Using global genome approaches to address problems in cod mariculture*

Frederick W. Goetz[1], Linda McCauley[2], Giles W. Goetz[1], Birgitta Norberg[3]

[1] Great Lakes WATER Institute
University of Wisconsin Milwaukee
600 E. Greenfield Ave.
Milwaukee, WI 53204
rick@uwm.edu

[2]Woods Hole Oceanographic Institution
Biology Department
Woods Hole, MA 02543
lmccauley@whoi.edu

[3] Institute of Marine Research
Austevoll Aquaculture Research Station
N–5392 Storebo
Norway
birgitta.norberg@imr.no

Correspondence should be directed to:

Frederick William Goetz
Great Lakes WATER Institute
University of Wisconsin Milwaukee
600 E. Greenfield Ave.
Milwaukee, WI 53204
414-382-1700 (phone); 414-382-1705 (fax)
rick@uwm.edu

* This work was supported in part by grant #139630/140 from the Research Council of Norway to BN and USDA grant #2004-35204-14232 to FWG.

Abstract

A number of techniques including expressed sequence tag (EST) analysis, serial analysis of gene expression, and microarrays are available to study the global expression and regulation of genes. Many of these techniques are being used for intensively reared fish such as trout, salmon and catfish to study genes involved in growth, reproduction and health. In contrast, relatively little is known about the composition and regulation of transcriptomes in gadids. However, several bottlenecks in cod mariculture might benefit from the discovery and analysis of genes involved in reproduction, growth and disease. As a result, we have begun EST analysis of genes in the cod ovary. Complimentary DNA (cDNA) libraries of cod ovaries taken from females at oocyte final maturation and ovulation have been constructed, and 1,361 ESTs have been analyzed. As expected, several oocyte-related genes were observed including various zona pellucida egg membrane proteins. However, pivotal cell cycle regulators such as cyclins, genes involved in the regulation of apoptosis such as the Bcl-2-related ovarian killer protein, and hormone receptor components were also observed. Finally, a cDNA for a potential novel cod antifreeze protein was observed 12 times, suggesting the existence of a cod egg-specific antifreeze protein.

1. Introduction

Worldwide decreases in wild stocks of cod have led to enhanced interest in the cultivation of this species (Brown *et al.*, 2003). While many aspects of cod mariculture have been addressed, there are still issues that will be pivotal for the successful economic development of cod culture including the enhancement of growth, resistance to disease, and gamete quality. Many of these issues will require the selection of desirable traits in brood stocks or an understanding of the basic physiology and genetics of cod. Genetic selection will require a good understanding of the cod genome. In certain intensively reared fish groups such as salmonids, there has been significant progress in developing genomic tools such as complimentary DNA (cDNA) and bacterial artificial chromosome (BAC) libraries, expressed sequence tags (ESTs), microarrays and genetic maps (Nichols *et al.*, 2003; Rexroad *et al.*, 2003; Rise *et al.*, 2004). However, with the exception of mitochondrial sequencing for phylogenetic investigations (Johansen and Bakke, 1996; Sigurgislason and Arnason, 2003), there have been virtually no comparable tools developed for cod.

Several aspects of reproduction, including egg quality, out-of-season spawning, and the effects of gonadal development (e.g., puberty) on somatic growth are important issues for successful cod mariculture. While cod reproduce seasonally, females can theoretically spawn multiple batches of eggs during the yearly reproductive period. For example, in captive cod, females shed from 1 to 19 batches of eggs depending on the individual (Kjesbu, 1989). In addition, wild stocks of cod living in very similar temperate regions can spawn at very different times of the year. To acquire a more complete understanding of cod reproduction, we have been studying the endocrine regulation of the gonad. For those investigations we first developed

stage-specific cDNA libraries of the cod ovary and used them to isolate and characterize genes involved in the regulation of steroid synthesis (Goetz *et al.*, 2004). Here we describe the use of these libraries to develop ESTs for cod reproduction. Expressed sequence tags are the partial sequencing and analysis of clones obtained from the random plating of cDNA libraries. Since cDNA libraries are composed of the RNA transcripts expressed in a given tissue, then ESTs represent the sequences of these transcripts; hence the name "expressed." For organisms in which gene sequence information is lacking, the analysis of ESTs is particularly useful to provide an initial survey of genes expressed in a given tissue. That information can then be used for more directed studies of gene expression (e.g., candidate gene approaches) or for more global expression analysis (e.g., cDNA microarrays).

2. Methods and Methods

2.1. Animals

Larvae of coastal Atlantic cod were hatched in 1999 (winter) and reared in a seawater pond (Parisvatnet) located outside Bergen. The cod were transported to Austevoll (Norway) in September 1999 and kept in indoor tanks under a simulated natural photoperiod until May the following year when they were transferred to an outdoor, 5 x 5 m cage. The fish were fed *ad libitum* and kept in the sea cage under a natural photoperiod until sampling. From these stocks, mature, two-year-old female cod were sacrificed in February 2001 and ovarian samples were collected. Fish were sampled at stage IV (late vitellogenesis), V (spawning) and VI (spent), defined according to a previously reported classification (Kjesbu, 1991). Cod are periodic (batch) spawners that may undergo 15-20 ovulatory cycles during their spawning period

(Kjesbu, 1989; Kjesbu *et al.*, 1990). Thus, the fish in stage V were further divided into 3 sub-stages: 1) females sampled early in the spawning period containing ovaries with oocytes undergoing final oocyte maturation (V1); 2) ovulating fish (V2); and 3) females sampled in the mid to late stages of the spawning period with ovaries containing oocytes undergoing final oocyte maturation (V3). All ovaries were snap frozen on dry ice and maintained frozen at -80^{o}C prior to RNA extraction.

2.2. RNA Extraction and Library Construction

Ovarian tissue from several reproductive stages was homogenized with a TissueTearor (Biospec) and RNA was extracted using Tri Reagent (Molecular Research Center, Inc.) at a ratio of 100 mg tissue/1.0 ml of reagent as previously described (Chomcynski, 1987; Chomcynski, 1993). Poly(A+) RNA was isolated using the PolyAtract mRNA Isolation System (Promega). Messenger ribonucleic acid (5.0 ug) derived from 4-5 females/reproductive stage, was used to construct individual cDNA libraries of each reproductive stage (I-V) in Zap Express (Stratagene). Libraries obtained from stages II, III and IV, were mass excised to pBK-CMV phagemids and plated at low density. Individual colonies were randomly picked and plasmid preparations made using the RevPrep Orbit (GeneMachines). Plasmid preparations were sequenced from the 5' end using the dideoxy chain termination method with "Big Dye Terminator" (Applied Biosystems) and the BK reverse vector primer. The reactions were precipitated and resuspended in "Hi-Di Formamide with EDTA" (Applied Biosystems) and run on an ABI Prism 3730 automated sequencer (Applied Biosystems).

2.3. Sequence Data Analysis

Sequence chromatogram files were trimmed for quality using phred (http://www.phrap.org/phrap.docs/phred.html), vector screened using cross match (http://www.phrap.org/phrap.docs/phrap.html) and analyzed locally using 1) blastx against the NCBI nonredundant (nr) protein database, 2) blastn against the NCBI nucleotide (nt) database and 3) blastn against the NCBI EST (dbEST) database. In general, "blast" analyses can be used to compare the translated and/or nucleotide sequences of an unknown sequence (e.g., an EST) against known NCBI sequences to provide an initial identification based on the degree of identity. Specifically, the blastx analysis provides an amino acid sequence comparison of each cod EST translated in 6 reading frames against the NCBI nonredundant protein database. The blastn analysis against the nt database compares the nucleotide sequence of each cod EST against all nucleotide sequences at NCBI, and the blastn analysis against the NCBI EST database compares the nucleotide sequence of all cod ESTs against the entire NCBI EST database. All sequences were grouped by category (GO database) and tentative identification was based initially on a blastx similarity score of $<10^{-3}$ or, in the case of blastx scores of $>10^{-3}$, a blastn score of $<10^{-5}$. All sequences were collectively analyzed for redundancy using CAP3 (Huang and Madan, 1999). Sequences for all of the cod ESTs are currently available at NCBI (accession numbers CO542794 - CO541452, AY70611-4, AY584595).

3. Results and Discussion

3.1. General

Of the 1,417 clones initially sequenced from the cod ovarian libraries, there were 1,361 sequences ≥ 100 bp that were annotated and reported here. The average sequence length was 645 bp and when analyzed for redundancy by CAP3, there were 208 contigs (of 2 or greater) and 659 singletons. As might be expected, a significant number of sequences were homologous to genes encoding proteins involved in metabolism, and ribosomal and structural proteins (Table 1). In addition, approximately 14 % of the ESTs coded for peptides that could be grouped as "Regulatory Proteins" involved in the control of cell division, transcription, translation, signaling, apoptosis, and development (Table 1).

3.2. Structural Proteins

A major group of structural genes coding for egg membrane proteins was observed; there were 18 ESTs for egg proteins that, based on contig analysis, theoretically represent 5 unique sequences aligning to choriogenin L (CO542190), ZPC2 (CO542098), ZPAX (CO542010), ZPC (CO542603) and ZPA (CO542593). In fish, the nomenclature for these egg proteins is inconsistent but they are extremely important since they constitute the inner portion or zona radiata of the egg envelope (Arukwe and Goksoyr, 2003). Other structural genes observed several times were beta 1 tubulin (CO541590) and a profiling-2 like EST (CO541946). ESTs (AY584595) for a protein that had greatest similarity with an antifreeze protein described from the longhorn sculpin (Zhao *et al.*, 1998) were observed 12 times in the ovarian libraries (Table 2). There are several classes of naturally occurring antifreeze proteins that act to depress the temperature that ice grows in a noncolligative manner, thereby exhibiting what is termed thermal hysteresis. In gadids the antifreeze proteins that have been reported are representative of a group

called “antifreeze glycoproteins” (Harding et al., 2003). These glycoproteins are distinct from the type of antifreeze proteins reported in the longhorn sculpin and the protein presumably encoded by the ESTs observed in the cod ovarian cDNA libraries. However, the cod ovarian protein also has sequence similarity with apolipoproteins so it is unclear if it acts as an antifreeze protein or has other functions in the egg/ovary.

3.3. Cell Division/Cell Cycle Regulators

Since the cod ovaries that were used for library construction were taken from fish at several stages of oocyte maturation and ovulation, we expected to see genes encoding proteins that are involved with the control of meiosis and cell division. Prominent among these proteins were cyclin B1 (CO541540) and B2 (CO541792), each observed several times (Table 2). Cyclins are factors expressed in cells that modulate the kinase activity of cyclin-dependent kinases, directing the progression through the cell cycle (Obaya and Sedivy, 2002). They are unified by the presence of a “cyclin box” in their amino acid sequence. Cyclin B and the Cdc2 kinase make up the maturational promoting factor (MPF) that induces oocyte final maturation and the resumption of meiosis (Kishimoto, 2003). Specifically, MPF acts at the G2/M transition of the cell cycle. In zebrafish, cyclin B1 was expressed in all tissues observed (including the ovary), but cyclin B2 expression was only observed in the ovary and testes, consistent with a role in the control of meiosis (Bauer, 2001). It is likely that one of the cyclin B sequences observed in the ESTs represent the cyclin component of the MPF in the cod ovary and, therefore, is of considerable importance.

Other ESTs that potentially are involved in regulating the cell cycle include a Wee1-like cDNA (CO542429), one of the kinases that phosphorylates the Cdc2 kinase of MPF (Mueller *et al.*, 1995); Pin 1 (CO541689), an enzyme required in mammals for mitosis progression (Fujimori *et al.*, 1999) and recently shown to regulate primordial germ cell proliferation in mice (Atchison *et al.*, 2003); and several genes encoding proteins associated with mitotic spindles including the Protein Regulator of Cytokinesis (PRC1; CO542004) (Jiang *et al.*, 1998) and a mitotic spindle assembly checkpoint protein homolog (MAD2B; CO542663) (van den Hurk *et al.*, 2004).

3.4. Cell Signaling/Apoptosis

ESTs for a number of gene involved in apoptosis and cell proliferation were also identified. These included several nucleoside diphosphate kinases (CO542766 & CO542513) that are biochemically responsible for the synthesis of nucleoside triphosphates but also have been implicated in a number of regulatory roles including oncogenic transformation and metastasis (Roymans *et al.*, 2002). Other ESTs encoding proteins theoretically involved with apoptosis were a cDNA (CO542330) encoding a novel Bcl-2 family member called Bok (Bcl-2 related ovarian killer) that has restricted expression to reproductive tissues and is pro-apopotic (Hsu *et al.*, 1997); an EST (CO542111) similar to SARPs (secreted apoptosis related proteins) that are anti-apoptotic (Melkonyan *et al.*, 1997); an EST (CO542251) encoding a protein with similarity to caspases (Uren *et al.*, 2000); and a cDNA (CO541487) for BID (BH3 interacting death domain agonist) that promotes cell death (Wang *et al.*, 1996).

3.5. Developmental Regulators

Since maternal oocyte mRNA undoubtedly made up a large portion of the transcripts in the ovarian cDNA libraries, we also expected to see a number of genes that have already been specifically implicated in the regulation of embryonic development. These included genes such as derriere (CO542177), a TGF-beta superfamily member involved in directing the fate of posterior development (Sun *et al.*, 1999) and geminin (CO542557), a protein that plays an early role in establishing the neural plate during gastrulation (Kroll *et al.*, 1998). However, there were also several genes found in high copy number that have only recently been established in the regulation of development. These included cDNAs (CO541612) with strong similarity to the zygote arrest 1 gene. The 7 ESTs of this gene could be separated into two distinct contigs. In mice and humans the zygote arrest gene is oocyte-specific and in mice lacking this gene, embryos are arrested at the one cell stage (Wu *et al.*, 2003a). In other species it may not be specific to the oocyte (Wu *et al.*, 2003b). Several copies of ESTs for stathmin (CO542152), that appears to play a pivotal role in early neurogenesis directing the early migration of neurons in the brain (Jin *et al.*, 2004) were observed. While stathmin is preferentially expressed in the brain, it also appears to be involved in implantation of the embryo in the uterus (Tamura *et al.*, 2003).

3.6. Others

Given the extensive hormonal control of the ovary, we expected to see ESTs for hormone receptors or for genes involved in hormone synthesis. Guanine nucleotide-binding proteins (G-proteins) are involved in the signal transduction of stimuli directed through receptors present on

the outer surface of a cell membrane. They are composed of three subunits; alpha, beta and gamma. Cod homologs of all three subunits (gamma: CO541764, beta: CO542763, alpha: CO541729) of these G-proteins were present in the ESTs (Table 2). An EST (CO542479) for the cAMP response element binding (CREB) protein transcription factor that binds to CRE elements in DNA and modulates gene transcription under cAMP stimulation was observed several times. The cod homolog (CO541987) of the regulatory subunit of AMP activated protein kinase was also identified.

There were several ESTs that had sequence similarity to genes encoding cytochrome P450 enzymes, a superfamily of proteins that are involved in the oxidative, peroxidative, and reductive metabolism of many compounds, including steroids, bile acids, fatty acids, prostaglandins, and xenobiotics (Hukkanen, 2000). Thus, some P450 enzymes are involved in steroid hormone synthesis in the gonads or other steroidogenic tissues (e.g., CYP11 & CPY17), while others are involved in the metabolism of environmental pollutants (e.g., CYP1A). Based on amino acid sequence similarity, one of these, cytochrome P450 11A1 (AY706102), is most likely the cod homolog of the cholesterol side-chain cleavage enzyme. Another EST (AY706101) is very similar in structure to a novel 3-beta hydroxysteroid dehydrogenase (3-beta HSD) that was characterized as the causative gene in several mouse mutations that involve abnormalities in cholesterol synthesis (Liu *et al.*, 1999). Whether the EST is truly a novel form of 3-beta HSD or the actual form responsible for the conversion of 3-beta-hydroxy-5-ene steroids into 3-keto-4-ene steroids, will require complete sequence analysis. ESTs (AY706104 & AY706103) for two other cytochrome P450 enzymes of unknown function were also observed.

Finally, ESTs for novel transcripts encoding proteins that might have a specific involvement with changes in the egg at oocyte maturation/ovulation/fertilization were observed (Table 2). These include an EST (CO542550) for hyosophorin, a protein that has been been shown to be cleaved and exocytosed into the perivitelline space during the cortical reaction at fertilization (Tsao *et al.*, 1999); an EST (CO542059) for alveolin, a metalloproteinase that is believed to induce chorion hardening at fertilization in medaka (Shibata *et al.*, 2000); and an EST (CO542075) for a transcript encoding a protein that has been shown to regulate chloride conductance through the membrane (Krapivinsky *et al.*, 1994) and, therefore, may be involved in oocyte volume regulation.

4. Conclusion

To our knowledge the ESTs reported here are the first to be submitted to Genbank for cod. Even though there were relatively few ESTs in this project compared to other surveys, several pivotal ovarian genes such as the B cyclins and steroidogenic enzymes have already been uncovered that can be used in the future for research on the endocrine control of reproduction in cod. When we initiated this EST project, we did not consider the proportion of mRNA from the oocyte versus follicle and stromal tissue that would be present in the libraries. However, after sequencing a number of clones, it is now clear that the contribution of oocyte mRNA is extremely high in comparison to other tissue sources such as the follicle wall. As a result, we probably obtained ESTs for transcripts that were mainly of oocyte origin. To obtain more transcripts that would come from the follicle wall or stroma, it might be necessary to sequence a very large number of ESTs from the libraries. It is clear that follicle wall genes are present in

these cod libraries since, for example, we have used the same cDNA libraries to isolate follicle cell regulators of steroidogenesis such as StAR (steroidogenic acute regulatory protein) (Goetz *et al.*, 2004). However, the magnitude of the disparity in genes originating from the oocyte versus the follicle wall is demonstrated by the fact that StAR is very highly upregulated at oocyte maturation in cod, yet it was not identified in the ~1,300 sequences in this study. Alternatively, we might need to prepare new libraries in which egg cytoplasm has been removed first so that the extra-oocyte components can be enriched. We did this previously with yellow perch *(Perca flavescens*), to enrich libraries for follicle wall genes involved in ovulation (Langenau *et al.*, 1999). However, this procedure is difficult for species in which the oocytes are small.

5. References

Arukwe, A., and Goksoyr, A. 2003. Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. Comparative Hepatology, 2: 4-24.

Atchison, F. W., Capel, B., and Means, A.R. 2003. Pin1 regulates the timing of mammalian primordial germ cell proliferation. Development, 130: 3579-86.

Bauer, M. 2001. Zebrafish mutagenesis: The isolation and characterization of reproductive mutations in the adult zebrafish. Ph.D. Thesis. The University of Notre Dame, Notre Dame, IN. 172 pp.

Brown, J. A., Minkoff, G., and Puvanendran, V. 2003. Larviculture of Atlantic cod (*Gadus morhua*): progress, protocols and problems. Aquaculture, 227: 357-372.

Chomcynski, P. 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques, 15: 532-537.

Chomcynski, P. a. S., N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry, 162: 156-159.

Fujimori, F., Takahashi, K., Uchida, C., and Uchida, T. 1999. Mice lacking Pin1 develop

normally, but are defective in entering cell cycle from G(0) arrest. Biochemical and Biophysical Research Communications, 265: 658-63.

Goetz, F. W., Norberg, B., McCauley, L.A., and Iliev, D.B. 2004. Characterization of the cod (*Gadus morhua*) steroidogenic acute regulatory protein (StAR) sheds light on StAR gene structure in fish. Comparative Biochemistry and Physiology-Part B: Biochemistry Molecular Biology, 137: 351-62.

Harding, M.M., Anderberg, P.I., and Haymet, A.D.J. 2003. 'Antifreeze' glycoproteins from polar fish. Eurpean Journal of Biochemistry, 270: 1381–1392 .

Hsu, S. Y., Kaipia, A., McGee, E., Lomeli, M., and Hsueh, A.J. 1997. Bok is a pro-apoptotic Bcl-2 protein with restricted expression in reproductive tissues and heterodimerizes with selective anti-apoptotic Bcl-2 family members. Proceedings of the National Academy of Sciences USA, 94: 12401-6.

Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Research, 9: 868-77.

Hukkanen, J. 2000. Xenobiotic-metabolizing cytochrome P450 enzymes in human lung. Academic Dissertation. Department of Pharmacology and Toxicology, University of Oulu, Finland. 69 pp. (http://herkules.oulu.fi/isbn9514258649/)

Jiang, W., Jimenez, G., Wells, N.J., Hope, T.J., Wahl, G.M., Hunter, T., and Fukunaga, R. 1998. PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. Molecular Cell, 2: 877-85.

Jin, K., Mao, X.O., Cottrell, B., Schilling, B., Xie, L., Row, R.H., Sun, Y., Peel, A., Childs, J., Gendeh, G., Gibson, B.W., and Greenberg, D.A. 2004. Proteomic and immunochemical characterization of a role for stathmin in adult neurogenesis. Federation of Experimental Societies for Experimental Biology Journal, 18: 287-99.

Johansen, S., and Bakke, I. 1996. The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): relevance to taxonomic studies among codfishes. Molecular Marine Biology and Biotechnology, 5: 203-14.

Kishimoto, T. 2003. Cell-cycle control during meiotic maturation. Current Opinions in Cell Biology, 15: 654-63.

Kjesbu, O. S. 1989. The spawning activity of cod, *Gadus morhua* L. Journal of Fish Biology, 34: 195-206.

Kjesbu, O. S. 1991. A simple method for determining the maturity stages of Northeast Arctic cod (*Gadus morhua* L.) by in vitro examination of oocytes. Sarsia, 75: 335-338.

Kjesbu, O. S., Witthames, P.R., Solemdal, P., and Walker, M.G. 1990. Ovulatory rhythm and a method to determine the stage of spawning in Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Science, 47: 1185-1193.

Krapivinsky, G. B., Ackerman, M.J., Gordon, E.A., Krapivinsky, L.D., and Clapham, D.E. 1994. Molecular characterization of a swelling-induced chloride conductance regulatory protein, pICln. Cell, 76: 439-48.

Kroll, K. L., Salic, A.N., Evans, L.M., and Kirschner, M.W. 1998. Geminin, a neuralizing molecule that demarcates the future neural plate at the onset of gastrulation. Development, 125: 3247-58.

Langenau, D. M., Goetz, F.W., and Roberts, S.B. 1999. The upregulation of messenger ribonucleic acids during 17alpha, 20beta-dihydroxy-4-pregnen-3-one-induced ovulation in the perch ovary. Journal of Molecular Endocrinology, 23: 137-52.

Liu, X. Y., Dangel, A.W., Kelley, R.I., Zhao, W., Denny, P., Botcherby, M., Cattanach, B., Peters, J., Hunsicker, P.R., Mallon, A.M., Strivens, M.A., Bate, R., Miller, W., Rhodes, M., Brown, S.D., and Herman, G.E. 1999. The gene mutated in bare patches and striated mice encodes a novel 3beta-hydroxysteroid dehydrogenase. Nature Genetics, 22: 182-7.

Melkonyan, H. S., Chang, W.C., Shapiro, J.P., Mahadevappa, M., Fitzpatrick, P.A., Kiefer, M.C., Tomei, L.D., and Umansky, S.R. 1997. SARPs: a family of secreted apoptosis-related proteins. Proceedings of the National Academy of Sciences USA, 94: 13636-41.

Nichols, K. M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A., and Thorgaard, G.H. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). Animal Genetics, 34: 102-15.

Obaya, A. J., and Sedivy, J.M. 2002. Regulation of cyclin-Cdk activity in mammalian cells. Cellular and Molecular Life Sciences, 59: 126-42.

Rexroad, C. E., 3rd, Lee, Y., Keele, J.W., Karamycheva, S., Brown, G., Koop, B., Gahr, S.A., Palti, Y., and Quackenbush, J. 2003. Sequence analysis of a rainbow trout cDNA library and creation of a gene index. Cytogenetic Genome Research, 102: 347-54.

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

Genome Research, 14: 478-90.

Roymans, D., Willems, R., Van Blockstaele, D.R., and Slegers, H. 2002. Nucleoside diphosphate kinase (NDPK/NM23) and the waltz with multiple partners: possible consequences in tumor metastasis. Clinical and Experimental Metastasis, 19: 465-76.

Shibata, Y., Iwamatsu, T., Oba, Y., Kobayashi, D., Tanaka, M., Nagahama, Y., Suzuki, N., and Yoshikuni, M. 2000. Identification and cDNA cloning of alveolin, an extracellular metalloproteinase, which induces chorion hardening of medaka (*Oryzias latipes*) eggs upon fertilization. Journal of Biological Chemistry, 275: 8349-54.

Sigurgislason, H., and Arnason, E. 2003. Extent of mitochondrial DNA sequence variation in Atlantic cod from the Faroe Islands: a resolution of gene genealogy. Heredity, 91: 557-64.

Sun, B. I., Bush, S.M., Collins-Racie, L.A., LaVallie, E.R., DiBlasio-Smith, E.A., Wolfman, N.M., McCoy, J.M., and Sive, H.L. 1999. derriere: a TGF-beta family member required for posterior development in *Xenopus*. Development, 126: 1467-82.

Tamura, K., Hara, T., Yoshie, M., Irie, S., Sobel, A., and Kogo, H. 2003. Enhanced expression of uterine stathmin during the process of implantation and decidualization in rats. Endocrinology, 144: 1464-73.

Tsao, C. C., Huang, F.L., and Chang, Y.S. 1999. Purification, characterization, and molecular cloning of carp hyosophorin. Molecular Reproduction and Development, 52: 288-96.

Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M.. 2000. Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Molecular Cell, 6: 961-7.

van den Hurk, W. H., Martens, G.J., Geurts van Kessel, A., and van Groningen, J.J. 2004. Isolation and characterization of the *Xenopus laevis* orthologs of the human papillary renal cell carcinoma-associated genes PRCC and MAD2L2 (MAD2B). Cytogenetic Genome Research, 106: 68-73.

Wang, K., Yin, X.M., Chao, D.T., Milliman, C.L., and Korsmeyer, S.J. 1996. BID: a novel BH3 domain-only death agonist. Genes and Development, 10: 2859-69.

Wu, X., Viveiros, M.M., Eppig, J.J., Bai, Y., Fitzpatrick, S.L., and Matzuk, M.M. 2003a. Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. Nature Genetics, 33: 187-91.

Wu, X., Wang, P., Brown, C.A., Zilinski, C.A., and Matzuk, M.M. 2003b. Zygote arrest 1 (Zar1) is an evolutionarily conserved gene expressed in vertebrate ovaries. Biology of Reproduction,

69: 861-7.

Zhao, Z., Deng, G., Lui, Q., and Laursen, R.A. 1998. Cloning and sequencing of cDNA encoding the LS-12 antifreeze protein in the longhorn sculpin, *Myoxocephalus octodecimspinosis*. Biochimica et Biophysica Acta, 1382: 177-80.

Table 1: Percent of cod ovarian ESTs grouped as structurally or functionally similar

Sequence Category	% of 1,361 ESTs	
Ribosomal Proteins	15.8	
Metabolic	9.9	
Structural Proteins	8.3	
Cell Signaling/Apoptosis	4.6	"Regulatory Proteins"
Cell Division/Cell Cycle Regulators (replication, mitosis, meiosis)	2.1	
Transcription/Translation/Protein processing	5.4	
Developmental Regulators	2.0	
Proteases/Antiproteases	1.3	
Immune-Related Proteins	1.1	
Others[1]	3.7	
Similar but unknown[2]	19.0	
No alignment[3]	26.8	

[1] Sequences with significant Blast hits but could not be classified into groupings above

[2] Sequences with significant Blast hits but of "unknown, hypothetical protein," etc. classification

[3] Sequences with no significant Blast hits against the NCBI nr and nt databases

Table 2: Selected genes from cod ovarian cDNA ESTs including most significant blastx hit against NCBI nr database

Putative Name/Function	EST/cDNA Accession #	Size (bp)	Blastx similarity score	Species most similar to	Accession # of similar protein	# of clones/ 1,361 seq
Structural Proteins						
choriogenin L	CO542190	779	2.0E-15	*O. latipes*	AAM91819	6
ZPC domain containing protein 1	CO542603	764	1.0E-66	*O. latipes*	AAD38906	3
ZPC domain containing protein 2	CO542098	779	1.0E-75	*O. latipes*	AAN31189	2
ZPA domain containing protein	CO542593	810	2.0E-55	*O. latipes*	AAD38904	5
ZPAX	CO542010	870	5.0E-70	*O. latipes*	AAN31186	2
beta 1 tubulin	CO541590	755	1.0E-143	*G. morhua*	AAC78686	4
profilin 2-like	CO541946	713	2.0E-55	*D. rerio*	AAH78650	7
antifreeze protein	AY584595	647	1.0E-18	*M. octodecem-spinosus*	P80961	12
Cell Division/Cycle Cycle Regulators						
G2/mitotic-specific cyclin B2	CO541792	871	6.0E-81	*O. latipes*	Q9DG96	6
G2/mitotic-specific cyclin B1	CO541540	678	1.0E-75	*O. latipes*	Q9DG97	4
mitotic cyclin a2-type	CO541990	703	2.0E-06	*Arabidopsis*	A96803	1
Wee1	CO542429	768	4.0E-98	*C. auratus.*	BAB70752	1
protein NIMA-interacting 1 (Pin 1)	CO541689	817	6.0E-69	*M. musculus.*	NP_075860	1
protein regulator of cytokinesis 1 isoform 1 (PRC1)	CO542004	746	5.0E-54	*D. rerio*	AAH46886	1
securin	CO541892	832	3.0E-08	*X. laevis*	AAF32357	1
mitotic spindle assembly checkpoint protein (MAD2B)	CO542663	672	2.0E-68	*X. laevis*	CAC86900	1
GCIP-interacting protein p29 isoform 1	CO541943	769	3.0E-97	*M. musculus*	NP_081056	1
Cell Signalling/Apoptosis						
estrogen responsive zinc finger protein	CO542304	494	2.0E-07	*D. rerio*	AAH45450	1
Ras association domain family 1 isoform C	CO541723	734	4.00E-47	*H. sapiens*	NP_733831	1
RAS, dexamethasone-induced 1	CO542446	839	1.0E-110	*M. musculus*	NP_033052	1
G-protein: G(I)/G(S)/G(O) gamma-5 like	CO541764	540	3.0E-26	*H. sapiens*	AAP36465	1
G-protein: beta subunit-like protein	CO542763	802	1.0E-136	*D. rerio*	NP_571519	5
G-protein G(i) alpha subunit	CO541729	729	1.0E-131	*O. latipes*	P87383	1
G-protein G(o)	CO541666	615	6.0E-31	*G. cydonium*	Q9XZV3	1
AMP-activated protein kinase beta subunit	CO541987	763	5.0E-74	*H. sapiens*	NP_006244	1
GTPase cRhoA	CO542194	696	1.0E-106	*G. gallus*	AAC18962	1
CREB protein	CO542479	820	6.0E-47	*H. sapiens*	AAC51331	4
tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase	CO541865	767	2.0E-39	*R. norvegicus*	NP_113791	1

Cell Signalling/Apoptosis (continued)						
S-100 protein alpha chain	CO542405	764	1.00E-38	*M. fossilis*	S35985	4
annexin III (lipocortin III)	CO542785	734	6.0E-50	*H. sapiens*	NP_005130	1
calmodulin 3	CO542018	756	5.0E-14	*H. sapiens*	CAD79597	1
Bcl-2-related ovarian killer protein	CO542330	766	2.0E-60	*G. gallus*	AAF81282	1
Ran protein	CO542337	726	4.0E-72	*S. salar*	CAA10039	1
regulator of G-protein signalling 19 interacting protein 1	CO541743	815	1.0E-122	*H. sapiens*	NP_005707	1
secreted apoptosis related protein 1	CO542111	835	8.0E-49	*M. musculus*	AAB70795	1
nucleoside-diphosphate kinase 2	CO542766	708	3.0E-66	*O. mykiss*	AAO42980	1
nucleoside diphosphate kinase 6	CO542513	800	2.0E-64	*D. rerio*	NP_571672	1
myeloid cell leukemia sequence 1	CO541777	620	3.0E-12	*D. rerio*	NP_919375	1
mucosa associated lymphoid tissue lymphoma translocation protein 1	CO542251	789	5.0E-14	*D. rerio*	NP_694508	1
inhibitor of growth family, member 5	CO541466	720	1.0E-68	*D. rerio*	NP_937854	1
angiopoietin-like 4 protein	CO542358	896	2.0E-25	*D. rerio*	NP_571893	1
BH3 interacting domain death agonist (BID)	CO541487	758	2.0E-09	*G. gallus*	NP_989883	1
Developmental Regulators						
zygote arrest 1	CO541612	856	1.0E-57	*H. sapiens*	NP_783318	7
stathmin 1	CO542152	873	2.0E-47	*R. norvegicus*	NP_058862	5
integrin beta 1 binding protein 3	CO542430	817	5.0E-70	*D. rerio*	AAH47858	2
Nedd4 family interacting protein 1	CO541596	822	2.0E-82	*H. sapiens*	NP_085048	1
geminin H	CO542557	645	2.0E-39	*X. laevis*	AAC41304	1
derriere	CO542177	479	5.0E-63	*D. rerio*	NP_571023	1
Immune Related						
cyclophilin A	CO542617	738	8.0E-63	*B. bassiana*	AAN39296	4
CD209	CO542029	783	1.0E-18	*H. sapiens*	NP_055072	1
heat shock protein hsp90 beta	CO541906	767	1.0E-111	*S. salar*	AAD30275	2
HSP binding immunophilin	CO542295	776	1.0E-80	*D. rerio*	AAH45387	1
Others						
cytochrome P450 11A1	AY706102	599	4.0E-53	*O. mykiss*	Q07217	1
cytochrome P450 2C33	AY706104	766	2.0E-50	*S. scrofa*	BAB85663	1
cytochrome P450 1C2	AY706103	659	4.0E-91	*S. chrysops*	AAL78299	1
NAD(P) dependent steroid dehydrogenase-like	AY706101	738	1.0E-85	*M. musculus*	AAH52834	1
prostaglandin D synthase	CO541950	847	4.0E-51	*D. rerio*	BAB88223	1
karyopherin (importin) alpha 2	CO542687	677	4.0E-58	*O. niloticus*	AAD51751	3
hyosophorin	CO542550	853	8.0E-12	*C. carpio*	AAC27329	2
alveolin	CO542059	764	2.0E-42	*O. latipes*	BAA90750	1
cold-inducible RNA binding protein 2	CO541897	779	1.0E-29	*X. laevis*	BAB19129	1
swelling-induced chloride conductance regulatory protein	CO542075	835	7.0E-53	*D. rerio*	AAH52141	1