with concurrent inhibition of the growth hormone-insulin-like growth factor 1 axis in a euryhaline teleost, the tilapia (*Oreochromis mossambicus*). *Biol. Reprod.* 2007, *77*, 614–625.

[25] Tyler, C. R., van Aerle, R., Hutchinson, T. H., An *in vivo* testing system for endocrine disruptors in fish early life stages using induction of vitellogenin. *Environ. Toxicol. Chem.* 1999, *18*, 337–347.

[26] Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C. et al., A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica*, in the South Atlantic bight. *Mol. Ecol.* 2009, *18*, 2415–2425.

[27] Chapman, R. W., Mancia, A., Beal, M., Veloso, Rathburn, C. et al., The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Mol. Ecol.* 2011, *20*, 1431–1449.

[28] Chapman, R. W., Reading, B. J., Sullivan, C. V., Ovary transcriptome profiling via artificial intelligence reveals a transcriptomic fingerprint predicting egg quality in striped bass, *Morone saxatilis*. *PLoS ONE* 2014, *9*, e96818.

[29] Wang, R.-L., Bencic, D., Biales, A., Flick, R., Lazorchak, J. et al., Discovery and validation of gene classifiers for endocrine-disrupting chemicals in zebrafish (*danio rerio*). *BMC Genomics* 2012, *13*, 358.

[30] Reading, B. J., Williams, V. N., Chapman, R. W., Williams, T. I., Sullivan, C. V., Dynamics of the Striped Bass (*Morone saxatilis*) Ovary Proteome Reveal a Complex Network of the Translasome. *J. Proteome Res.* 2013, *12*, 1691−1699.

[31] King, W., V, Berlinsky, D., Sullivan, C. V., Involvement of gonadal steroids in final oocyte maturation of white perch (*Morone americana*) and white bass (*M. chrysops*): in vivo and in vitro studies. *Fish Physiol. Biochem.* 1995, *14*, 489–500.

[32] Jackson, L. F., Sullivan, C. V., Reproduction of White Perch: The Annual Gametogenic Cycle. *T. Am. Fish. Soc.* 1995, *124*, 563–577.

[33] Heppell, S. A., Jackson, L. F., Weber, G. M., Sullivan, C. V., Enzyme-linked immunosorbent assay (ELISA) of vitellogenin in temperate basses (genus *Morone*): plasma and in vitro analyses. *T. Am. Fish. Soc.* 1999, *128*, 532–541.

[34] Baker, H. J. and Shapiro, D. J., Rapid Accumulation of Vitellogenin Messenger RNA during Secondary Estrogen Stimulation of *Xenopus laevis*. *J Biol. Chem*. 1978. *253*, 4521-4524.

[35] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976, *72*, 248–254.

[36] Noble, J. E., Bailey, M. J. A., Quantitation of Protein. In *Methods in Enzymology*, Burgess, R. R., Deutscher, M. P., Eds. Elsevier 2009, pp 73–95.

[37] Karp, N. A. and Lilley, K. S., Design and Analysis Issues in Quantitative Proteomics Studies. *Proteomics* 2007, *7*, 42-50.

[38] Wiśniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat. Meth.* 2009, *6*, 359–362.

[39] Andrews, G. L., Shuford, C. M., Burnett, J. C., Hawkridge, A. M., Muddiman, D. C., Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chromatogr. B.* 2009, *877*, 948–954.

[40] Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A. et al., Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 2011, *10*, M111.011015.

[41] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

[42] Stanke, M., Morgenstern, B., AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.*, 2005, *33*, W465–W467.

[43] Weatherly, D. B., Atwood, J. A., Minning, T. A., Cavola, C., Tarleton, R. L. et al., A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results. *Mol. Cell. Proteomics* 2005, *4*, 762-772.

[44] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[45] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, *74*, 5383–5392.

[46] Wu, X., Tseng, C. W., Edwards, N., HMMatch: peptide identification by spectral matching of tandem mass spectra using hidden Markov models. *J. Comput. Biol.* 2007, *14*, 1025-1043.

[47] Gokce, E., Shuford, C. M., Franck, W. L., Dean, R. A., Muddiman, D. C., Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation during Proteomic Workflows. *J. Am. Soc. Mass Spectrom.* 2011, *22*, 2199–2208.

[48] Williams, V. N., Reading, B. J., Hiramatsu, N., Amano, H., Glassbrook, et al., Multiple vitellogenins and product yolk proteins in striped bass, *Morone saxatilis*: molecular characterization and processing during oocyte growth and maturation. *Fish Physiol. Biochem.* 2014, *40*, 395–415.

[49] Stone, E. A., Ayroles, J. F., Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet.* 2009, *5*, e1000479.

[50] Dennis, G., Jr, Sherman, B. T., Hosack, D. A., Yang, J., DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003, *4*, R60.1-R60.11.

[51] Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D., GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 1998, *14*, 656–664.

[52] Witten, I. H., Frank, E., Hall, M. A., *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier: Boston, 2011.

[53] McGinn, T., Wyer, P. C., Newman, T. B., Keitz, S., Leipzig, R. et al., Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *Can. Med. Soc. J.* 2004, *171*, 1369-1373.

[54] Nikkilä, J., Sysi-Aho, M., Ermolov, A., Seppänen-Laakso, T., et al., Gender-dependent progression of systemic metabolic states in early childhood. *Mol. Syst. Biol.* 2008, *4*, 197.

[55] Anderson, N. L., The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* 2002, *1*, 845–867.

[56] Lange, I. G., Hartel, A., and H. H. D. Meyer. Evolution of oestrogen functions in vertebrates. *J. Steroid biochem.* 2002, *83*, 219-226.

[57] Wallimann, T., Wyss, M., Brdiczka, D., Nicolay, K., Eppenberger, H. M., Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: the 'phosphocreatine circuit' for cellular energy homeostasis. *Biochem. J.* 1992, *281*, 21-40.

[58] Babin, P. J., Conservation of a vitellogenin gene cluster in oviparous vertebrates and identification of its traces in the platypus genome. *Gene* 2008, *413*, 76–82.

[59] Tata, J. R., The expression of the vitellogenin gene. *Cell* 1976.

[60] Sumpter, J. P., Jobling, S., Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environ. Health Persp.* 1995, *103*, 173-178.

[61] Denslow, N. D., Chow, M. C., Kroll, K. J., Green, L., Vitellogenin as a Biomarker of Exposure for Estrogen or Estrogen Mimics. *Ecotoxicology* 1999, *8*, 385–398.

[62] Davis, L. K., Hiramatsu, N., Hiramatsu, K., Reading, B. J., Matsubara, T. et al., Induction of Three Vitellogenins by 17beta-Estradiol with Concurrent Inhibition of the Growth Hormone-Insulin-Like Growth Factor 1 Axis in a Euryhaline Teleost, the Tilapia (*Oreochromis mossambicus*). 2007, *77*, 614-625.

[63] Hiramatsu, N., Matsubara, T., Fujita, T., Sullivan, C. V., Multiple piscine vitellogenins: biomarkers of fish exposure to estrogenic endocrine disruptors in aquatic environments. *Mar. Biol.* 2006, *149*, 35-47.

[64] Rose, J., Holbech, H., Lindholst, C., Nørum, U., Povlsen, A., et al., Vitellogenin induction by 17β-estradiol and 17α-ethinylestradiol in male zebrafish (*Danio rerio*). *Comp. Biochem. Phys. C* 2002, *131*, 531–539.

[65] Wang, H., Tan, J. T. T., Emelyanov, A., Korzh, V., Gong, Z., Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene* 2005, *356*, 91–100.

[66] Potempa, J., Korzus, E., Travis, J., The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* 1994, *269*, 15957-15960.

[67] Jo, G., Lim, W., Bae, S.-M., Bazer, F. W., Song, G., Avian SERPINB12 expression in the avian oviduct is regulated by estrogen and up-regulated in epithelial cell-derived ovarian carcinomas of laying hens. *PLoS ONE* 2014, *9*, e99792.

[68] Lim, W., Ahn, S. E., Jeong, W., Kim, J.-H., Kim, J., et al., Tissue specific expression and estrogen regulation of SERPINB3 in the chicken oviduct. *Gen. Comp. Endocrinol.* 2012, *175*, 65–73.

[69] Ulbrich, S. E., Frohlich, T., Schulke, K., Englberger, E., Waldschmitt , N. et al., Evidence for Estrogen-Dependent Uterine Serpin (SERPINA14) Expression During Estrus in the Bovine Endometrial Glandular Epithelium and Lumen. *Biol. Reprod.* 2009, *81*, 795–805.

[70] Dietzel, E., Wessling, J., Floehr, J., Schäfer, C., Ensslen, S. et al., Fetuin-B, a liver-derived plasma protein is essential for fertilization. *Dev. Cell* 2013, *25*, 106–112.

[71] Ylönen, A., Rinne, A., Herttuainen, J., Bøgwald, J., Jarvinen, M. et al., Atlantic salmon (*Salmo salar* L.) skin contains a novel kininogen and another cysteine proteinase inhibitor. *Eur. J. Biochem.* 1999, *266*, 1066–1072.

[72] Ylönen, A., Kalkkinen, N., Saarinen, J., Bøgwald, J., Helin, J. Glycosylation analysis of two cysteine proteinase inhibitors from Atlantic salmon skin: di-O-acetylated sialic acids are the major sialic acid species on N-glycans. *Glycobiology* 2001, *11*, 523–531.

[73] Ylönen, A., Helin, J., Bøgwald, J., Jaakola, A., Rinne, A. et al., Purification and characterization of novel kininogens from spotted wolffish and Atlantic cod. *Eur. J. Biochem.* 2002, *269*, 2639–2646.

[74] Zhou, L., Li-Ling, J., Huang, H., Ma, F., Li, Q. Phylogenetic analysis of vertebrate kininogen genes. *Genomics* 2008, *91*, 129–141.

[75] Zhou, L., Liu, X., Jin, P., Li, Q. Cloning of the kininogen gene from *Lampetra japonica* provides insights into its phylogeny in vertebrates. *J. Genet. Genomics* 2009, *36*, 109–115.

[76] Chen, L. M., Chung, P., Chao, S., Chao, L., Chao, J. Differential regulation of kininogen gene expression by estrogen and progesterone *in vivo*. *Biochim. Biophys. Acta* 1992, *1131*, 145–151.

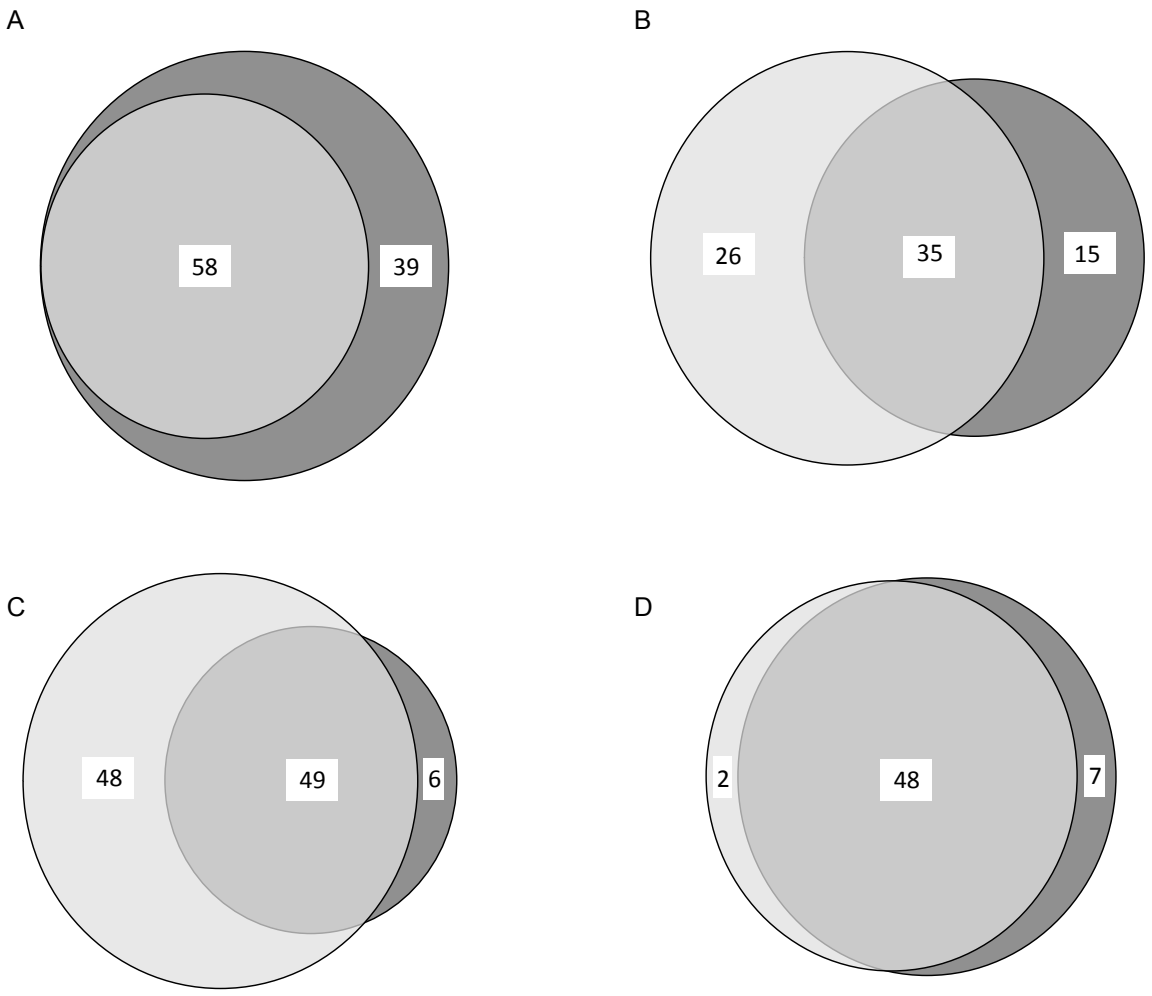


Figure 1: Venn diagrams depicting the total number of white perch plasma proteins detected by nanoLC-MS/MS in (A) male (left) and female (right) Initial Control (IC) plasmas, (B) IC male plasma (left) and 17β-estradiol (E_2)-induced male plasma (right), (C) IC female plasma (left) and E_2-induced female plasma (right), and (D) IC male (left) and E_2-induced female plasma (right) using the AUGUSTUS predicted open reading frames from the striped bass genome as a reference database.

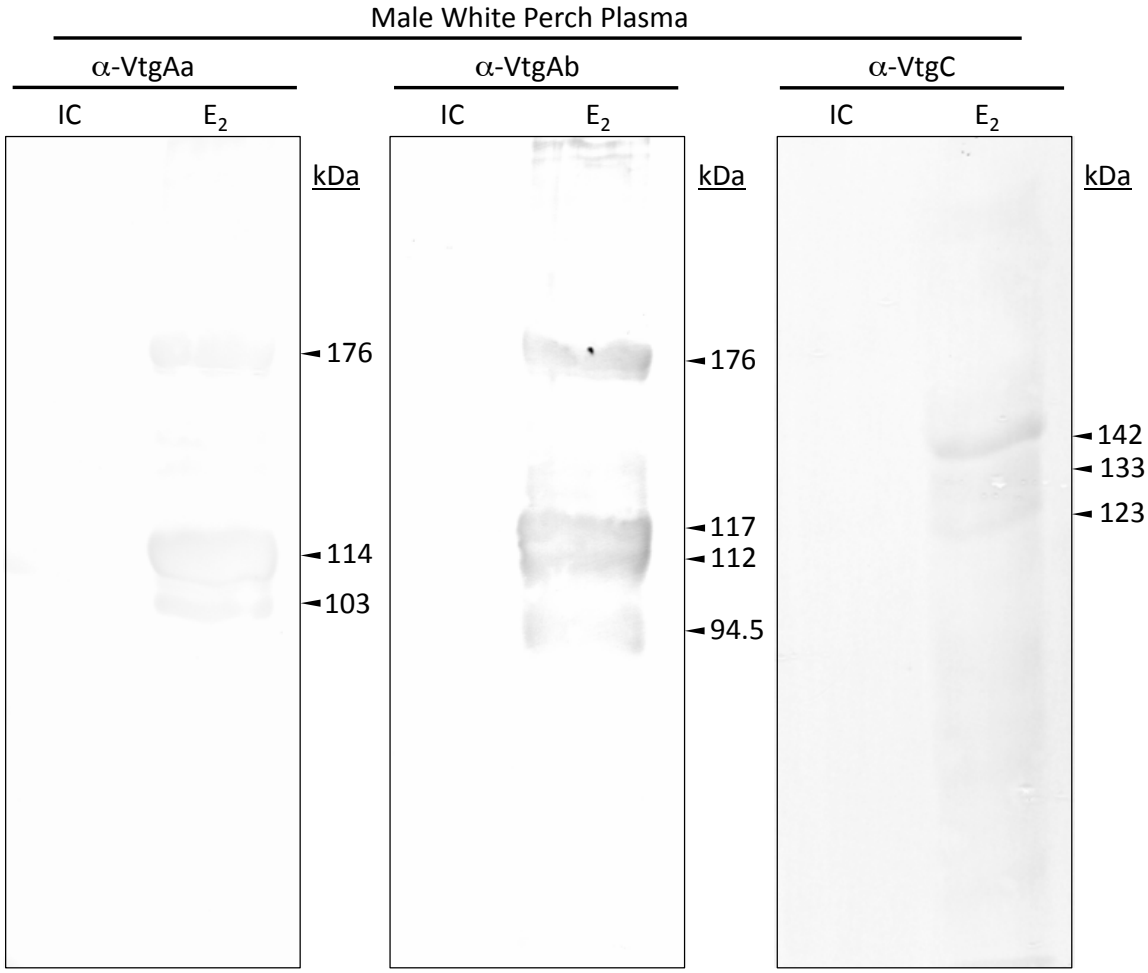


Figure 2. Western blotting of male white perch IC and E_2-induced plasma using polyclonal antisera raised against two unique synthetic peptides per target protein (VtgAa, VtgAb, and VtgC). No reactivity for any of the vitellogenins was seen in IC plasma, while all three forms of vitellogenin were detected in male white perch plasma following induction with E_2.

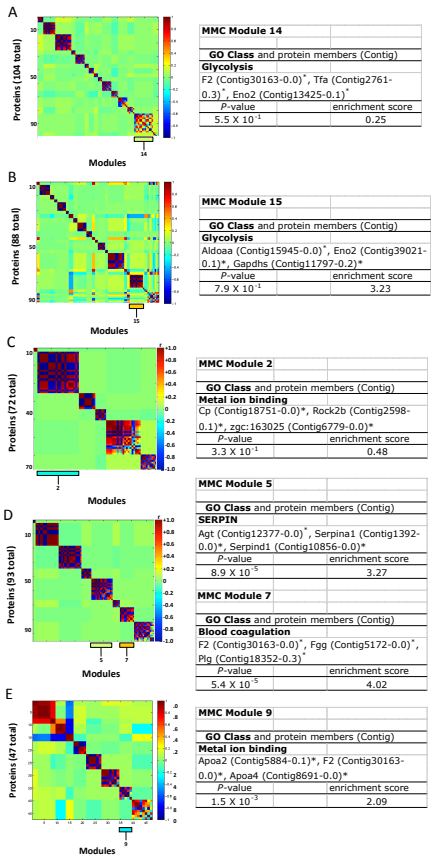


Figure 3. Modulated modularity clustering (MMC) of two-way ANOVA residual values from the white perch plasma protein expression dataset (left) and a table listing MMC module(s) with significant Gene Ontology (GO) enrichment within MMC modules using DAVID (right) for: The entire white perch plasma protein expression data set (A), IC male and IC female white perch plasma protein expression data (B), IC male and E_2-induced male white perch plasma protein expression data (C), IC female and E_2-induced female white perch plasma protein expression data (D), and E_2-induced male and E_2-induced female white perch plasma protein expression data (E). An asterisk after the Contig ID indicates that the protein varied significantly by at least one two-way ANOVA comparison after Benjamini-Hochberg false discovery rate correction.

Table 1. Number of proteins detected in the initial control (IC) male, IC female, 17β-estradiol (E_2)-induced male, and E_2-induced female white perch plasma samples by tandem mass spectrometry. Proteins with ≥ 0.25 average normalized spectral counts ($NSpC_{avg}$) across all twelve technical replicates (three technical replicates per treatment) were further analyzed by two-way ANOVA.

Plasma Sample	Total Proteins Identified	Proteins With $NSpC_{avg} \geq 0.25$
IC male	58	58
IC female	97	88
E_2-induced male	50	49
E_2-induced female	55	55

Table 2. Two-way ANOVA *p*-value range from comparisons between male and female white perch plasma proteins detected in internal control (IC) and 17β-estradiol (E_2)-induced samples by tandem mass spectrometry using the predicted proteins from the striped bass genome as a reference database. The number of proteins determined to be significantly different after the Bonferroni familywise error (BFW) and Benjamini-Hochberg false discovery rate (BH) corrections for multiple tests also are reported.

p-value range	Treatment (IC/E_2-induced)	Gender (Male/female)	Interaction (Treatment by Gender)
< 0.0001	47	14	23
0.0001-0.05	31	33	36
> 0.05	16	21	34
Significant by BFW	57	26	32
Significant by BH	77	44	57
Total Proteins	94	68	93

Table 3. Performance of support vector machines (SVMs) classifiers used to evaluate white perch plasma proteomes. Treatment groups were initial control (IC) male, IC female, 17β-estradiol (E_2)-induced male, and E_2-induced female. Data were clustered into either 3 or 2 clusters by *k*-means clustering or this preprocessing step was omitted. 'Correct Classification' refers to SVMs classifier performance during 66%-split or 10-fold cross-validations (CV) and is given as a percent (%). 'AUROC' is the area under the receiver operating characteristic curve.

Treatment groups and machine learning preprocessing steps	Classifier Performance Statistics	66%-split CV	10-fold CV
No preprocessing (IC male, IC female, E_2-induced male or E_2-induced female plasmas)	Correct classification	100.0	100.0
	Kappa statistic	1.0	1.0
	AUROC	1.0	1.0
***k*-means 3 clusters** (IC male, IC female, or E_2-induced plasmas of both genders)	Correct classification	100.0	100.0
	Kappa statistic	1.0	1.0
	AUROC	1.0	1.0
***k*-means 2 clusters** (E_2-induced male and E_2-induced female plasmas)	Correct classification	100.0	*not performed*
	Kappa statistic	1.0	*not performed*
	AUROC	1.0	*not performed*

Table 4. Performance of support vector machines (SVMs) classifiers used to evaluate white perch plasma proteomes. Treatment groups were initial control (IC) male, IC female, 17β-estradiol (E_2)-induced male, and E_2-induced female. Data were clustered into either 3 or 2 clusters by *k*-means clustering or this preprocessing step was omitted. 'Correct Classification' refers to SVMs classifier performance during 66%-split or 10-fold cross-validations (CV) and is given as a percent (%). 'AUROC' is the area under the receiver operating characteristic curve. Values reported represent the average ± standard deviation of SVMs models generated from 10 randomly ordered datasets.

Treatment groups and machine learning preprocessing steps	Classifier Performance Statistics	66%-split CV	10-fold CV
No preprocessing (IC male, IC female, E_2-induced male or E_2-induced female plasmas)	Correct classification	10.0 ± 24.2%	14.2 ± 14.7%
	Kappa statistic	0.022 ± 0.128	-0.144 ± 0.196
	AUROC	0.707 ± 0.190	0.389 ± 0.141
***k*-means 3 clusters** (IC male, IC female, or E_2-induced plasmas of both genders)	Correct classification	5.0 ± 21.1%	13.3 ± 12.5%
	Kappa statistic	0.00 ± 0.105	-0.156 ± 0.167
	AUROC	0.635 ± 0.240	0.376 ± 0.121
***k*-means 2 clusters** (E_2-induced male and E_2-induced female plasmas)	Correct classification	40.0 ± 20.1%	*not performed*
	Kappa statistic	0.000 ± 0.000	*not performed*
	AUROC	0.400 ± 0.210	*not performed*

Supporting information

Supplemental Table 1. A total of 104 unique proteins detected in white perch plasma samples when the nanoLC-MS/MS and ProteoIQ data were searched against the striped bass protein sequences predicted from the striped bass genome as a reference database using AUGUSTUS.

Supplemental Table 2. Nucleotide sequences predicted from the striped bass genome by AUGUSTUS for all 104 unique proteins detected in white perch plasma samples when the nanoLC-MS/MS and ProteoIQ data were searched against the predicted striped bass protein sequences as a reference database.

Supplemental Table 3. Striped bass AUGUSTUS predicted Contig ID from open reading frames (ORF) 0.0-0.5, approved gene name, and *p*-value for white perch plasma proteins that varied significantly by two-way ANOVA after a correction by Bonferroni familywise error (green) and Benjamini-Hockberg false discovery rate (yellow).

Supplemental Table 4. Striped bass AUGUSTUS predicted protein sequence ID, approved gene name, and residual values for the complete white perch plasma protein expression dataset from each of the replicates.

Supplemental Table 5. Outputs from each modulated modularity clustering (MMC) performed on white perch plasma protein expression data using two-way ANOVA 1) all residuals, 2) IC male and IC female residuals, 3) IC male and E_2-induced male residuals, 4) IC female and E_2-induced female residuals, and 5) E_2-induced male and E_2-induced female residuals.

Supplemental File 1. Glycolysis/gluconeogenesis KEGG Pathway analysis of the entire white perch plasma protein dataset as measured by mass spectrometry using the AUGUSTUS predicted open reading frames from the striped bass genome as a reference database. Detected proteins are indicated by a red box.

CHAPTER 4

Differences in the Proportional Accumulation of Vitellogenins and Implications on Early Life Histories in Two Closely Related Fish Species [1]

Justin Schilling[3], Philip L. Loziuk[4], David C. Muddiman[4,5], Harry V. Daniels[3], and Benjamin J. Reading[2,3]

[3]Department of Applied Ecology, North Carolina State University, Raleigh, North Carolina 27695, United States

[4]W. M. Keck FTMS Laboratory for Human Health Research, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States

[5]Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina 27607, United States

Running title: Multiple Vitellogenins and Their Receptors

Keywords: Vitellogenin, Mass Spectrometry, Selected Reaction Monitoring, Stable Isotope Labeled, Absolute Quantification

[1]**Grant Support**: This research was funded by North Carolina Hatch Project NC 06819 and the North Carolina Agricultural Foundation.

[2]**Correspondence**: Benjamin J. Reading, 127 David Clark Laboratories, Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695-7617, USA

E-mail: bjreadin@ncsu.edu

Summary sentence: Vitellogenin Ab is the predominant vitellogenin in white perch and is quantifiable in pre-vitellogenic ovary tissue while vitellogenin C is quantifiable in pre-

vitellogenic liver and localizes exclusively to lipid droplets in growing oocytes and does not bind any oocyte membrane receptor.

Presented in part at Aquaculture America, The Annual International Conference & Exposition of U.S. Aquaculture Society, National Aquaculture Association and U.S. Aquaculture Suppliers Association, 19-22 February 2015, New Orleans, Louisiana.

Abstract

We employed selected reaction monitoring tandem mass spectrometry to accurately quantify the three white perch (*Morone americana*) vitellogenins (VtgAa, VtgAb, VtgC) in the liver, plasma, and ovary during pre-, early-, mid-, and post-vitellogenic oocyte growth. Western blotting using antibodies raised against two highly unique antigenic peptides per target protein generally mirrored the selected reaction monitoring tandem mass spectrometry results. VtgC and VtgAb, however, were quantifiable by selected reaction monitoring in pre-vitellogenic ovary and liver tissues, respectively, but were not detected by Western blotting. Immunohistochemistry also indicated that VtgC was present within pre-vitellogenic oocytes and localized to lipid droplets within vitellogenic oocytes. Affinity purification coupled to tandem mass spectrometry using highly purified VtgC as a bait protein revealed a single specific interacting protein that eluted with suramin (Y-box binding protein 2a-like) and confirmed that VtgC does not bind the ovary vitellogenin receptors (LR8 and Lrp13). Western blotting for LR8 and Lrp13 showed that both receptors were expressed during vitellogenesis with LR8 and Lrp13 expression highest in early- and mid-vitellogenesis, respectively. As measured by selected reaction monitoring, the VtgAa within the ovary peaked during post-vitellogenesis, while VtgAb and VtgC peaked during early-vitellogenesis. The proportional abundance of white perch vitellogenins (VtgAa, VtgAb, VtgC) within post-

vitellogenic ovary tissues collected in May as measured by selected reaction monitoring were 5:35.7:1, respectively.

Introduction

Vitellogenins (Vtgs) are the predominant egg yolk precursor proteins in oviparous vertebrates. Vtgs are synthesized in the liver under the influence of estrogen, secreted into the bloodstream as ~500 kDa homodimers, and bind specific receptors on the surface of growing oocytes during a process termed *vitellogenesis*. Multiple vitellogenins have been described in a number of fish species including mummichog (*Fundulus heteroclitus*) (LaFleur et al., 1995; 2005), barfin flounder (*Verasper moseri*) (Matsubara, Ohkubo, Andoh, Sullivan, & Hara, 1999), haddock (*Melanogrammus aeglefimus*) (Reith, Munholland, Kelly, Finn, & Fyhn, 2001), zebrafish (*Danio rerio*) (Wang, Tan, Emelyanov, Korzh, & Gong, 2005; Wang, Yan, Tan, & Gong, 2000), mosquitofish (*Gambusia affinis*) (Sawaguchi, Koya, Yoshizaki, Ohkubo, Andoh, et al., 2005a; Sawaguchi, Ohkubo, Koya, & Matsubara, 2005), red seabream (*Pagrus major*) (Sawaguchi et al., 2006), Atlantic halibut (*Hippoglossus hippoglossus*) (Finn, 2007a,b), gray mullet (*Mugil cephalus*) (Amano et al., 2007a), goldsinny wrasse (*Ctenolabrus rupestris*) (Kolarevic, Nerland, Nilsen, & Finn, 2008), striped bass (*Morone saxatilis*) (Williams et al., 2014), and white perch (*Morone americana*) (Hiramatsu et al., 2002; Reading et al., 2009).

Complete type Vtgs of highly derived fish species (Acanthomorpha) are classified as either VtgAa or VtgAb based upon their structures using the nomenclature system proposed by Finn and Kristofferson (Finn & Kristoffersen, 2007). Complete type Vtgs contain five distinct yolk protein domains: Lipovitellin heavy chain (LvH), phosvitin (Pv), lipovitellin light

chain (LvL), beta-component (β'-c), and C-terminal domain (C-t). The third form of Acanthomorph vitellogenin, VtgC, is an incomplete type that lacks or has reduced Pv, β'-c, and C-t domains (Finn, 2007; Wang et al., 2000). In marine fishes that spawn floating eggs, for instance, the yolk protein domains of these three Vtgs undergo varying degrees of proteolysis into free amino acids (FAAs) during ovarian maturation. The LvH domain of VtgAa (LvHAa), for instance, is extensively proteolyzed, while those of VtgAb (LvHAb) and VtgC (LvHC) remain largely intact (Finn, 2007; Matsubara et al., 1999; Reith et al., 2001). The FAAs derived from VtgAa form an osmotic gradient that contributes to oocyte hydration and egg buoyancy in fish species that spawn pelagic eggs (Matsubara et al., 1999). These LvHAa-derived FAAs also serve as diffusible nutrients during embryogenesis while LvHAb remains largely intact and is consumed by later stage embryos (Amano et al., 2008; Amano, Fujita, Hiramatsu, Sawaguchi, et al., 2007a; Finn, Kolarevic, Kongshaug, & Nilsen, 2009; Matsubara et al., 1999; Sawaguchi et al., 2006; Sawaguchi, Ohkubo, Koya, & Matsubara, 2005). The LvHC supplies nutrients to developing yolk sac larvae (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005) while supplying nutrients to developing yolk sac larvae (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005). Varying ratios of these three Vtg forms accumulated in the egg yolk in turn lead to varying degrees of oocyte hydration and egg buoyancy that may relate to spawning strategies and likely have influence for early life history and development (Schilling et al., 2014; Williams et al., 2014).

Complete type Vtgs (VtgAa and VtgAb) have been shown to bind to two lipoprotein receptors on the surface of growing oocytes, LR8 and Lrp13. We have recently shown that white perch VtgAa binds primarily to Lrp13 while VtgAb binds to LR8 (Hiramatsu et al., 2012; Reading, Hiramatsu, & Sullivan, 2011; Reading, Williams, Chapman, Williams, & Sullivan,

2013). The exact mode of oocyte entry for VtgC has remained unclear, however it does not bind either of these oocyte membrane lipoprotein receptors (Reading et al., 2011).

Previous studies have provided both semiquantitative and quantitative analyses of vitellogenins in assorted tissues of oviparous vertebrates including fishes, sea turtle, and chickens by mass spectrometry (MS)-based proteomics (Andrews Kingon, Petitte, Muddiman, & Hawkridge, 2013; Cohen, Jahouh, & Sioud, 2009; Cohen, Mansour, & Banoub, 2006; Plumel et al., 2013; Schilling et al., 2014; Williams et al., 2014; 2013; Zhang, Bartels, Brodeur, & Woodburn, 2004). This study represents, to the best of our knowledge, the first absolute quantification of vitellogenins by selected reaction monitoring (SRM) liquid chromatography tandem mass spectrometry (LC-MS/MS) in liver, plasma, and ovary tissues across an entire reproductive cycle.

Materials and methods

Tissue collection

Adult female white perch were reared under natural photothermal conditions at the NC State Pamlico Aquaculture Field Laboratory (Aurora, NC) (Jackson & Sullivan, 1995). All experiments in the present study were carried out in accordance with an approved NC State University Animal Care and Use Committee (IACUC) protocol. Liver, plasma, and ovary were sampled at four time points across one year: August, November, February, and May. At each time point, three fish were anesthetized with Finquel MS-222 (Argent Chemical Laboratories, Redmond, WA) and tissues individually collected and separately stored. Blood was collected from the caudal vein by heparinized needle and syringe and transferred into 12 x 75 mm tubes containing 25 µL of aprotinin solution (40 TIU/mL aprotinin in 0.9% NaCl) (Sigma Aldrich, St. Louis, MO). Following centrifugation, the plasma was collected and

stored at -80 °C until use. Liver tissue was collected and stored at -80 °C until use. Ovary tissue was collected and stored at -80 °C until use. Additional ovary tissue was fixed in Bouin's in preparation for histology and immunohistochemistry (IHC). Ovary tissues were staged based upon the most advanced clutch of oocytes at each sampling time point (Jackson & Sullivan, 1995).

Filter-aided sample preparation of white perch tissues

Plasma, liver, and ovary samples were thawed and diluted with Tris-buffered saline (20 mM Tris-HCL pH 8.0, 150 mM NaCl, and 2 mM $CaCl_2$) to a final protein concentration of 1 mg/mL as determined by Bradford assay (Bradford, 1976; Noble & Bailey, 2009). Liver, plasma, and ovary samples from each fish at each time point were individually prepared for tandem mass spectrometry using a modified filter-aided sample preparation (FASP) protocol (Karp & Lilley, 2007; Schilling et al., 2014). Samples were incubated for 30 min at 56 °C with 15 µL of 50 mM DTT per 200 µL (200 µg) of sample to reduce disulfide bonds. Proteins were denatured by adding 200 µL of 8 M urea in 0.1 M Tris-HCl pH 8.0 to each sample. Each sample was then transferred onto a Vivacon 500 30 kDa MW cutoff filter (Sartorius Stedim Biotech, Goettingen, Germany) and centrifuged at 14,000 × g for 15 min at 21 °C. To ensure denaturation, an additional 200 µL of 8 M urea in 0.1 M Tris-HCl pH 8.0 was added to each sample before centrifuging again as above. Each sample was on-filter alkylated by adding 64 µL of 50 mM iodoacetamide (200mM final) prepared in 8 M urea atop each filter. Samples were incubated in the dark at 37 °C for 1 hr and then centrifuged at 14,000 × g for 15 min at 21 °C. Each filter was washed three times with 100 µL of 2 M urea/10mM $CaCl_2$ by centrifugation for 10 min at 14,000 × g followed by three washes with 100 µL of 0.1M Tris pH 7.5. Each sample was then placed in a new centrifuge tube and modified trypsin freshly

prepared in 0.1M Tris pH 7.5 was added to each sample at an enzyme to protein ratio of 1:5. Following eight hours of digestion at 37 °C, trypsinization was quenched with 50 µL of 0.001% zwittergent/1% formic acid and tryptic peptides were collected by centrifugation at 14,000 × g for 10 min at 21 °C. A second quench/wash step was carried out to maximize tryptic peptide recovery using 400 µL of 0.001% zwittergent/1% formic acid and centrifugation at 14,000 × g for 30 min at 21 °C. Samples were dried using a SpeedVac (Thermo Fisher Scientific, San Jose, CA) and stored at -20 °C.

LC-MS/MS materials

Unless otherwise stated, all reagents were purchased from Sigma-Aldrich (St. Louis, MO). All solvents were HPLC-grade from Honeywell Burdick & Jackson (Muskegon, MI).

Stable isotope-labeled peptide standards and transition characterization

Stable isotope-labeled (SIL) peptides uniquely identifying their respective protein were synthesized by New England Peptide (Gardner, MA). Newly obtained standards from New England Peptide were subjected to direct infusion into a TSQ Quantum triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose, CA) at 50 µM in order to identify the most abundant charge state for each peptide precursor. Standards were then fragmented and the 6 most abundant product ions were chosen for further characterization. Collision energy optimization experiments were performed by testing a range of collision energies and choosing the energy for which maximum signal intensity of each peptide was obtained.

LC-MS/MS analysis

Samples were analyzed in triplicate using a nanoLC-2D (AB SCIEX, Framingham, MA) equipped with an AS1 autosampler and coupled with TSQ Quantum triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose, CA) via a direct inject configuration in which 10 µL was injected onto a C18 reverse phase analytical column, 150 X 0.5 mm (Agilent Technologies, Santa Clara, CA) both packed with Zorbax SB-C18 (5 µm particle size). Mobile phases A and B were composed of water/acetonitrile/formic acid (98/2/0.2% and 2/98/0.2%, respectively). The sample was loaded at 12 µL/min for 3 min using mobile phase A. A 30-minute elution was then performed at a flow rate of 16 µL/min utilizing a 5-40% mobile phase B gradient. Compounds were ionized using an ion max source with a spray potential of 3.5 kV, a sheath gas pressure of 30 units, a capillary temperature of 270°C, and a skimmer offset of 8 V. Targets were analyzed using selected reaction monitoring (SRM) mode. Instrument parameters included an argon collision gas pressure of 1.5 mTorr, a full width half maximum of 0.7 *m/z* for Q1 and Q3, a scan width of 0.01 *m/z*, a scan time of 0.050 sec/transition and a chrom filter peak width of 45 seconds. Data from targeted LC-MS/MS experiments were imported into Skyline v.2.5.0.6079 (MacLean et al., 2010) where reproducible co-elution of native and SIL peptides along with their respective transitions was used to confirm the presence of the target peptide. Peak integration was performed after manual adjustment of integration windows. After performing consistent integration of all peptides across all samples, data were exported into Excel. Peak area ratios of native-to-SIL peptides were multiplied by the amount of each SIL peptide added during digestion to obtain absolute quantities in femtomoles (fmol) of target protein/µg of total protein. Since the amount of VtgC detected in the PreVG ovary samples was toward

the lower limit of confident quantification, these samples were run twice (Supplemental Table 1).

Experimental replication and statistical analysis

SRM data from three technical replicates of each independent biological replicate were analyzed by ANOVA using JMP Pro 10 (Copyright © 2013 SAS Institute Inc.). Mean values were considered statistically different when $P \leq 0.05$.

Antibody production

Antibodies against the two known white perch vitellogenin receptors (LR8 and Lrp13) and the three vitellogenins (VtgAa, VtgAb, and VtgC) were raised against two synthetic peptides for each protein in chickens (GeneTel, Madison, WI). Great care was taken to select highly unique antigenic sequences for each target protein (Supplemental Table 2). For the vitellogenins, each antigenic peptide was selected from the LvH domain, as our previous studies have shown that this region undergoes less proteolysis than the other domains (Reading et al., 2009; Williams et al., 2013). The resulting white perch vitellogenin antibodies are referred to hereafter as α-White Perch VtgAa, α-White Perch VtgAb, and α-White Perch Vtg C. Preimmune chicken IgY from each immunized hen was also purified to serve as a negative control for Western blotting and IHC validation.

Western blotting

While each tissue from each fish was individually prepared and analyzed by LC-MS/MS, samples from the same tissues and time points were pooled for Western blotting. Samples were resuspended and boiled in Laemmli's buffer and electrophoresed under

reducing and denaturing conditions through BIO-RAD Mini-PROTEAN TGX 4-15% Precast Gels with 10-well comb and 30 µL per well capacity (Catalog # 456-1083). All samples were loaded at 15 µg per well. Molecular weights were calculated based upon the migration distances of the BIO-RAD Kaleidoscope Precision Plus Protein Standards (Catalog # 161-0375). Primary antibodies were used at 1:5,000 and R&D Systems biotinylated goat α-chicken IgY secondary antibody (Catalog # BAF010) was used at 1:10,000. Blots were developed using Vector Laboratories VECTASTAIN Elite ABC-peroxidase per kit (Catalog # PK-6100) instructions.

White perch vitellogenin C purification

White perch vitellogenin C (VtgC) was purified from 1.8mL of 17β-estradiol (E_2)-induced male white perch plasma using a combination of anion exchange chromatography and size exclusion chromatography. Briefly, pooled E_2-induced plasma was diluted in an equal volume of 20 mM Tris-Bis-Propane, pH 9.0 containing 150 mM NaCl, 3.5 TIU/L aprotinin, 5% sucrose and applied to a POROS50 column. While complete type vitellogenins (VtgAa and VtgAb) bound to the POROS50 resin, VtgC did not and was present in the flow through. The flow through material was concentrated using a Centricon 10 kDa MW cutoff filter and applied to Superdex 200 gel filtration column in order to separate VtgC from remaining contaminating proteins. VtgC purity was assessed by Coomassie total protein staining and α-White Perch VtgC Western blotting.

White perch vitellogenin C affinity purification coupled to tandem mass spectrometry

Fractions containing highly purified white perch VtgC were coupled to for an affinity purification coupled to tandem mass spectrometry experiment (AP-MS/MS). Approximately 1

mg of highly purified white perch VtgC in 3 mL of binding buffer (20 mM Tris-HCl, 2 mM $CaCl_2$, and 150 mM NaCl, pH 8.0, containing 1 mM phenylmethyl-sulfonyl fluoride and 4 IU/L aprotinin) was coupled to 1 mL of Affi-Gel 15 slurry (~500 µl resin) per kit instructions. An uncoupled control resin was generated by incubating 1 mL of Affi-Gel 15 slurry (~500 µl resin) with 3 mL of binding buffer. Following coupling, any remaining reactive esters were blocked by adding 5 µl of 1 M ethanolamine HCl (pH 8.0) to each resin bed on a rotating wheel at 21 ºC for 1 hr. Both resin beds were then equilibrated with 10 column volumes of binding buffer. Solubilized membrane proteins were prepared from 5 g of vitellogenic white perch ovaries according to our previous studies (Reading et al., 2011; 2014) and incubated individually with control and VtgC-labeled affinity media for 4 hrs at 4 ºC in batch on a rotating wheel. The VtgC-labeled and unlabeled resins incubated with ovary solubilized membrane fraction were then washed with 15 volumes of binding buffer at 4 ºC prior to elution with 20 mM Tris-HCl, 10 mM suramin, 5 mM EDTA, 150 mM NaCl, pH 6.0. The column was equilibrated for 30 min at 4ºC before eluting proteins with 5 volumes of elution buffer (~1 mL). After additional washing with 15 volumes of binding buffer at 4 ºC, a second elution step followed using 20 mM Tris-HCl, 5 mM EDTA, 1.5 M NaCl, pH 6.0 with which the resin was equilibrated for 30 min at 4 ºC before eluting proteins with 5 volumes of elution buffer (~1mL).

White perch vitellogenin C AP-MS/MS filter-aided sample preparation

Elution fractions were reduced with DTT at a final concentration of 5mM at 37 °C for 30 mins and then mixed with 200 µl of 8 M urea in 0.1 M Tris-HCL pH 7.0 and placed atop Vivacon 500 30 kDa MW cutoff filters and centrifuged at 14,000 x g for 15 mins. An additional 200 µl of 8 M urea in 0.1 M Tris-HCL pH 7.0 was added to each filter unit and they

were centrifuged again at 14,000 x g for 15 mins. The flow through was discarded from each collection tube. Samples were alkylated with 100 µl of 0.05 M iodoacetamide prepared in 8 M urea added to each sample, mixed for 1 min and incubated in the dark at 21 °C for 20 mins before centrifuging at 14,000 x g for 15 mins. Samples were further denatured with 100 µl 8 M urea by centrifuging at 14,000 x g for 10 mins. This was repeated two additional times. Samples were then washed with 100 µl of 0.1 M Tris-HCL and centrifuged at 14,000 x g for 15 mins. This was repeated twice. A fresh collection vial was placed beneath each filter unit and 40 µl 0.1 M Tris-HCl containing modified trypsin (1:100 ratio of trypsin:total protein) was added to each sample and mixed for 1 min. Filter units were then sealed with parafilm and incubated at 37 °C for 18 hrs. Following trypsinization, 1 µl of 2% formic acid was added to each sample. All samples were then centrifuged at 14,000 x g for 15 mins. To maximize recovery of tryptic peptides, an additional 40 µl of 0.1 M Tris-HCl was added atop each filter unit and they were again centrifuged at 14,000 x g for 15 mins. Final trypsinized protein concentrations were measured using a Nanodrop at A280 (Thermo Scientific, Wilmington, DE).

White perch vitellogenin C AP-MS/MS nanoReversed phase LC-MS/MS

Protein concentrations of the digests were obtained using a Nanodrop at A280 (Thermo Scientific, Wilmington, DE). Samples were reconstituted to a protein concentration of 0.2 µg/µL using mobile phase A (98/2/0.2% water/acetonitrile/ formic acid), and a total of 5 µL was injected onto the trap for desalting. Separation of peptides was performed using a Thermo Scientific EASY nLC II (Thermo Scientific, San Jose, CA) in line with a cHiPLC nanoflex system (AB Sciex, Framingham, MA). A vented column configuration, a ChromXP C18-CL 3 µm trap column, and a ChromXP C18- CL 75 µm × 15 cm analytical column were

used for these experiments. The initial condition of 2% mobile phase B (2/ 98/0.2% water/acetonitrile/formic acid) was increased to 35% over 201 min, then steeply ramped to 95% mobile phase B over 10 min and maintained at 95% mobile phase B for 10 min to wash the column. The column was then equilibrated at 2% mobile phase B for 12 min. Each sample was analyzed in triplicate.

Results

Histology and oocyte staging

The most advanced clutch of oocytes within white perch ovary tissues sampled in August was of the pre-vitellogenic stage (PreVG) (Figure 1 A). The most advanced clutch of oocytes within white perch ovary tissues sampled in November was of the early-vitellogenic stage (EVG) (Figure 1 B). The most advanced clutch of oocytes within white perch ovary tissues sampled in February was of the mid-vitellogenic stage (MVG) (Figure 1 C). The most advanced clutch of oocytes within white perch ovary tissues sampled in May was of the post-vitellogenic stage (PostVG) (Figure 1 D). Length and weight statistics are provided in Table 1.

Selected reaction monitoring

In white perch, VtgAb is the predominant vitellogenin detected in liver, plasma, and ovary during vitellogenesis (Figures 2 and 3; Table 2). VtgAb was most abundant during MVG in the liver, plasma, and ovary as measured by SRM (5.706 ± 0.330, 61.757 ± 3.209, and 1,557.167 ± 27.513 fmol/ug total protein, respectively).

As measured by SRM, VtgAa was not quantifiable in any PreVG tissue and was most abundant in PostVG liver, plasma, and ovary tissues (1.419 ± 0.118, 17.529 ± 0.757, and 165.617 ± 6.815 fmol/ug total protein) (Figures 2 and 3; Table 2).

Vitellogenin C was not quantifiable by SRM in liver tissues at any stage of oocyte growth (Figures 2 and 3; Table 2). VtgC was most abundant in PostVG plasma as measured by SRM (3.847 ± 0.475 fmol/ug total protein) (Figures 2 and 3; Table 2), while it was most abundant in EVG ovary tissues (49.212 ± 4.052 fmol/ug total protein) (Figures 2 and 3; Table 2).

In PreVG tissues, only VtgAb could be confidently quantified in liver (Figures 2 and 3) while only VtgC could be confidently quantified in the ovary (Figures 2 and 3). None of the vitellogenins were quantifiable in PreVG plasma (Figures 2 and 3).

The proportional abundance of white perch VtgAa, VtgAb, and VtgC within PostVG ovary tissues as measured by SRM were 5:35.7:1, respectively (Table 2). An example extraction ion chromatogram depicting the co-elution of heavy and light VtgAb peptides is given in Supplemental Figure 1.

Western blotting

The Western blotting results from white perch liver, plasma, and ovary tissues for VtgAa, VtgAb, and VtgC are presented in Figure 4. In the liver, protein bands of 126.2 kDa and 76.4 kDa reacted with α-White Perch VtgAa in EVG and MVG tissues collected in November and February, respectively. Protein bands of 128.5, 116.8, and 83.7 kDa reacted with α-White Perch VtgAb in EVG, MVG, and PostVG liver tissues. Protein bands of 134.2 and 80.5 kDa reacted with α-White Perch VtgC in MVG and EVG liver tissues.

Western blotting of plasma for VtgAa revealed a reactive protein band of 111 kDa from during vitellogenesis. There is also a 29.8 kDa band visible in the PostVG plasma sample. Multiple protein bands were visible in vitellogenic plasma samples stained with α-White Perch VtgAb ranging from 236.9 to 44.5 kDa. There are faint bands of 66.9 and 29.8 kDa in plasma stained with α-White Perch VtgC.

All three vitellogenins are detectable during vitellogenesis in ovary tissues by Western blotting. Staining with α-VtgAa produced two bands of 84.1 and 32.9 kDa from early-, mid-, and post-vitellogenic ovary tissues, with a larger 135.8 kDa band present in EVG ovary tissues. Vitellogenin Ab was detected as a band of 83.8 kDa from EVG through PostVG, with a 32.4 visible in MVG. Similarly, VtgC was detected mainly as one band of 79.4 from EVG through PostVG with a 29 kDa band visible in MVG.

Western blotting for white perch LR8 and Lrp13 indicate that LR8 expression was highest in the ovary during EVG, while Lrp13 expression was highest in the MVG ovary (Figure 5). Neither LR8 nor Lrp13 was detectable in plasma or liver by Western blotting (Figure 5).

White perch vitellogenin C immunohistochemistry

Confocal microscopy imaging of white perch PreVG ovary sections stained with α-White Perch VtgC covalently coupled to DyLight633 indicates that VtgC is present within primary growth perinucleolar primary growth oocytes in PreVG ovary tissues (Figure 6). As oocyte growth progresses into vitellogenesis, it is apparent that VtgC localizes exclusively to lipid droplets (Figure 6).

White perch vitellogenin C AP-MS/MS

Tandem mass spectrometry confirmed that the VtgC bait protein was highly purified (*data not shown*). Elution from the VtgC-labeled Affi Gel using suramin yielded a single specific binding protein, Y-box binding protein 2a-like (ybx2al). Total peak area for Y-box binding protein 2a-like was 9.75 X 10^6 for the VtgC-labeled Affi Gel and 0.00 for the control Affi Gel (Table 3).

Discussion

In this study we present the first absolute quantification by selected reaction monitoring tandem mass spectrometry of white perch vitellogenins in liver, plasma, and ovary sampled at four stages across one reproductive year. Western blotting using antibodies raised against unique, Vtg-specific synthetic peptides largely supports the quantitative mass spectrometry data. In further agreement with the quantitative mass spectrometry data, immunohistochemistry reveals that VtgC is present in PreVG ovary tissues and localizes to lipid droplets in vitellogenic oocytes. Affinity purification coupled to tandem mass spectrometry identified a specific interaction between VtgC and Y-box-binding protein 2a-like that was disrupted by 10mM suramin. Finally, we show that VtgAb is the predominant vitellogenin in the post-vitellogenic perch ovary (85.6%), with VtgAa and VtgC comprising 11.99% and 2.39%, respectively.

While VtgAb is the predominant perch vitellogenin in liver, plasma, and ovary during vitellogenesis, SRM also indicates that the proportion of VtgAa accumulated in the PostVG ovary increases from MVG (Figure 3). A similar pattern of vitellogenin uptake in ovary tissues of striped bass has been reported recently (Williams et al., 2014), suggesting that the oocytes substantially accumulate VtgAb from EVG through MVG and preferentially

accumulate VtgAa from MVG through PostVG. As in the striped bass, accumulation VtgAa and VtgAb by the oocyte is not solely dependent upon the circulating plasma Vtg concentrations. Their accumulation does, however, correlate with the expression patterns of the Vtgrs LR8 and Lrp13.

Neither VtgAa nor VtgAb could be confidently quantified in the PreVG white perch ovary. We were unable to confidently quantify VtgC in white perch liver tissues at any stage of oocyte growth. In our previous study of striped bass vitellogenins, VtgC was also undetectable by nanoLC-MS/MS in liver tissues sampled across one reproductive year, despite comprising nearly a quarter (26.1%) of the total vitellogenin accumulated in late-vitellogenic oocytes (Williams et al., 2014). Although it is present at low levels (17.005 ± 3.209 fmol/ug total protein; Table 2; Figure 3), we were able to confidently quantify VtgC in PreVG ovary tissues. Given 1) its limited estrogen responsiveness in white perch and other fish species (Ohkubo et al., 2003), 2) its continuous deposition into perch oocytes from Pre- through Post-VG (Figures 2 and 3) (Reading et al., 2008), 3) its inability to bind the LR8 and Lrp13 vitellogenin receptors in perch (Table 3) (Reading et al., 2011; 2014) and 4) its species-specific high variation in percent contribution to the mature egg yolk (Williams et al., 2014), VtgC from Acanthomorph fishes appears to behave more similarly to other lipoproteins than to complete type vitellogenins (Babin & Vernier, 1989). It is interesting that striped bass, with 26.1% VtgC in their mature oocytes, spawn neutrally buoyant eggs, while white perch, with just 2% VtgC in their mature oocytes, ovulate adhesive eggs (Williams et al., 2014). Further investigation will be required to understand the molecular basis underlying these differences in life history and whether or not they might relate to highly variable amounts of VtgC within mature oocytes of different fish species.

Western blotting results for white perch VtgAa, VtgAb, and VtgC across all tissues and time points generally confirm the findings of quantitative mass spectrometry (Figures 2 and 3). Two exceptions, however, were that VtgAb and VtgC were not clearly detected by Western blotting in the PreVG liver and ovary, respectively (Figure 3), whereas they were detected by mass spectrometry in these tissues at low levels (Figure 2). Additionally, VtgC was detected in PreVG ovary by immunohistochemistry (see below). Such discrepancies between quantitative mass spectrometry data and Western blotting data are not surprising and have been reported for striped bass vitellogenins as well (Williams et al., 2014). Given the highly variable performance of antibodies in immunoassays and the high sensitivity of mass spectrometry, the practical utility of validating quantitative mass spectrometry results by Western blotting has been brought into question (Aebersold, Burlingame, & Bradshaw, 2013).

Our Western blotting results are consistent with those in our previous study on striped bass vitellogenins (Williams et al., 2013; 2014). The antibodies used in the present study were prepared using two unique antigenic synthetic peptides within the LvH domain of each vitellogenin. As such, the α-White Perch Vtg Western blots in the present study generally showed fewer reactive bands than those in our previous studies, which utilized antibodies raised against both the LvH and lipovitellin light (LvL) domains of vitellogenins purified from plasma of E_2-induced mullet (*Mugil cephalus*) in the case of α-Mullet Vtgs (Amano et al., 2008; Amano, Fujita, Hiramatsu, Sawaguchi, et al., 2007a; Amano, Fujita, Hiramatsu, Shimizu, et al., 2007b; Williams et al., 2013; 2014), and against purified white perch VtgAa, VtgAb, and VtgC in the case of the α-WP Vtgs used by Hiramatsu et al. in 2002 (Hiramatsu et al., 2002). No reactivity was seen by Western blotting in PreVG liver,

PreVG plasma, or PreVG ovary samples with any of the antibodies used in this study (Figures 3 and 4).

Western blotting with α-White Perch VtgAa resulted in a major band of 126.2 kDa in the EVG and MVG liver samples, with a faint 76.4 kDa band present in the EVG sample. This banding pattern is consistent with the α-Mullet VtgAa Western blotting of striped bass liver samples and the α-WP Vtgs Western blotting of striped bass post-vitellogenic and ovulated oocyte extracts in our previous studies (Williams et al., 2013; 2014). Western blotting of perch plasma showed an apparent VtgAa band of 111 kDa was present in EVG, MVG, and PostVG samples, with a 29.8 kDa band visible in the PostVG sample. Vitellogenins are progressively proteolyzed into yolk proteins during oocyte growth and α-Vtg Western blotting, both in the present study and our previous studies, reflect this (Williams et al., 2013; 2014). Specifically, α-White Perch VtgAa Western blotting of ovary tissues revealed two reactive bands of 84.1 and 32.9 kDa from EVG through PostVG, with an additional band of 135.8 kDa visible in the EVG sample.

Western blotting with α-White Perch VtgAb resulted bands of 128.5 116.8, and 83.7 kDa in EVG, MVG, and PostVG white perch liver samples. EVG, MVG, and PostVG plasma samples showed reactive bands of 236.9, 218.2, 127.7, 121.7, and 115.9 kDa when stained with α-White Perch VtgAb. The MVG plasma sample had additional reactive bands of 169.7 and 162.9 kDa, while the PostVG sample had an additional faint band of 44.5 kDa. In ovary tissues, α-White Perch VtgAb reacted with an 83.8 kDa band from EVG through PostVG, with a faint 32.4 kDa appearing in the MVG and PostVG samples.

Western blotting with α-White Perch VtgC revealed a band of 134.2 kDa in EVG, MVG, and PostVG white perch liver samples. An additional 80.5 kDa band was faintly visible

in the EVG and MVG samples. A diffuse band of 66.9 kDa was faintly visible in PreVG, EVG, MVG, and PostVG plasma samples stained with α-White Perch VtgC. There was an additional faint 29.8 kDa band in the PostVG plasma sample. A 79.4 kDa band was found in EVG, MVG, and PostVG ovary samples when stained with α-White Perch VtgC with an additional 29 kDa band faintly visible in the MVG ovary sample.

Immunohistochemisty (IHC) of PreVG perch ovary tissues using α-White Perch VtgC covalently linked to DyLight633 indicates that VtgC is present in the ooplasm of secondary growth oocytes within the PreVG ovary (Figure 6). As ovarian maturation progresses into vitellogenesis, IHC reveals that VtgC is primarily sequestered into a ring of lipid droplets in the peripheral ooplasm (Figure 6). To our knowledge, this represents the first report of VtgC localization of within growing oocytes.

Using the striped bass ovary transcriptome as a reference database, a single protein, Y-box binding protein 2a-like (Ybx2al), was found to interact specifically with VtgC by AP-MS/MS. This protein specifically eluted with the known disruptor of lipoprotein-lipoprotein receptor interactions, suramin (Brown, Via, Gotto, Bradley, & Gianturco, 1986). Y-box-binding proteins are highly expressed in germ cells and bind Y-box elements in the promoters of certain genes and also bind to mRNA transcribed from these genes (Safran et al., 2010). A ClustalW dendrogram of vertebrate Y-box-binding proteins (Ybx1, Ybx2, Ybx3, and Ybx2al) places Ybx2al proteins in a distinct group of largely uncharacterized predicted proteins in a number of fish species (Supplemental Figure 2). While Ybx2al may be a nuclear contaminating protein that co-purified during the ovary membrane protein preparation, it was detected primarily in the membrane fraction in our previous compartment proteomics study of the white perch ovary (Schilling et al., 2014). Sequence analysis of the eleven Yb2al protein sequences presented in Supplemental Figure 2 indicates that they all

lack a transmembrane domain and are likely soluble proteins. Further investigation into the sub-cellular localization, function, and interactions of this protein are needed.

In the goldsinny wrasse, a fish species which spawns hyper-pelagic eggs, the VtgAa comprises nearly all of the vitellogenin derived egg yolk (Kolarevic et al., 2008). The proportional ratios of VtgAa : VtgAb, on the other hand, appears to fall within a range of 1 : 1 to 1 : 7.1 for striped bass and white perch, respectively. Therefore, differences in VtgAa and VtgAb appear to correlate to differences in egg buoyancy between these two closely related species, as the striped bass spawns neutrally buoyant eggs and the white perch spawns demersal adhesive eggs. In barfin flounder, a marine species that spawns pelagic eggs, the ratio of VtgAa : VtgAb is 9 : 15 (Sawaguchi et al., 2008), however the substantially higher proportion of VtgAb in the egg yolk of white perch may relate to its characteristic demersal eggs. More information regarding the egg yolk compositions of various fishes will be required to fully understand the influence of vitellogenin on egg buoyancy.

The three Perciform Vtgs are known to undergo different degrees of proteolysis within growing oocytes (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005). While VtgAa and, to a lesser extent, VtgAb are proteolyzed into free amino acids or small peptides, VtgC appears to remain largely intact through oocyte maturation and ovulation and remains available to developing embryos as a food source before they are able to feed on their own. In mosqitofish, for example, the VtgC-derived yolk proteins are the last components of the egg yolk that remain to be consumed by yolk-sac fry (Sawaguchi, Ohkubo, Koya, & Matsubara, 2005). Considering the data that are currently available, it appears that VtgC is the most variable form of vitellogenin within the post-vitellogenic oocytes of Acanthomorph fishes, ranging from ~2.5% in perch to 26% in striped bass (Figure 7) (Schilling et al., 2014;

Williams et al., 2014). Therefore, VtgC composition may relate to other aspects of early life history in these fishes.

White perch reared indoors in recirculating aquaculture systems (< 5 ppt salinity) had proportional vitellogenin ratios of 7.7 : 16 : 1 in PostVG ovary tissues as reported in our previous study (Schilling et al., 2014). In the present study, white perch were reared outdoors in a flow through system supplied with estuary creek water (7-15 ppt salinity) and the proportional vitellogenin ratios in PostVG ovary tissues were 5 : 35.7 : 1 (Table 2). The observed proportional ratio differences may be due to differences in culture conditions, differences in salinity, and/or differences in detection methods (i.e. semiquantitative tandem mass spectrometry versus absolute quantification by selected reaction monitoring tandem mass spectrometry). These differences in the proportional accumulation of Vtgs suggest that there is considerable plasticity in the Vtg-Vtgr system in white perch ovary tissues. Such plasticity may allow fine-tuning of egg buoyancy based upon the specific gravity of the water into which the eggs are to be spawned, which can vary considerably year-to-year depending upon wind, freshwater flow, and the location of the salt front (North & Houde, 2003). While the Vtgs were not analyzed, eggs in the striped bass varied significantly in size, density, oil globule diameter, and the size of the oil globule relative to oocyte diameter depending upon the watershed from which they were collected (Bergey et al., 2003).

Fertilized white perch and striped bass larvae both hatch around 48 hrs post-fertilization (Eldridge, Whipple, Eng, Bowers, & Jarvis, 1981; Mansueti, 1964; North & Houde, 2003). When food is then restricted or deprived, white perch larvae survive for up to 15 days post-fertilization while striped bass larvae survive for up to 31 days post-fertilization (Figure 8) (Eldridge et al., 1981; Mansueti, 1964). Additionally, the time to first feeding of these two species differs, with the white perch and the striped bass larvae beginning to feed

at 3-5 days and 7-9 days post-fertilization, respectively. Therefore, although these closely related species share similar time frames during the earliest stages of development (i.e. from fertilization to hatch), the striped bass larvae appear to have an extended developmental window from hatching to first feeding, and these larvae have yolk stores that allow them to survive in the absence of food for twice as long as white perch after hatch. This disparity in early developmental stages post-hatch may relate to differences in VtgC yolk content of the white perch and striped bass eggs, which is 2.5-5% and 26%, respectively. As more data from a wider range of fish species become available, the complexity of the egg yolk system in general and the Vtg-Vtgr system in particular and their relation to the diversity of early life history strategies among fishes may become apparent.

Acknowledgements

We thank Dr. Andy McGinty and Michael Hopper for their efforts caring for the white perch.

References

Aebersold, R., Burlingame, A. L., & Bradshaw, R. A. (2013). Western Blots versus Selected Reaction Monitoring Assays: Time to Turn the Tables? *Molecular & Cellular Proteomics, 12*(9), 2381–2382. doi:10.1074/mcp.E113.031658

Amano, H., Fujita, T., Hiramatsu, N., Kagawa, H., Matsubara, T., Sullivan, C. V., & Hara, A. (2008). Multiple vitellogenin-derived yolk proteins in gray mullet (*Mugil cephalus*): disparate proteolytic patterns associated with ovarian follicle maturation. *Molecular Reproduction and Development*, *75*(8), 1307–1317. doi:10.1002/mrd.20864

Amano, H., Fujita, T., Hiramatsu, N., Sawaguchi, S., Matsubara, T., Sullivan, C. V., & Hara, A. (2007a). Purification of multiple vitellogenins in grey mullet (*Mugil cephalus*). *Marine Biology*, *152*(6), 1215–1225. doi:10.1007/s00227-007-0768-z

Amano, H., Fujita, T., Hiramatsu, N., Shimizu, M., Sawaguchi, S., Matsubara, T., et al. (2007b). Egg yolk proteins in gray mullet (*Mugil cephalus*): purification and classification of multiple lipovitellins and other vitellogenin-derived yolk proteins and molecular cloning of the parent vitellogenin genes. *Journal of Experimental Zoology. Part a, Ecological Genetics and Physiology*, *307*(6), 324–341. doi:10.1002/jez.388

Andrews Kingon, G. L., Petitte, J. N., Muddiman, D. C., & Hawkridge, A. M. (2013). Multi-peptide nLC-PC-IDMS-SRM-based assay for the quantification of biomarkers in the chicken ovarian cancer model. *Methods*, *61*(3), 323–330. doi:10.1016/j.ymeth.2013.04.004

Babin, P. J., & Vernier, J. M. (1989). Plasma lipoproteins in fish. *Journal of Lipid Research*, *30*, 467-489.

Bergey, L., Rulifson, R. A., Gallagher, M. L., & Overton, A. S. (2003). Variability of Atlantic Coast Striped Bass Egg Characteristics. *North American Journal of Fisheries Management*, *23*(2), 558-572. doi: 10.1577/15488675(2003)023<0558:VOACSB>2.0.CO;2

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254.

Brown, S. A., Via, D. P., Gotto, A. M., Bradley, W. A., & Gianturco, S. H. (1986). Apolipoprotein E-mediated binding of hypertriglyceridemic very low density lipoproteins to isolated low density lipoprotein receptors detected by ligand blotting. *Biochemical and Biophysical Research Communications*, *139*(1), 333–340.

Cohen, A. M., Jahouh, F., & Sioud, S. (2009). Quantification of Greenland halibut serum vitellogenin: a trip from the deep sea to the mass spectrometer. *Rapid Communications in Mass Spectrometry*, *23*, 1049–1060. doi: 10.1002/rcm.3966

Cohen, A. M., Mansour, A. A. H., & Banoub, J. H. (2006). Absolute quantification of Atlantic salmon and rainbow trout vitellogenin by the “signature peptide” approach using electrospray ionization QqToF tandem mass spectrometry. *Journal of Mass*

Spectrometry, *41*(5), 646–658. doi:10.1002/jms.1023

Eldridge, M. B., Whipple, J. A., Eng, D., Bowers, M. J., & Jarvis, B. M. (1981). Effects of Food and Feeding Factors on Laboratory-Reared Striped Bass Larvae. *Transactions of the American Fisheries Society*, *110*(1), 111–120. doi:10.1577/1548-8659(1981)110<111:EOFAFF>2.0.CO;2

Finn, R. N. (2007). The maturational disassembly and differential proteolysis of paralogous vitellogenins in a marine pelagophil teleost: a conserved mechanism of oocyte hydration. *Biology of Reproduction*, *76*(6), 936–948. doi:10.1095/biolreprod.106.055772

Finn, R. N., & Kristoffersen, B. A. (2007). Vertebrate Vitellogenin Gene Duplication in Relation to the “3R Hypothesis”: Correlation to the Pelagic Egg and the Oceanic Radiation of Teleosts. *PloS One*, *2*(1), e169. doi:10.1371/journal.pone.0000169

Finn, R. N., Kolarevic, J., Kongshaug, H., & Nilsen, F. (2009). Evolution and differential expression of a vertebrate vitellogenin gene cluster. *BMC Evolutionary Biology*, *9*, 2. doi:10.1186/1471-2148-9-2

Hiramatsu, N., Luo, W., Reading, B. J., Sullivan, C. V., Mizuta, H., Ryu, Y. W., et al. (2012). Multiple ovarian lipoprotein receptors in teleosts. *Fish Physiology and Biochemistry*. doi:10.1007/s10695-012-9612-6

Hiramatsu, N., Matsubara, T., Hara, A., Donato, D. M., Hiramatsu, K., Denslow, N. D., &

Sullivan, C. V. (2002). Identification, purification and classification of multiple forms of vitellogenin from white perch (*Morone americana*). *Fish Physiology and Biochemistry*, *26*(4), 355–370. doi:10.1023/B:FISH.0000009266.58556.9a

Jackson, L. F., & Sullivan, C. V. (1995). Reproduction of white perch: the annual gametogenic cycle. *Transactions of the American Fisheries Society*, *124*(4), 563–577. doi: 10.1577/1548-8659(1995)124<0563:ROWPTA>2.3.CO;2

Karp, N. A., & Lilley, K. S. (2007). Design and Analysis Issues in Quantitative Proteomics Studies. *Proteomics*, *7*(S1), 42–50. doi:10.1002/pmic.200700683

Kolarevic, J., Nerland, A., Nilsen, F., & Finn, R. N. (2008). Goldsinny wrasse (*Ctenolabrus rupestris*) is an extreme vtgAa-type pelagophil teleost. *Molecular Reproduction and Development*, *75*(6), 1011–1020. doi:10.1002/mrd.20845

LaFleur, G. J., Byrne, B. M., Kanungo, J., Nelson, L. D., Greenberg, R. M., & Wallace, R. A. (1995). *Fundulus heteroclitus* vitellogenin: the deduced primary structure of a piscine precursor to noncrystalline, liquid-phase yolk protein. *Journal of Molecular Evolution*, *41*(4), 505–521.

LaFleur, G. J., Raldúa, D., Fabra, M., Carnevali, O., Denslow, N., Wallace, R. A., & Cerdà, J. (2005). Derivation of major yolk proteins from parental vitellogenins and alternative processing during oocyte maturation in *Fundulus heteroclitus*. *Biology of Reproduction*, *73*(4), 815–824. doi:10.1095/biolreprod.105.041335

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, *26*(7), 966–968. doi:10.1093/bioinformatics/btq054

Mansueti, R. J. (1964). Eggs, Larvae, and Young of the White Perch, *Roccus americanus*, with Comments on Its Ecology in the Estuary. *Chesapeake Science*, *5*(1/2), 3-45. doi:10.2307/1350789

Matsubara, T., Ohkubo, N., Andoh, T., Sullivan, C. V., & Hara, A. (1999). Two forms of vitellogenin, yielding two distinct lipovitellins, play different roles during oocyte maturation and early development of barfin flounder, *Verasper moseri*, a marine teleost that spawns pelagic eggs. *Developmental Biology*, *213*(1), 18–32. doi:10.1006/dbio.1999.9365

Noble, J. E., & Bailey, M. J. A. (2009). Chapter 8 Quantitation of Protein. In *Methods in enzymology* (Vol. 463, pp. 73–95). Elsevier. doi:10.1016/S0076-6879(09)63008-1

North, E. W., & Houde, E. D. (2003). Linking ETM physics zooplankton prey and fish early life histories to striped bass *Morone saxatilis* and white perch *M. americana* recruitment. *Marine Ecology Progress Series*, *260*, 219-236.

Ohkubo, N., Mochida, K., Adachi, S., Hara, A., Hotta, K., Nakamura, Y., & Matsubara, T. (2003). Development of enzyme-linked immunosorbent assays for two forms of

vitellogenin in Japanese common goby (*Acanthogobius flavimanus*). *General and Comparative Endocrinology*, *131*(3), 353–364.

Plumel, M. I., Wasselin, T., Plot, V., Strub, J.-M., Van Dorsselaer, A., Carapito, C., et al. (2013). Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (*Dermochelys coriacea*). *Journal of Proteome Research*, *12*(9), 4122–4135. doi:10.1021/pr400444m

Reading, B. J., Hiramatsu, N., & Sullivan, C. V. (2011). Disparate binding of three types of vitellogenin to multiple forms of vitellogenin receptor in white perch. *Biology of Reproduction*, *84*(2), 392–399. doi:10.1095/biolreprod.110.087981

Reading, B. J., Hiramatsu, N., Sawaguchi, S., Matsubara, T., Hara, A., Lively, M. O., & Sullivan, C. V. (2008). Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. *Marine Biotechnology (New York, N.Y.)*, *11*(2), 169–187. doi:10.1007/s10126-008-9133-6

Reading, B. J., Hiramatsu, N., Sawaguchi, S., Matsubara, T., Hara, A., Lively, M. O., & Sullivan, C. V. (2009). Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. *Marine Biotechnology (New York, N.Y.)*, *11*(2), 169–187. doi:10.1007/s10126-008-9133-6

Reading, B. J., Hiramatsu, N., Schilling, J., Molloy, K. T., Glassbrook, N., Mizuta, H., et al. (2014). Lrp13 is a novel vertebrate lipoprotein receptor that binds vitellogenins in teleost fishes. *The Journal of Lipid Research*, *55*(11), 2287–2295. doi:10.1194/jlr.M050286

Reading, B. J., Williams, V. N., Chapman, R. W., Williams, T. I., & Sullivan, C. V. (2013). Dynamics of the striped bass (*Morone saxatilis*) ovary proteome reveal a complex network of the translasome. *Journal of Proteome Research*, *12*(4), 1691–1699. doi:10.1021/pr3010293

Reith, M., Munholland, J., Kelly, J., Finn, R. N., & Fyhn, H. J. R. (2001). Lipovitellins derived from two forms of vitellogenin are differentially processed during oocyte maturation in haddock (*Melanogrammus aeglefinus*). *Journal of Experimental Zoology*, *291*(1), 58–67. doi:10.1002/jez.5

Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., et al. (2010). GeneCards Version 3: the human gene integrator. *Database*, *2010*(0), baq020–baq020. doi:10.1093/database/baq020

Sawaguchi, S., Kagawa, H., Ohkubo, N., Hiramatsu, N., Sullivan, C. V., & Matsubara, T. (2006). Molecular characterization of three forms of vitellogenin and their yolk protein products during oocyte growth and maturation in red seabream (*Pagrus major*), a marine teleost spawning pelagic eggs. *Molecular Reproduction and Development*, *73*(6), 719–736. doi:10.1002/mrd.20446

Sawaguchi, S., Koya, Y., Yoshizaki, N., Ohkubo, N., Andoh, T., Hiramatsu, N., et al. (2005a). Multiple vitellogenins (Vgs) in mosquitofish (*Gambusia affinis*): identification and characterization of three functional Vg genes and their circulating and yolk protein products. *Biology of Reproduction*, *72*(4), 1045–1060. doi:10.1095/biolreprod.104.037895

Sawaguchi, S., Ohkubo, N., Amano, H., Hiramatsu, N., Hara, A., Sullivan, C. V., & Matsubara, T. (2008). Controlled accumulation of multiple vitellogenins into oocytes during vitellogenesis in the barfin flounder, *Verasper moseri*. *Cybium Int J Ichthyol*, *32*(suppl 2), 262.

Sawaguchi, S., Ohkubo, N., Koya, Y., & Matsubara, T. (2005). Incorporation and utilization of multiple forms of vitellogenin and their derivative yolk proteins during vitellogenesis and embryonic development in the mosquitofish, *Gambusia affinis*. *Zoological Science*, *22*(6), 701–710.

Schilling, J., Nepomuceno, A., Schaff, J. E., Muddiman, D. C., Daniels, H. V., & Reading, B. J. (2014). Compartment proteomics analysis of white perch (*Morone americana*) ovary using support vector machines. *Journal of Proteome Research*, *13*, 1515−1526. doi:10.1021/pr401067g

Wang, H., Tan, J. T. T., Emelyanov, A., Korzh, V., & Gong, Z. (2005). Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene*, *356*, 91–100. doi:10.1016/j.gene.2005.03.041

Wang, H., Yan, T., Tan, J., & Gong, Z. (2000). A zebrafish vitellogenin gene (vg3) encodes a novel vitellogenin without a phosvitin domain and may represent a primitive vertebrate vitellogenin gene. *Gene*.

Williams, V. N., Reading, B. J., Amano, H., Hiramatsu, N., Schilling, J., Salger, S. A., et al. (2014). Proportional accumulation of yolk proteins derived from multiple vitellogenins is precisely regulated during vitellogenesis in striped bass (*Morone saxatilis*). *Journal of Experimental Zoology*, *321*(6), 301–315. doi:10.1002/jez.1859

Williams, V. N., Reading, B. J., Hiramatsu, N., Amano, H., Glassbrook, N., Hara, A., & Sullivan, C. V. (2013). Multiple vitellogenins and product yolk proteins in striped bass, *Morone saxatilis*: molecular characterization and processing during oocyte growth and maturation. *Fish Physiology and Biochemistry*, 1–21. doi:10.1007/s10695-013-9852-0

Zhang, F., Bartels, M. J., Brodeur, J. C., & Woodburn, K. B. (2004). Quantitative measurement of fathead minnow vitellogenin by liquid chromatography combined with tandem mass spectrometry using a signature peptide of vitellogenin. *Environmental Toxicology and Chemistry*, *23*(6), 1408. doi:10.1897/03-425

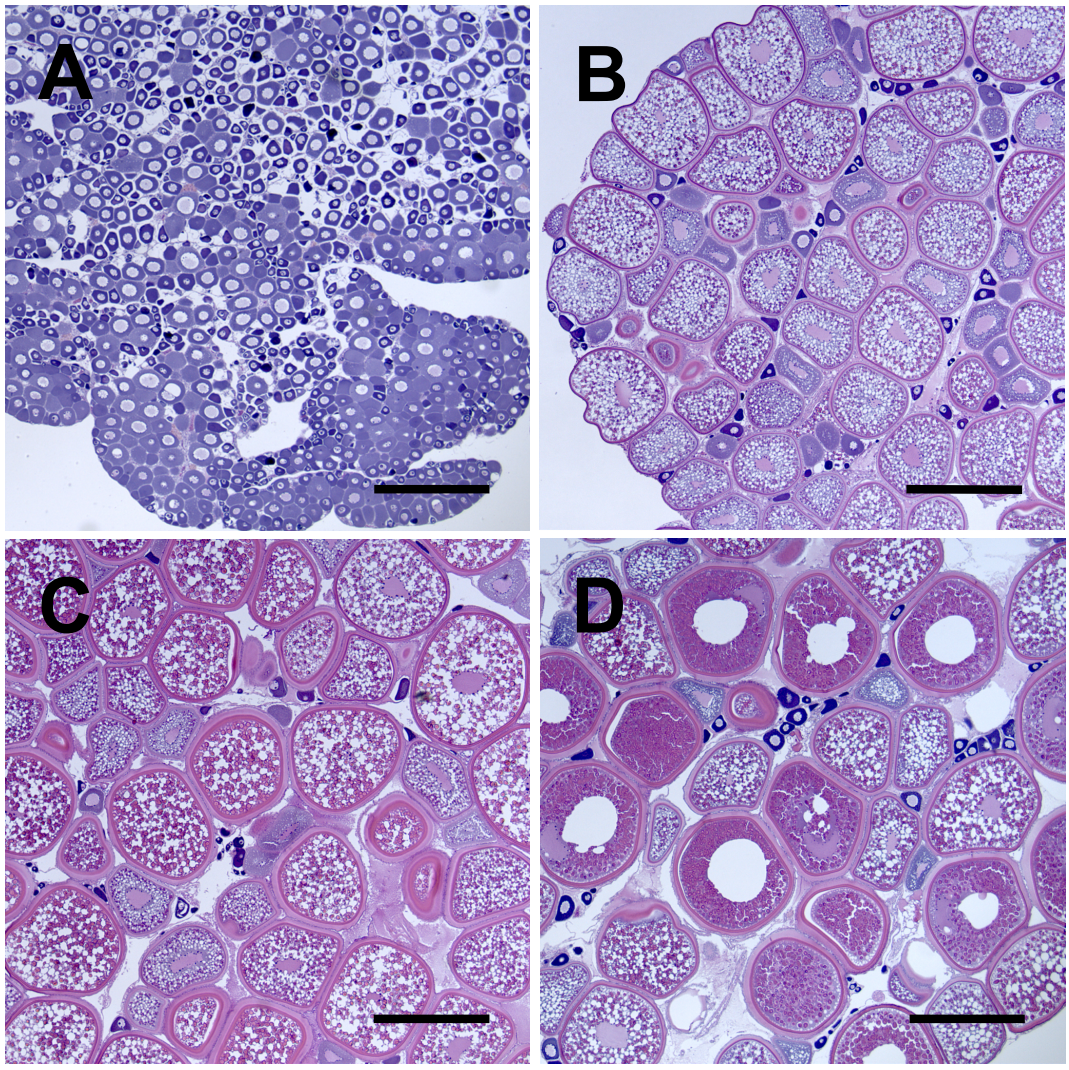


Figure 1. Hematoxylin and eosin staining of representative white perch ovary tissue sections sampled at four key time points across one reproductive year during (A) pre-vitellogenesis, (B) early-vitellogenesis, (C) mid-vitellogenesis, and (D) post-vitellogenesis. Scale bar is 500 microns.

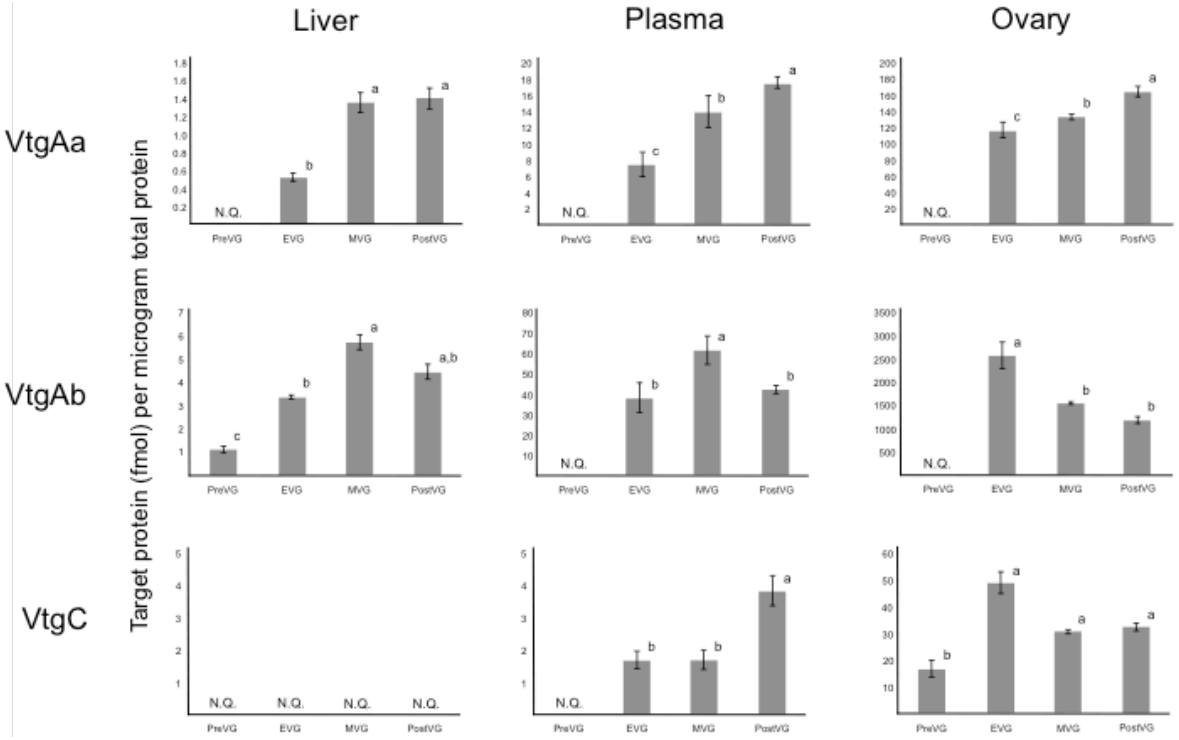


Figure 2. Results from selected reaction monitoring (SRM) tandem mass spectrometry for the three white perch vitellogenins (VtgAa, VtgAb, and VtgC) in female liver, plasma, and ovary tissues sampled across one reproductive year during pre-vitellogenesis (PreVG), early-vitellogenesis (EVG), mid-vitellogenesis (MVG), and post-vitellogenesis (PostVG). The mean ± standard error of the mean is shown. "N.Q." indicates that the native peptide was not quantifiable. Levels not connected by the same letter are significantly different at a = 0.05.

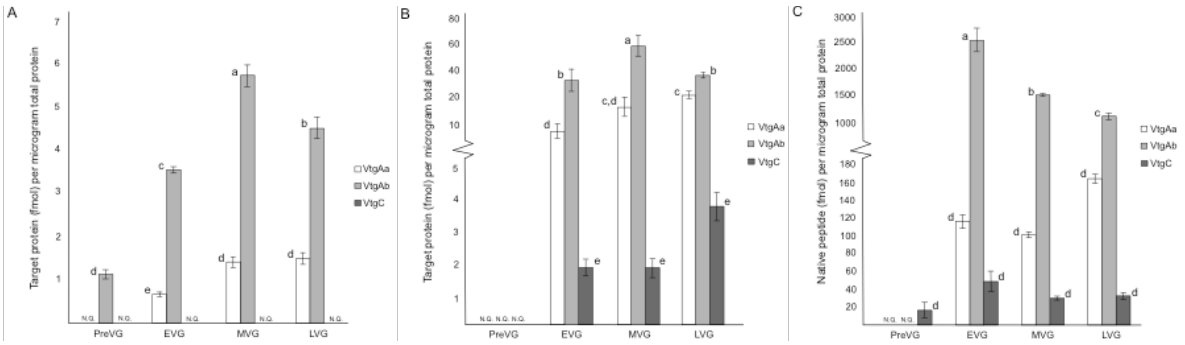


Figure 3. Absolute quantification by selected reaction monitoring of white perch vitellogenins in the A) liver, B) plasma, and C) ovary tissues sampled across one reproductive year during pre-vitellogenesis (PreVG), early-vitellogenesis (EVG), mid-vitellogenesis (MVG), and post-vitellogenesis (PostVG). The mean ± standard error of the mean is shown. "N.Q." indicates that the native peptide was not quantifiable. Levels not connected by the same letter are significantly different at a = 0.05.

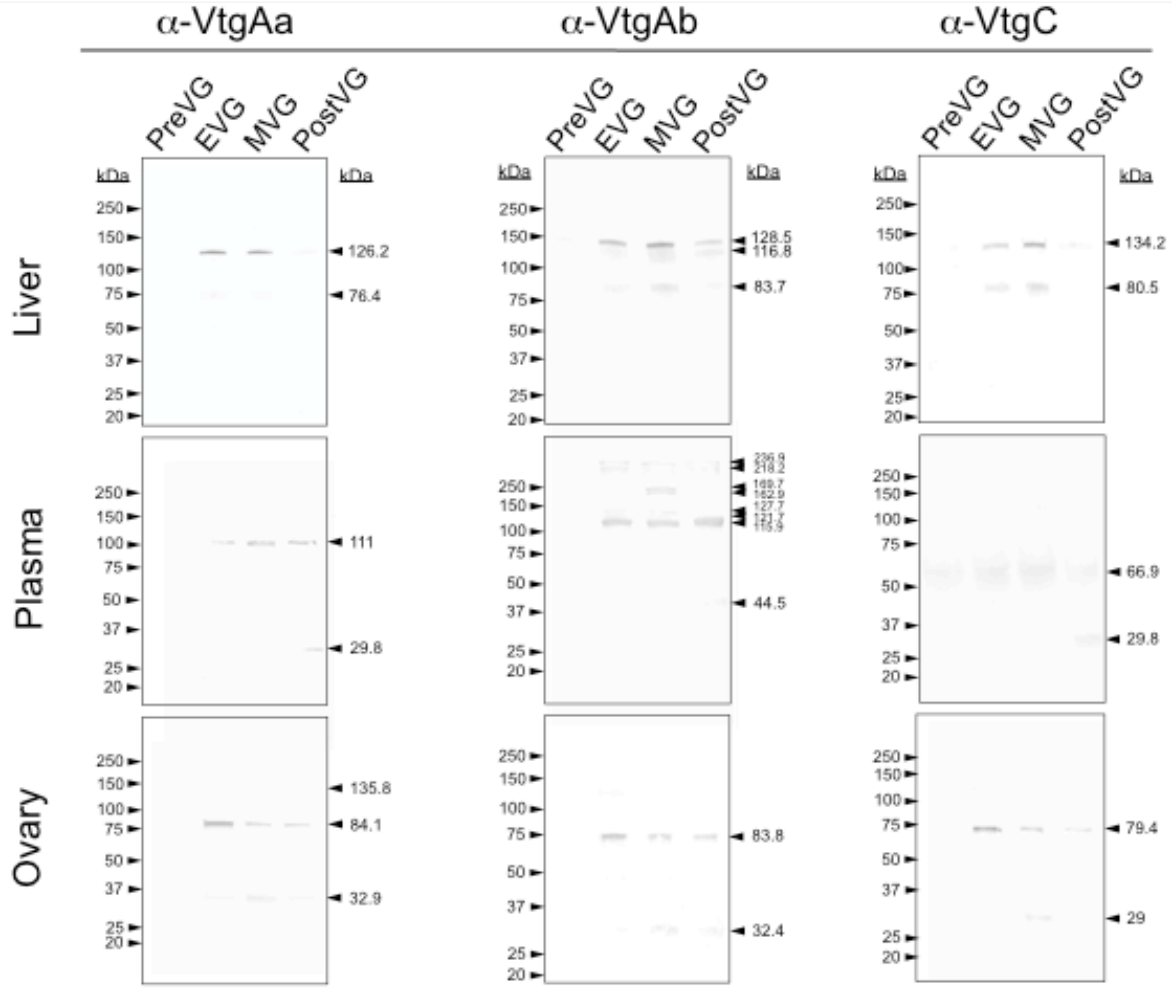


Figure 4. Results of Western blotting for the three white perch vitellogenins in female liver, plasma, and ovary tissues sampled across one reproductive year during pre-vitellogenesis (PreVG), early-vitellogenesis (EVG), mid-vitellogenesis (MVG), and post-vitellogenesis (PostVG).

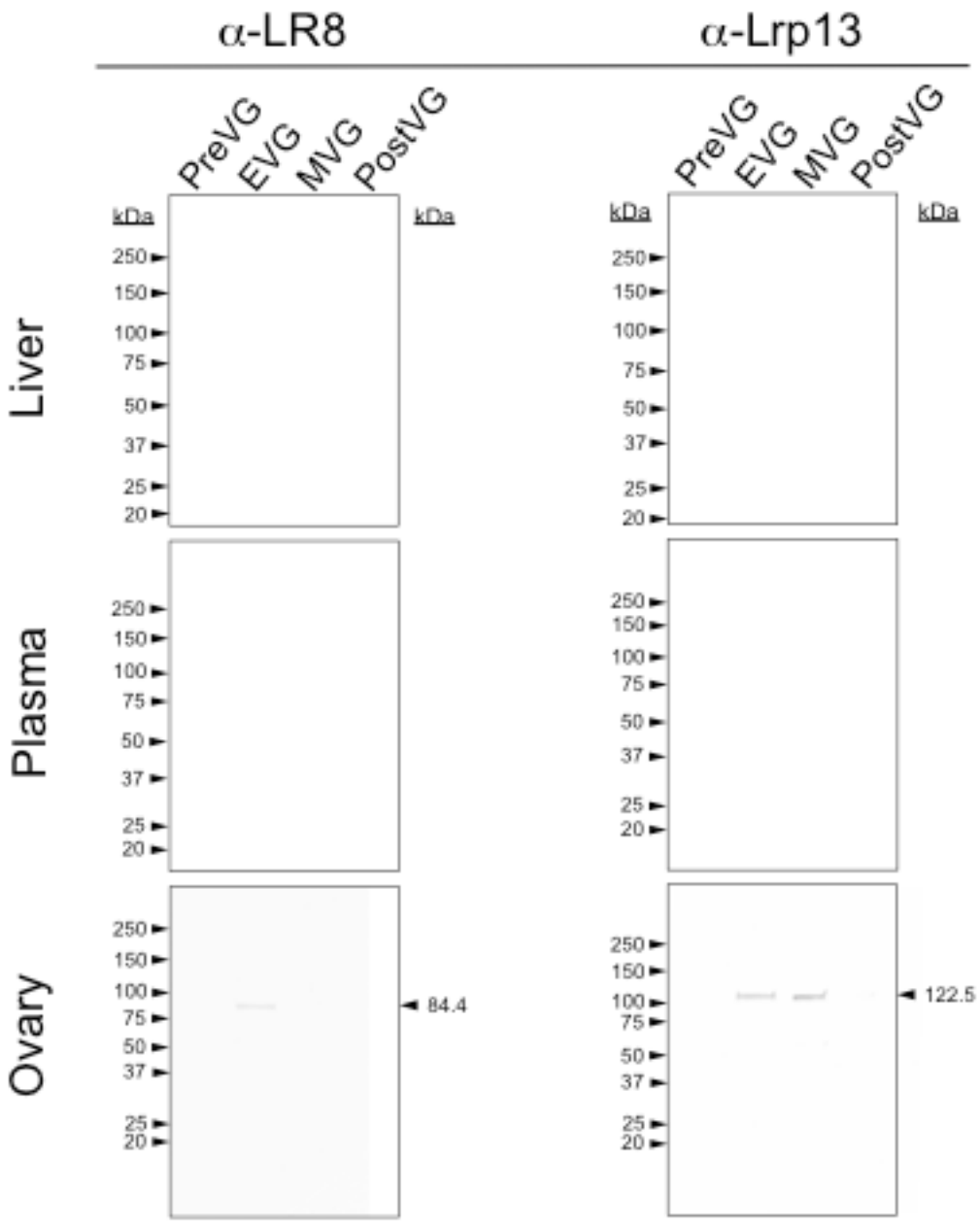


Figure 5. Results of Western blotting for the two white perch vitellogenin receptors in female liver, plasma, and ovary tissues sampled across one reproductive year during pre-vitellogenesis (PreVG), early-vitellogenesis (EVG), mid-vitellogenesis (MVG), and post-vitellogenesis (PostVG).

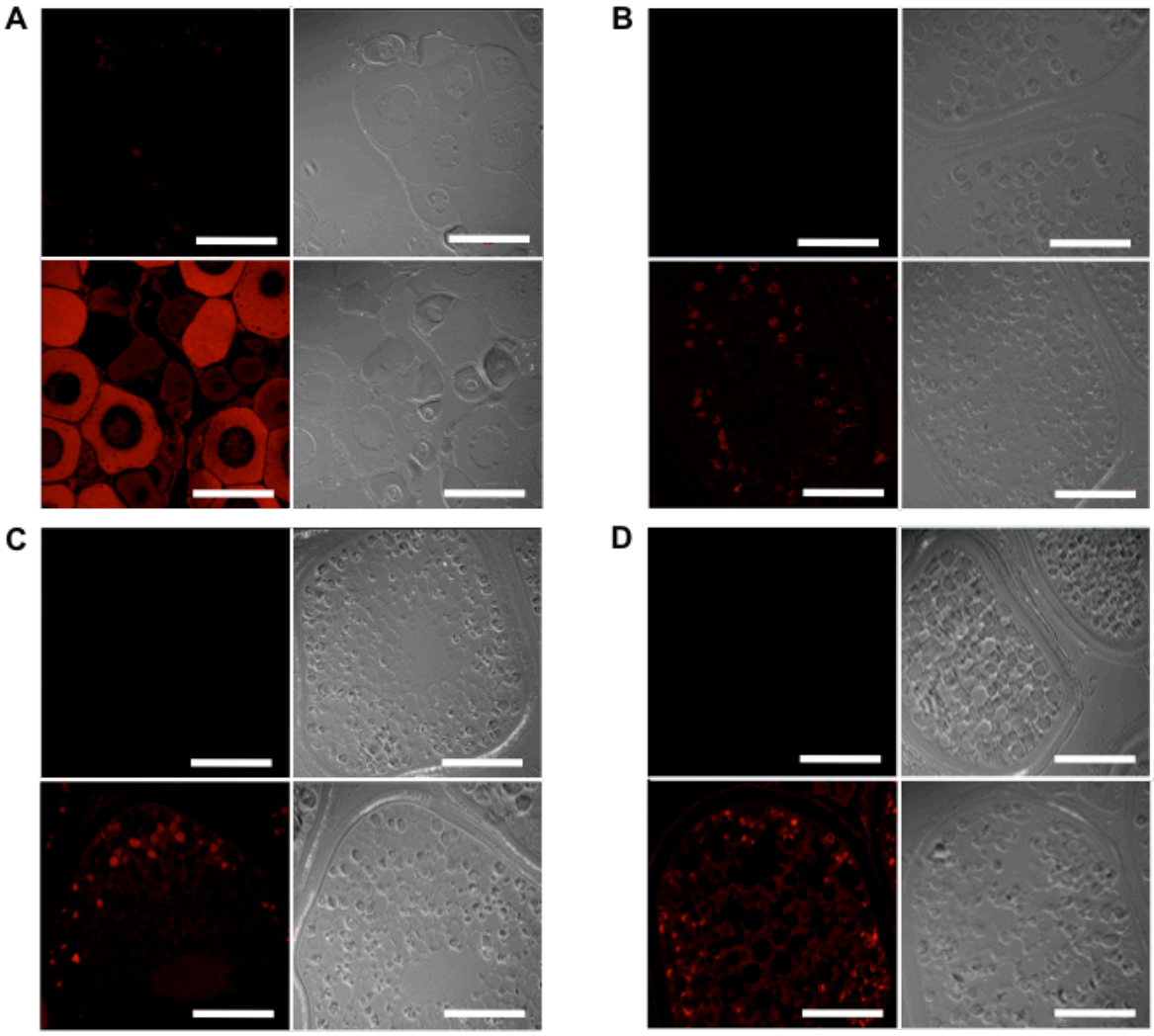


Figure 6. Confocal microscopy images of immunohistochemistry of mature female white perch ovary tissues across one reproductive year stained with anti-VtgC coupled to DyLight633: (A) pre-vitellogenic, (B) early-vitellogenic, (C) mid-vitellogenic, and (D) post-vitellogenic ovary sections.

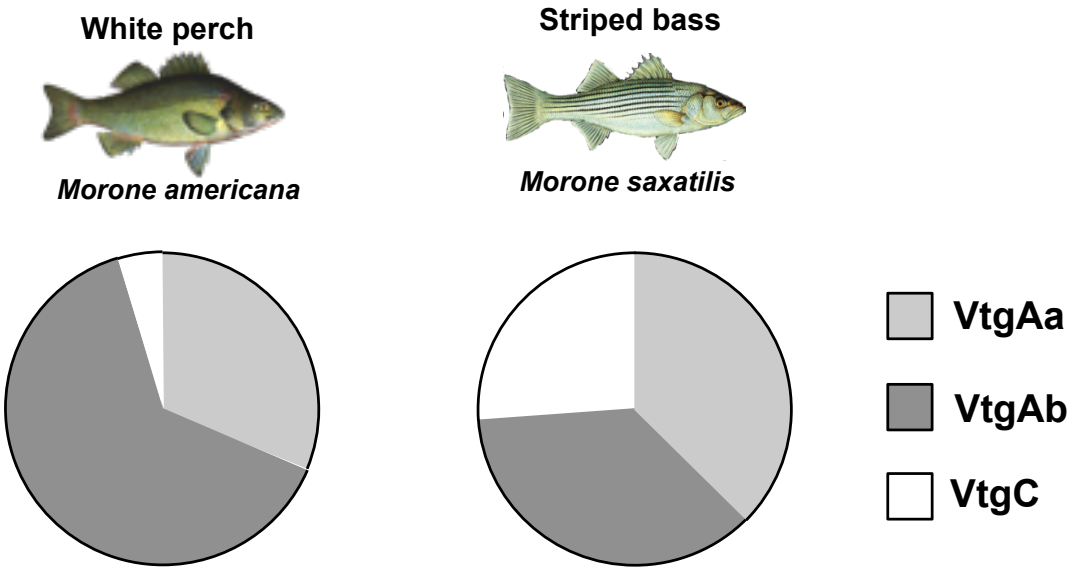


Figure 7. In white perch, yolk proteins derived from VtgC are minor components of the total egg yolk (< 5%), whereas in striped bass they are major components of the egg yolk (~ 25%). [Williams et al., 2014 (J Exp Zool Part A); Schilling et al., 2014 (J Proteome Res)].

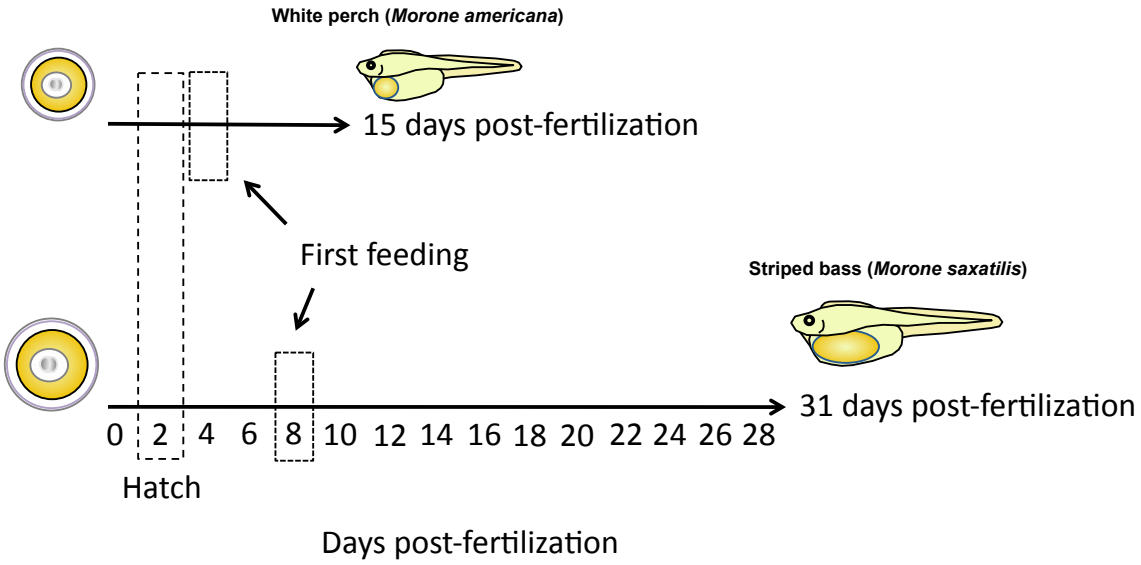


Figure 8. Average survival duration of food-restricted white perch and striped bass larvae. Dashed boxes indicate approximate time of hatching (~2 days) and onset of first feeding (~4 days in white perch, ~8 days in striped bass). [Mansuetti, 1964; Eldridge, et al., 1981; North & Houde, 2003].

Table 1. Sexually mature female white perch sampling month, oocyte stage, weight, and length statistics.

Sampling Month	Oocyte Stage	Average Weight (g) ± STD DEV	Average Length (mm) ± STD DEV
August	PreVG	114 ± 59	188 ± 27
November	EVG	183 ± 42	216 ± 16
February	MVG	201 ± 59	227 ± 10
May	PostVG	182 ± 15	217 ± 7

Table 2. Average amount (femtomoles per microgram total protein) of the three white perch vitellogenin proteins (VtgAa, VtgAb, VtgC) in liver, plasma, and ovary during pre-vitellogenesis (PreVG), early-vitellogenesis (EVG), mid-vitellogenesis (MVG), and post-vitellogenesis (PostVG).

	VtgAa	VtgAb	VtgC	VtgAa/VtgAb/VtgC
Liver				
PreVG	**0±0**	**1.185±0.124**	**0±0**	**0:1:0**
EVG	**0.531±0.047**	**3.463±0.061**	**0±0**	**1:6.5:0**
MVG	**1.372±0.109**	**5.796±0.330**	**0±0**	**1:4.2:0**
PostVG	**1.419±0.118**	**4.499±0.344**	**0±0**	**1:3.2:0**
Plasma				
PreVG	**0±0**	**0±0**	**0±0**	**0:0:0**
EVG	**7.557±1.489**	**38.454±7.480**	**1.705±0.271**	**4.4:22.6:1**
MVG	**13.956±1.984**	**61.757±7.124**	**1.721±0.307**	**8.1:36.9:1**
PostVG	**17.529±0.757**	**42.488±2.023**	**3.847±0.475**	**4.6:11:1**
Ovary				
PreVG	**0±0**	**0±0**	**17.005±3.209**	**0:0:1**
EVG	**116.650±9.841**	**287.969±49.212**	**49.212±4.052**	**2.4:5.9:1**
MVG	**134.883±3.870**	**1557.167±27.513**	**30.949±0.790**	**4.4:50.3:1**
PostVG	**165.617±6.815**	**1173.956±65.035**	**32.870±1.412**	**5:35.7:1**

Table 3. Results from affinity purification coupled to tandem mass spectrometry using purified white perch VtgC as a bait protein.

Elution Buffer	Bait Protein	Striped Bass Ovary Transcriptome ID	Protein ID	Protein MW (kDa)	Sequence Coverage (%)	# Unique Peptides	Peak Area
Suramin	VtgC	Contig00578	Y-box binding protein 2a-like	37.9	5.56	2	9.75E6
	Blank				0.00	0	0.00E0

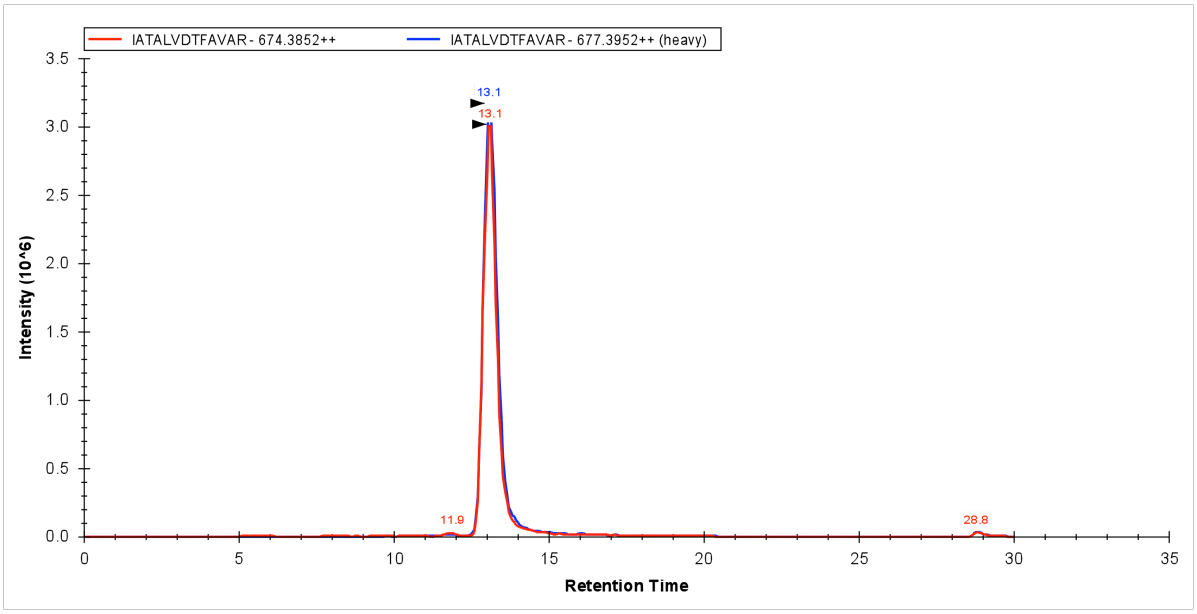
Table 4. White perch vitellogenin gene names, protein identifications, Uniprot accession numbers, peptide sequences, selected reaction monitoring transitions, and collision energies.

				SRM Transitions		
Gene Name	Protein ID	Uniprot Accession #	Peptide Sequence	Q1 *(m/z)*	Q3 *(m/z)*	Collision Energy (eV)
VtgAa	Vitellogenin Aa	A5GXQ1_MORAM	TEGLQEALLK	551.3111 (light)	871.5247	21
					701.4192	
					573.3606	
					231.0975	
					288.119	
				554.3212 (heavy)	877.5449	21
					707.4393	
					579.3808	
					231.0975	
					288.119	
VtgAb	Vitellogenin Ab	A5GXQ2_MORAM	IATALDTFAVAR	674.3852 (light)	991.5571	24
					878.473	
					779.4046	
					664.3777	
					185.1285	
				677.3952 (heavy)	997.5772	24
					884.4932	
					785.4248	
					670.3978	
					185.1285	
VtgC	Vitellogenin C	A5GXQ3_MORAM	YFQATTLGLPLEISK	840.964 (light)	1070.6456	27
					969.5979	
					856.5138	
					686.4083	
					283.1441	
				843.9741 (heavy)	1076.6657	27
					975.618	
					862.534	
					692.4284	
					283.1441	

Supplemental Table 2. Antigenic peptides for white perch vitellogenin Aa, vitellogenin Ab, vitellogenin C, LR8, and Lrp13 used for antibody production.

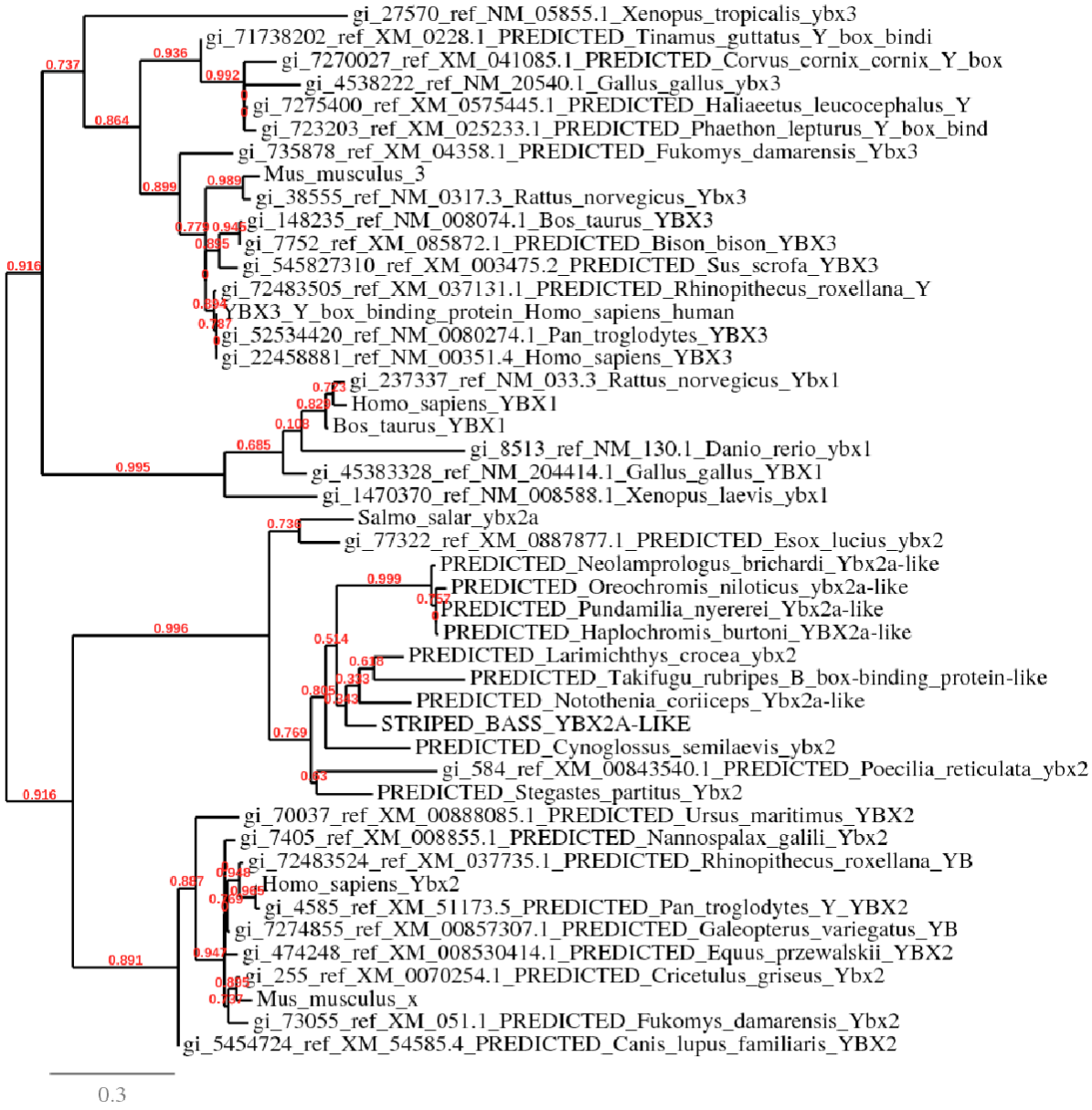
Target Protein	Antigenic Peptide Sequence	Amino Acid Position within Target Protein
VtgAa	CAKCQQDSKNLR	204-215
	DVAVKPIIKSKAAQFEKKYC	963-982
VtgAb	LPRNIASKLKLPKAFQKKMC	970-989
	EILEDKNMLMKL	1063-1074
VtgC	CVESELRREYYHEEYP	977-992
	RLPSDSASSVRGSPNNHHHHL	1053-1072
LR8	CSVDLNGDNRKKVLQS	624-639
	CRPEANVSTSIQVDSTARGSA	748-768
Lrp13	TTLNESSQLRNLATQDC	1085-1101
	CSLGYSGDSCQDHLLKT	1123-1139

Supplemental Figure 1



Supplemental Figure 1. Extraction ion chromatogram depicting the co-elution of heavy (blue) and light (red) VtgAb peptides from post-vitellogenic white perch ovary tissues.

Supplemental Figure 2



Supplemental Figure 2. ClustalW dendrogram showing relationships between y-box-binding protein family polypeptide sequences. GenBank accession numbers are provided. Numbers above each branch are p-distances.

CHAPTER 5

Future Directions

Molecular basis of Vtg endocytosis and compartmentalization

While we were able to trace the sub-cellular localization of the incomplete type vitellogenin, VtgC, in white perch oocytes, the central question of how the complete type Vtgs are compartmentalized remains unanswered. I tired to answer this question by having the cytoplasmic tails of LR8 and Lrp13, unphosphorylated and phosphorylated, synthesized by New England Peptide. After months of failing to synthesize full-length Lrp13 and failing to phosphorylate LR8 as requested, we elected to cancel the order. It is interesting that these cytoplasmic tails were so difficult to synthesize and phosphorylate and there are undoubtedly biological implications embedded in these facts.

Tracing Vtg compartmentalization and trafficking

I had been planning to use the LR8 and Lrp13 synthetic cytoplasmic tails to 'fish' specific interacting proteins out of soluble ovary preparations. In these AP-MS/MS experiments, I planned to identify the specific interacting proteins by tandem mass spectrometry, as I did in the VtgC AP-MS/MS experiment detailed in Chapter 4. With this approach not longer viable at this time, one way I might try to visualize VtgAa and VtgAb localization within growing oocytes would be to:

1) Purify VtgAa and VtgAb using our established protocols involving anion exchange and size exclusion chromatography.

2) Label the purified VtgAa and VtgAb with distinct fluorophores with no spectral overlap.

3) Add the purified, labeled Vtgs to oocyte cultures for which we have established protocols.

4) Image the oocytes as the fluorescently labeled Vtgs are endocytosed.

While such an experiment would not elucidate the molecular identities of the specific proteins and cellular machinery that are involved in LR8/VtgAb or Lrp13/VtgAa endocytosis and subsequent compartmentalization, we might be able to determine whether or not VtgAa and VtgAb are separately compartmentalized. Tracing the fluorescence further into oocyte maturation may allow us to see whether later movements of these Vtgs share common or distinct patterns. Given that we know VtgAa is extensively proteolyzed in free amino acids and peptides while VtgAb undergoes much less proteolysis, it would not be surprising to see these Vtgs compartmentalized separately.

Since I have established a protocol for purifying VtgC to near homogeneity, the experiment outlined above might provide additional clues to how this protein enters the oocyte and possibly provide additional evidence to confirm the patterns of VtgC localization within growing oocytes that were found by IHC as described in Chapter 4.

Machine learning analysis of RNASeq data from striped bass 4 hr post-fertilization viable and inviable embryos

While not included in this dissertation, the transcriptomic data from viable and inviable 4hr post-fertilization striped bass embryos has emerged following extensive bioinformatics efforts. We made three comparisons to attempt to elucidate whether differences in maternal transcripts might underlie differences in embryo viability and whether or not there are transcriptome differences in viable or inviable embryos from females of high (>20%) fertility versus low (<5%) fertility. Initial parametric analyses (ANOVA) only showed significant (a = 0.05) differences in ~178 genes (out of ~24,500) between inviable embryos from high and low fertility females. Our next step with these data will be to apply various forms of machine learning to reveal differences in the global patterns within the transcriptomes of viable and inviable embryos.

Male and female white bass (*Morone chrysops*) genome sequences

We have separately sequenced the genomic DNA of a male and a female white bass and they are nearly fully annotated. For the female white bass, there were a total of 57,533 contigs with 643,931,133 bases and a total GC content of 39.5%. Cegma indicates it is 84.68% complete and 97.98% partially complete, which is very good. For the male white bass, the assembly was very similar with a total of 644,366,547 bases in 56,818 contigs with a CG content of 39.5%. Cegma indicates this assembly is 82.66% complete, but also 97.98% partially complete, which is very good. By comparison, the striped bass genome, which is on the cusp of publication, has 35,010 contigs, and 585,184,717 bases with 40% GC content.

Striped bass and white bass are the progenitors of the hybrid striped bass. With both parental genomes in hand, we can begin to ask some fascinating evolutionary questions, such as:

1) Are there sex-specific differences in genome organization between the male and female white bass?

2) How do the parental genomes recombine in the hybrid offspring?

3) Can we see patterns of gene expression in hybrid offspring that could explain differences in phenotype? For example, if we select slow growing and fast growing hybrid offspring and perform RNAseq on their tissues, can we see patterns in gene expression that could explain differences in growth rate? If so, how do these genes relate to the parental genomes?

4) What changes have occurred in the immune genes of these two closely related species over the time that they have been reproductively isolated? Moreover, can we use the genomic sequences to more accurately date this divergence?

5) Given that white bass are found exclusively in fresh water while striped bass are anadromous, spawning in fresh water and living primarily in salt water, what genomic differences exist between white bass and striped bass that may be adaptive to their respective life histories?

Summary and concluding remarks

As is so often the case, the more closely we examine a biological system, the more we uncover unanticipated complexities. This has certainly been the case in our research into the Vtg-Vtgr system of the genus *Morone* with the recent descriptions of:

1) Lrp13 and the disparate binding affinities of the complete type Vtgs, VtgAa and VtgAb, for LR8 and Lrp13.

2) The presence of VtgC in the pre-vitellogenic perch ovary.

3) The localization of VtgC exclusively to lipid droplets within vitellogenic perch oocytes.

4) The distinctly different proportional accumulation of Vtgs between the closely related striped bass and white perch.

5) The apparent plasticity of the Vtg-Vtgr system in white perch, which seems to correlate to water salinity during vitellogenesis.

As additional data become available, I anticipate that the plasticity of the Vtg-Vtgr in response to salinity will become more apparent, particularly among fishes that spawn in estuarine waters that are susceptible to large, abrupt changes in salinity due primarily to dramatic wind and rainfall (or, conversely, drought) events.

If all Vtg types within a species can be purified to homogeneity with minimal loss of cargoes, I would also anticipate that rapidly advancing lipidomic analysis would reveal that each Vtg carries a distinct lipid cargo. While I might not anticipate striking differences between the lipid cargoes of complete type Vtgs, I would expect there to be considerable differences between the lipid cargos of complete and incomplete type Vtgs.

Since we have shown in perch that VtgC behaves more similarly to lipoproteins (i.e. LDL and VLDL) in that it steadily enters growing oocytes from pre-vitellogenesis onward and is not particularly responsive to estrogen induction, I would expect VtgC to carry a larger proportion of neutral lipids than the complete type Vtgs.

Confirmation of this would make sense in light of our finding and discussion in Chapter 4 of the striking differences in the amount of VtgC in post-vitellogenic perch and striped bass oocytes. In particular, as Sawaguchi et al. (2005) have reported, it appears that VtgC serves as a late stage food source for larvae before they are able to feed on their own. Striped bass post-vitellogenic oocyte yolk contains ~26% VtgC, while perch yolk contains just 2.5-5% VtgC, which makes sense as striped bass larvae can live for up to 31 days when food is restricted or deprived, while perch larvae live just 15 days. This differential apportioning of yolk nutrients between these two closely related species makes sense as striped bass spawn in freshwater, far from the plankton bloom that the developing larvae will need to reach when capable of exogenous feeding ~ 8 days post-fertilization. White perch, on the other hand, spawn their adhesive eggs in brackish waters that are typically lower in the estuary and closer to the plankton bloom. As such, their larvae may need less time and less food to reach the bloom.