**FISHEGG: FINAL PUBLISHABLE SUMMARY REPORT**

Poor and variable egg quality is a persistent problem limiting fish farming around the world. Most substances in ovulated eggs are derived from multiple yolk precursor proteins, vitellogenins (Vtgs). The different types of Vtg have special functions in oocyte growth and maturation, but their disparate role(s) in determination of egg quality and embryo developmental potential were unknown. This project sought to deliver a complete portrait of the identity, functions and importance of multiple Vtgs and their product yolk proteins in determining egg quality in the zebrafish. Specific objectives were: 1. Molecular characterization of the structure, expression, product proteins, and phylogeny of the multiple Vtg genes, 2. Proteomic detection and quantification of the Vtgs and related Vtg-derived yolk proteins in maternal females, oocytes, eggs and offspring of good and poor quality spawns, and 3. Identification of essential Vtgs in studies employing novel gene knock-out technology.

The zebrafish genome was scrutinized to identify *vtg* genes, gene expression sites were identified by mapping *vtg* mRNA sequences to their parent *vtg* genes, and tissue transcript data was obtained by high-throughput RNA sequencing (RNAseq) in another project (PHYLOFISH). The *in silico* and RNAseq analyses revealed that zebrafish possess 8 different *vtg* genes (*vtg1, 2, 3, 4, 5, 6, 7*, and *8*) mainly expressed in the liver and intestine of adult females. Synteny analyses localized *vtg1, 4, 5, 6,* and *7* to a cluster on chromosome 22, *vtg 2* and *8* to a neighboring cluster on the same chromosome, and *vtg3* isolated on chromosome 11. Sequence analyses revealed that the deduced Vtg1, 4, 5, 6 and 7 polypeptides are nearly identical (>90%), missing β’-component (βc) and C-terminal (Ct) yolk protein domains present in complete forms of teleost Vtg, and orthologous to primitive forms of Vtg in carps (e.g.VtgAo1). Zebrafish Vtg2 and 8 are highly similar and contain all known yolk protein domains, although Vtg8 possesses a shorter phosvitin (Pv) domain, and are orthologous to carp VtgAo2, salmonid VtgAsa-b and eel VtgAe. Zebrafish Vtg3 is missing Pv, βc and Ct domains, having the unique incomplete structure seen in C-type Vtgs from other teleosts. PCR using gene-specific primers verified that each *vtg* gene (except *vtg8*) is capable of producing mRNA. We conclude that *vtg*1, 4, 5, 6, and 7 are collectively responsible for Vtg1 production, that *vtg2* and 8 may be jointly responsible for Vtg2 production, and that *vtg*3 is solely responsible for production of Vtg3. Mapping RNAseq data to genomic sequences showed that *vtg*1 transcripts are most abundant in adult female liver, followed by *vtg*2, *vtg*3, and *vtg*4-8 transcripts. Combined with estimates of Vtg protein abundance from nano-scale, liquid chromatography and tandem mass spectrometry (nanoLC-MS/MS) analyses in our ‘Pooled samples experiment’ (see below), these findings indicate that Vtg1 is the dominant form of Vtg in zebrafish, while Vtg8 is rare and least abundant. Results of Western blotting (WB) performed using Vtg type-specific antibodies (Abs) verified maternal synthesis, transport and gonadal uptake of all three major forms of Vtg.

 The zebrafish *vtg1*, *2* and *3* genes were targeted for knock-out (KO) using CRISPR/Cas methodology. As zebrafish *vtg1, 4, 5, 6* and *7* all give rise to the same type of Vtg protein (Vtg1), we identified common target sites to knock them out simultaneously. Three targets sites per *vtg* gene were identified. Guide RNA and CRISPR/nCas9n constructs were produced and injected into the animal pole of single cell stage embryos in fertilized eggs. After 24 h, selected embryos were screened for mutations by PCR using genomic DNA as the template. Batches of embryos with positive signals for desired mutations were raised to adulthood and screened again using genomic DNA from finclips as the PCR template. The type and location of mutations obtained after KO of the zebrafish *vtg1*, *2* and *3* genes were identified via sequencing and alignment to available zebrafish genomic sequences. All *vtgs* were successfully mutated with various deletions of up to 8kb of genomic DNA in F0-generation embryos. Adults with expected mutations were kept for further breeding/screening trials, which were continued until purebred lines of zebrafish with mutation(s) of interest were obtained. Phenotyping experiments were conducted on homozygous (Hm) females from the F3 generation of each line to determine impacts of gene KO on female reproductive performance, and embryonic and larval development and survival. Targeted gene expression levels were measured in F3 Hm, heterozygous (Ht) and wild type (Wt) female liver by RT-qPCR. Western blots (WB) using monospecific polyclonal antibodies (Abs) developed against certain yolk protein domains in each type of Vtg were carried out on liver, ovary and eggs to verify KO success. NanoLC-MS/MS analyses (see below) of fertilized eggs and liver from F3 Hm and Wt females further verified gene KO.

Even though there was no apparent effect of *vtg3* KO on female fecundity, egg diameter, and larval size and hatching time and rate, the F4 generation of the *vtg3* KO purebred line has shown up to 95% mortality at late larval stage (past day 20). Any hatched larvae showed active swimming and feeding activities despite an up to 15% incidence of abdominal edema and spinal cord malformations. RT-QPCR results revealed a significant difference in *vtg3* gene expression levels in Hm *vtg3* KO and wild type female livers. WB using the specific Ab developed against Vtg3 lipovitellin light chain (a-LvL3) verified the absence of LvL3 in the ovary and eggs of Hm *vtg3* KO females. Apart from verifying the success of *vtg3* KO, the nanoLC-MS/MS findings indicated up-regulation of structural proteins related to cardiac muscle development and contraction, which could, in turn, be related to abdominal edema development. In the *type 1 vtg* purebred KOline, mortality at late larval stage was even more drastic with no survival by 22 days after fertilization and more severe abdominal edema (up to 70 %) and spinal cord malformations before that time. Larvae from this line have shown no active swimming and/or feeding activity. The first attempt at *vtg2* mutation transmission to later generations failed and the experiment was repeated until transmission was achieved. Due to the delay, the purebred line for this mutation is still under development. However, the *vtg2* KO F3 generation Ht embryo has shown a distinct phenotype at an early embryonic stage, with yolk leaking through the vitelline membrane resulting in death of the embryo within < 8 hours after fertilization. Although some data from the latter two experiments is still under analysis, a manuscript presenting results from the *vtg3* KO results has been submitted for publication. Collectively, these findings indicate that, while both Vtg type 1 and Vtg3 may have crucial functions in late larval life of zebrafish, the Vtg2 is essential for development and survival at very early embryonic stages.

We also employed nanoLC-MS/MS to explore the general proteomics of egg quality in zebrafish. Samples of good versus poor quality eggs were subjected to nanoLC-MS/MS after fractionation by SDS-PAGE, excision of gel fractions, and *in gel* tryptic digestion of the samples. A ‘Pooled samples experiment’ involved high-resolution fractionation of two samples pooled from 4 individual spawns per egg quality group (20 fractions per group, 40 total) before nanoLC-MS/MS. A ‘Replicated samples experiment’ involved low-resolution fractionation of samples from 4 different spawns per egg quality group, yielding 2 fractions per spawn (16 total); these fractions were excised from the gel so as to avoid the major Vtg-derived yolk protein bands in each lane. Mass spectra were searched against a target-decoy concatenated database and analyzed using ProteoIQ to identify and quantify group specific and common proteins based on normalized spectral counts (NSC). Analyses based on manual curation and annotation of individual proteins revealed that poor quality eggs are deficient in proteins involved in energy metabolism, lipid metabolism and protein synthesis, and have a higher representation of proteins related to apoptosis, oncogenes, endosome/lysosomes, and innate immune functions, while also being rich in lectins and zona pellucida proteins. Automated gene ontology-based protein enrichment analyses (PANTHER GO-Slim) were highly confirmative of these findings. We also detected several potential egg quality marker proteins that were differentially expressed in good versus poor quality eggs.

This project achieved its major goal of delivering a complete molecular portrait of the existence, significance, and contributions to egg quality of multiple Vtgs and their product yolk proteins in the zebrafish. Definitive verification that all three Vtgs play essential roles in maternal oogenesis and embryonic and larval development was obtained in the gene KO studies. Further research is needed to identify the specific molecular processes impaired by deletion of each form of Vtg. Our novel finding that such deletions have far ranging effects on egg proteome profiles will undoubtedly lead to identification of new molecular functions of the different forms of Vtg that underpin their importance to egg quality. This work also identified several proteins, including Vtgs, as potential egg quality markers for future research. Collectively, these developments significantly advance our basic understanding of the molecular details of fish oogenesis and open the door to discovery of the proteomic determinants of egg quality, and to optimization of egg quality, in farmed fishes. Thus the project will contribute to sustainable growth of fish farming and global food security.