LISEN LI

THE IMMUNOME AND EMBRYO QUALITY IN SEA BREAM AND SEA BASS

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2022

LISEN LI

THE IMMUNOME AND EMBRYO QUALITY IN SEA BREAM AND SEA BASS

Doutoramento em Ciências Biológicas

Especialidade em Genética, Genómica e Evolução

Orientadores:

Prof.ª Doutora Deborah M. Power

Doutora Liliana I.T. dos Anjos Guerreiro

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

Faro, 2022

LISEN LI

THE IMMUNOME AND EMBRYO QUALITY IN SEA BREAM AND SEA BASS

Doctoral Programme in Biological Sciences
Specialty in Genetics, Genomics and Evolution

Supervisors:
Prof.ª Doutora Deborah M. Power
Doutora Liliana I.T. dos Anjos Guerreiro

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

Faro, 2022

Título | Thesis title

The immunome and embryo quality in sea bream and sea bass

Declaração de autoria de trabalho:

Declaro ser o autor deste trabalho que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Declaration of authorship:

This work has not previously been submitted for a degree in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Lisen Li

Support

Financial support for the work reported in this PhD thesis is gratefully acknowledged as it was crucial for the good progress of the scientific work. A diversity of funding sources supported the thesis work either directly through the purchase of consumables (PerformFISH a European Union's Horizon 2020 research and innovation grant, agreement No 727610) or through funding CCMAR support services and equipment used during execution of the thesis work (Portuguese Foundation for Science and Technology (FCT) project UIDB/04326/2020, UIDP/04326/2020 and from the operational programmes CRESC Algarve 2020 and COMPETE 2020 through project EMBRC.PT ALG-01-0145-FEDER-022121).

Apoio

O apoio financeiro ao trabalho relatado na presente tese de doutoramento é reconhecido com gratidão, pois foi crucial para o progresso positivo e sucesso do trabalho científico. Uma diversidade de fontes de financiamento apoiou o trabalho desenvolvido nesta tese, quer diretamente através da compra de consumíveis/materiais (PerformFISH a European Union's Horizon 2020 research and innovation grant, agreement Nº 727610), quer indiretamente através do financiamento do CCMAR de apoio a serviços e equipamentos utilizados durante a execução do trabalho (projetos da Fundação para a Ciência e Tecnologia (FCT) - UIDB/04326/2020, UIDP/04326/2020 e dos programas operacionais CRESC Algarve 2020 e COMPETE 2020 através do projeto EMBRC.PT ALG-01-0145-FEDER-022121).

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

CCMAR
Centro de Ciências do Mar
Algarve

UAlg
UNIVERSIDADE DO ALGARVE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727610.

PerformFISH

Acknowledgements

Within my PhD academic years, many lab mates, colleagues in the university and collaborators from European aquaculture companies and academic institutions provided or helped me to collect and prepare experimental samples and materials, optimize the methodologies, and offered ideas.

First, I am grateful to my supervisor Prof. Deborah M Power for giving much help, support and guidance on my thesis and make it possible for me to carry out successfully the thesis. Although professor was always in the university with a busy timetable, she still offered me a valuable opportunity to start the work on sea bream and sea bass and arranged a meeting every week or whenever necessary to solve problems encountered during the thesis work and to clearly point out the priority of the thesis work.

Thanks to my co-supervisor Dr. Liliana Anjos who guided me in the workflow of protein extraction and quantification, as well as the proteomic analysis. Apart from the experimental work, Liliana also directed me in drafting the scientific paper and helped edit the thesis manuscript with a great deal of patience.

I would like to thank Dr. João Cardoso who gave me the guidance when I started my first thesis chapter work. João brought me to the "portal" of bioinformatic analysis and taught me how to use and interpret data using this means. In the end, we finished the first chapter and published the paper on lysozyme. He also offered me much help in modifying the thesis manuscript.

Thanks to Prof. Adelino Canário for teaching me about analysis of statistical significance in the thesis work and providing guidance during group meetings every week. Professor also provided me with knowledge and skill in organizing experimental records and the use of scientific software.

Thanks to my colleague Soraia Santos who helped to organize fish embryo samples and provided support and guidence in the preparation of protocols for RNA extraction, cDNA synthesis and quantitative PCR. Therefore, this guaranteed the smooth proceed of the thesis work. Thanks to Dr Ana Patrícia Mateus for helping in lysozyme enzymatic assays and the data analysis with much patience and Dr Rute Felix for guiding me in the use of PCR.

Many thanks to all nice colleagues in the lab: Cármen Sousa, Babak Najafpour, Sania Rubaiyat, Bruno Louro, Pedro Guerreiro, Patrícia Pinto, Rute Martins. Also thanks to Marta Valente Bernardo from the CCMAR sequencing platform that analysed quickly and efficiently my samples.

Thanks to Ana Margarida Amaral, Liliane Leman, Rita Costa Abecasis and other colleagues from CCMAR office for giving many helps in dealing with my student affairs.

Thanks to Prof. Jiale Li, Assoc. Prof. Yubang Shen and Master student Weining Yang from Shanghai Ocean Univeristy for giving the guidance and substantial support and help for the study reported in thesis chapter 3

Thanks to my friend David Canário and other friends from Faro for the good company and encouragement between each other during my four-year study.

In the end, I would like to thank all my family especially my parents who have always supported me and given me everything unconditionally, and for their encouragement, concern and help whenever I needed it during my PhD but also throughout my life. I will make myself worthy of my parents' care and upbringing by living a positive life and working with great enthusiasm!

Matter of the today do not wait tomorrow,

don't wait others what he had done.

Johann Wolfgang von Goethe

Abstract

Gilthead sea bream (*Sparus aurata*) and European sea bass (*Dicentrarchus labrax*) are teleosts belonging to Eupercaria and are the most important aquaculture fish species in the Mediterranean region. These two species are ranked second after the Atlantic salmon (*Salmo salar*) in production volume and value in the European Union (EU) aquaculture sector. Unpredictable fertilized egg/embryo quality and performance remain a bottleneck that threatens sustainability of sea bream and sea bass aquaculture, impeding the increased productivity of aquaculture that entirely depends on hatchery production. To address this issue, criteria and molecular markers linked to embryo quality that could be used to monitor and manage hatchery production were procured. Comparative molecular approaches using molecular biology, proteomics and transcriptomics were performed to analyze embryo performance and immunity in samples from several European commercial hatcheries. The core achievements were the: a) identification and characterization of lysozyme and complement 5 (C5) gene families and embryo and larval gene expression and enzyme activity from a diversity of hatcheries, b) characterization of the embryo proteome from three Mediterranean fish species [white sea bream (*Diplodus sargus*), meagre (*Argyrosomus regius*) and sea bream] 24h before hatch and at hatch and identification of common and species specific molecular patterns linked to biological function and putative quality-related proteins, c) comparative transcriptomics of good and poor quality sea bream embryos from several Mediterranean hatcheries. Quality-linked transcripts and some elements of the regulatory epitranscriptome (non-coding RNA) were identified as well as the contribution of maternal proteins to embryos. Taken together, the results provide a comprehensive description of the molecular basis of sea bream and sea bass embryo development and reveal that immune-related molecules in fertilized eggs are low abundance. The development (quality)-related candidate markers identified will be of value for management of fish embryos in aquaculture hatcheries.

Keywords: Aquaculture hatchery, embryo quality, Evolutionary developmental biology, Molecular biomarker, Omics technologies

Resumo

O mar Mediterrâneo cobre apenas 0,7% da área oceânica mundial, mas é um dos principais reservatórios de biodiversidade marinha e costeira com cerca de 28% de espécies endémicas. A dourada (*Sparus aurata*) e o robalo (*Dicentrarchus labrax*), são duas espécies de peixe que pertencem à série Eupercaria e das mais comercializadas pela indústria de aquicultura nesta região. No sector da aquicultura da União Europeia (EU), estas duas espécies ocupam o segundo lugar no "ranking" da cadeia de valor, depois do salmão do Atlântico. Contudo, as suas características morfológicas e de crescimento são os parâmetros de qualidade relevantes considerados pela indústria. Esta abordagem deve-se à sua comercialização em formato de peixe inteiro, o que reduz o valor de mercado na cadeia de valor e também à ausência de critérios de qualidade para seleção de ovos e embriões, bem como, de marcadores moleculares de qualidade com maior grau de sensibilidade. Todos estes obstáculos, limitam substancialmente o desenvolvimento das indústrias associadas á comercialização destas duas espécies de peixe, impedindo a expansão da aquicultura e das "hatcheries" (maternidades incubadoras de ovos/embriões controlados artificialmente para fins comerciais). Para mitigar este problema, utilizou-se técnicas de biologia molecular e tecnologias ómicas e estabeleceu-se uma abordagem comparativa direcionada á descoberta de moléculas e vias metabólicas funcionais de importância crítica para o sistema imune dos peixes associada á "performance" de desenvolvimento de ovos e embriões.

Esta tese está organizada em seis capítulos. Inicia-se com uma visão geral dos critérios morfológicos, físico-químicos e moleculares existentes para avaliar a qualidade de ovos e embriões para melhorar a gestão da aquicultura de peixes *(Capítulo 1)*. Subsequentemente, caracterizou-se a família de genes do sistema imune: a) a das lisozimas em peixes teleósteos com enfase na sua caracterização molecular e funcional em dourada *(Capítulo 2)* e análises moleculares estruturais e evolutivas e b) a do complemento C5 (C5) em peixes especialmente em espécies da faamilia Cyprinidae *(Capítulo 3)*. Os *capítulos 4 e 5*, integram abordagens de proteómica e transcriptómica em espécies de peixes mediterrânicos [pargo (*Diplodus sargus*), corvina (*Argyrosomus regius*) e dourada], focando os processos de desenvolvimento e de eclosão e na função da enzima "hatching enzyme" em dourada e robalo. Foi feita uma associação entre os capítulos e a análise integrada dos dados do transcritoma do embrião *(Capítulo 5)* revelou um padrão de expressão significativamente diferente (p-valor < 0,05) para o C5 *(Capítulo 3)* em diferentes lotes de embriões de dourada nas comparações entre graus de

qualidade (Boa *vrs* Má) e entre estágios de desenvolvimento (Pré- eclosão *vrs* Eclosão). A variação do C5 em relação ao lote de embriões não foi afetada pela origem da "hatchery", indicando que as prática de manejo ou os próprios reprodutores não influenciam a sua expressão. Os resultados sugerem que este gene e o seu produto proteico, são provavelmente importantes na proteção imunológica precoce e também em outras funções ainda não descritas na dourada ou em outras espécies de peixes. Também a integração dos resultados do proteoma *(Capítulo 4)* e do transcritoma *(Capítulo 5)* do embrião de dourada nos mesmos estágios de desenvolvimento, identificou um grupo de proteínas que se especula serem de origem materna. No último capítulo, sumarizou-se os resultados e são apresentadas perspetivas baseadas nos avanços e desafios atuais e propostas para o desenvolvimento de uma ferramenta integrada de monitorização da qualidade dos embriões e uma base biológica do desenvolvimento de ovos e embriões de peixes *(Capítulo 6)*.

Neste projeto foram: 1) identificadas duas importantes famílias de genes associadas à imunidade inata em peixes, a das lisozimas e a do C5. Caracterizou-se pela primeira vez a função das lisozimas através da sua expressão e atividade enzimática em embriões e em diferentes estágios larvares de uma diversidade de reprodutores de dourada. Estudou-se a função do C5 em peixes, através da construção de redes génicas, modelação por homologia e "docking" molecular entre o C5 e o seu receptor (C5R/CD88); 2) mapeou-se e caracterizou-se o proteoma do embrião de três espécies de peixes mediterrânicos (sargo, corvina e dourada) em duas fases do seu desenvolvimento (24h antes da eclosão e na eclosão) e identificou-se um grupo de proteínas potencialmente relacionadas com a imunidade e a qualidade dos embriões. Avaliou-se a função do gene para enzima "hatching enzyme", com base na sua expressão em embriões de dourada e robalo nos estágios acima referidos; 3) mapeou-se e caracterizou-se o transcritoma de embriões de dourada com origem em diferentes "hatcheries" na região mediterrânica através de uma abordagem comparativa entre qualidade e estágios de desenvolvimento (qualidade- Boa *vrs* Má; estágios- Pré- eclosão *vrs* Eclosão) e identificou-se uma diversidade de transcritos, vias metabólicas e elementos do epitranscriptoma regulatório do RNA-não codificante. Foram identificados em comum 42 candidatos a marcadores de qualidade e enriquecidas duas vias metabólicas relacionadas com o sistema imunológico e associadas às "hatcheries": a via de infeção por Salmonella (constituída por 7 genes relacionados com o sistema imune) e a via de sinalização MAPK (mitogen-activated protein kinase). Foram identificadas 543 proteínas que são expressas apenas no proteoma, sugerindo que podem ter origem materna e destas, 7 (diferencialmente expressas) estão potencialmente relacionadas com o sistema imune. Globalmente, os resultados forneceram um grupo de

marcadores relacionados à imunidade e ao desenvolvimento (qualidade), com potencial de se traduzirem em critérios de qualidade de ovos e embriões para a indústria da aquicultura. Estes resultados, foram amplamente estudados para descrever a base molecular biológica entre os diferentes estágios de desenvolvimento de ovos e embriões de peixes e também, entre diferentes lotes de embriões de diferente qualidade. As ferramentas biológicas e critérios desenvolvidos neste trabalho, oferecem uma orientação para as "hatcheries" de peixes e a sua aplicação contribuirá para melhorar no futuro o sector da aquicultura.

Palavras-chave: Maternidade de aquicultura, Biologia evolutiva do desenvolvimento, Biomarcadores moleculares, Viabilidade de ovos/embriões, tecnologias ómicas.

Table of contents

List of Tables

List of Figures

List of Abbreviations

18s - 18 ribosomal subunit

ActRIIB - activin receptor type-2B

Apoe - apolipoprotein E

Arf5 - ADP-ribosylation factor 5

Arf4a - ADP-ribosylation factor 4a

Acsl6 - long-chain-fatty-acid–ligase 6-like

Angptl2 - angiopoietin-like 2

Ano10 - Anoctamin 10

Abhd5 - Abhydrolase Domain Containing 5

Atp2a2 - ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2

Atp6v0e2 - ATPase H+ transporting V0 subunit e2

Atp1a2a - sodium potassium-transporting atpase subunit alpha-2-like

Atp1a1 - sodium potassium-transporting atpase alpha-1 subunit

Ap2a1 - ap-2 complex subunit alpha-2-like isoform 1

Atp6v1ab - ATPase H+ transporting V1 subunit Ab

Atp6v1ba - ATPase, H+ transporting, lysosomal, V1 subunit B, member a

Atp6v1h - ATPase H+ transporting V1 subunit H

Atp6v1e1b - ATPase H+ transporting V1 subunit E1b

Ap2a1 - adaptor related protein complex 2 subunit alpha 1

Aff3 - F4/FMR2 family member 3

AR- *Argyrosomus regius*

ANOVA - analysis of variance

Arpc1b - actin related protein 2/3 complex, subunit 1B

Buc - bucky ball

Bruno-like - zebrafish mutant involved in animal-vegetal polarity

B4galnt1 - beta-1,4-N-acetyl-galactosaminyltransferase 1

Bmp-2 - Bone morphogenetic protein gene

BI - Bayesian Inference

BH - before hatch

BP - biological process

CycB - cyclin B

Cdc25 - cell division cycle 25

Cdc42 - cell division control protein 42 homolog isoform 1

Crp2 - C-reactive protein 2

Ckap5 - cytoskeleton-associated protein 5

Ctsl - Cathepsin L

Ctsd/CatD - cathepsin D

Ctsla - cathepsin isoform a

Ctslb - cathepsin L gene isoform b

Card14 - caspase recruitment domain family member-14

Cspg5 - Chondroitin Sulfate Proteoglycan 5

Cpm - carboxypeptidase M

Cpsf6 - cleavage and polyadenylation specific factor 6

Cct2 - chaperonin containing TCP1 subunit 2

Ccni - cyclin I

Ccnt1 - Cyclin T1

Crb2 - crumbs family member 2

Cct2- Chaperonin containing TCP1 subunit 2

Cct4 - Chaperonin containing TCP1 subunit 4

Crfb17 - interferon gamma receptor alpha chain

Chst10 - carbohydrate sulfotransferase 10

C19orf67 - chromosome 19 open reading frame 67

Coa5 - cytochrome c oxidase assembly factor 5

Cltcb - clathrin, heavy chain b (Hc)

Chg - choriogenin

ceRNA - competing endogenous RNA

CCR4-NOT - carbon catabolite repression 4 - negative on TATA-less complex

C5 - complement C5

C5AR - C5a receptor

C6 - complement C6

Cpb2 - carboxypeptidase B2

CI - confidence interval

CC - Cellular component

Casp3b - caspase 3, apoptosis-related cysteine peptidase b

Cycsb - cytochrome c, somatic b

Dazl - DAZL, deleted in azoospermia-like

Dnajb4 - dnaJ heat shock protein family (Hsp40) member B4

Dnajc25 - dnaJ heat shock protein family (Hsp40) member C25

Dcc - DCC netrin 1 receptor

Dennd1a - DENN domain containing 1A

Dctn3 - Dynactin Subunit 3

dpf - days post fertilization

Ddx39a - ExD-box helicase 39A

DS - *Diplodus sargus*

DL - *Dicentrarchus labrax*

DEP - differentially expressed proteins

Eomesodermin - a zebrafish gene with mesoderm inducing activity

Eef2 - elongation factor-2

Eno1a - Enolase A

Elp6 - Elongator Acetyltransferase Complex Subunit 6

ESTs - Expressed Sequence Tags

Ef1a - elongation factor 1-alpha

Evi5l - ecotropic viral integration site 5 like

Elav1a - embryonic lethal-abnormal vision) - like protein 1a

Eif5b - eukaryotic translation initiation factor 5B

FAA - free amino acids

Foxr1 - Forkhead box 1

Fell - Fish egg lectin like isoform X1

Fel - Fish egg lectin like precursor

Flna - filamin-A

Fcgbp - Fc fragment of IgG-binding protein

Fbxw2 - F-box and WD repeat domain containing 2

Frs2 - fibroblast growth factor receptor substrate 2

Fam161a - FAM161 centrosomal protein A

FAO - Food and Agriculture Organization

FEL - Fixed Effects Likelihood

FDR - False Discovery Rates

Goosecoid - a homeobox protein, Wnt antagonist

Gnpat - Glyceronephosphate O-acyltransferase

GxE –Genotype x Environment interaction

Grp78 - glucose-regulated protein

GO – Gene Ontology

hpf - hours post fertilization

Hamp1 - Hepcidin-1

Hsp70 - heat-shock protein 70

Hsp90 beta - heat shock protein 90 beta

Hsp90aa1.1 - heat shock protein hsp 90-alpha 1-like

HGNC - human official gene nomenclature

H3f3a - Histone H3.3

Hdlbpa - high density lipoprotein binding protein a

Hdac1 - probable histone deacetylase 1-b-like

H – hatch

He - hatching enzyme

Hsp90aa1.2 - heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 2

Ids - iduronate 2-sulfatase

Igf2bps - Insulin-like growth factor 2 mRNA-binding proteins

I2s - iduronate 2-sulfatase

Irf7 - Interferon regulatory factor 7

IFN-γ - Interferon gamma

Il1b - Interleukin-1 beta

Igs – Immunoglobulins

KO - knockout

Kansl2 - KAT8 regulatory NSL complex subunit 2

KEGG - Kyoto Encyclopedia of Genes and Genomes

Lrrc10 - leucine rich repeat containing 10

LYZ - Lysozyme

LYZ-C - chicken or conventional type LYZ

LYZ-G - goose type LYZ

LALBA – Lactalbumin

Lonrf2 - LON peptidase N-terminal domain and ring finger 2

Lipt1 - lipoyltransferase 1

lncRNAs - long non-coding RNAs

LPS - lipopolysaccharide

LTA - lipoteichoic acid

Mago-nashi - a conserved protein of unknown function with expression in zebrafish blastula

miRNAs - micro-RNAs

Mhc2a - mhc class 2A chain

Mdm2 - MDM2 proto-oncogene

Mfsd14b - Major facilitator superfamily domain containing 14B

Mbd2 - ethyl-CpG binding domain protein 2

Mdm1 - Mouse double minute 1 homolog

Mdm2- MDM2 proto-oncogene

Ma7d1 - map7 domain containing protein 1

Map2k7 - thimitogen-activated protein kinase 7

Mcoln1 - mucolipin 1

Mrpl30 - mitochondrial ribosomal protein L30

Mitd1 - microtubule interacting and trafficking domain containing 1

Mgat4a - mannosyl (alpha-1,3-)-glycoprotein beta - 1,4 – N – acetylglucosaminyltransferase A

MSTN – Myostatin

Mvp - major vault

MZT - maternal-to-zygote transition

MBT - mid-blastula transition

MALT - Mucosa-Associated Lymphoid Tissue

ML - Maximum-Likelihood

MAC - membrane attack complex

MSA - multiple sequence alignment

MSTN – Myostatin

MF - Molecular function

Notch1a - a member of the Notch family

Nanos - Nanos2, nanos2, Nanos Homolog 2 (Drosophila)

Nup107 - nucleoporin 107

Nup133 - Nucleoporin 133

ncRNAs - Non-coding RNAs

Nod1 - nucleotide oligomerization domain 1

Narl - antibacterial and antiviral-related lncRNA

Otulina - OTU deubiquitinase with linear linkage specificity a

Pou2A - gene encoding POU domain

Pabpc1a - poly(A) binding protein cytoplasmic 1a

P62 - Nucleoporin p62

Pfn2l - Profilin 2 like

Pgm1 - Phosphogluco-mutase 1

Pgd - 6-phosphogluconate dehydrogenase

Psma2-A - proteasome subunit alpha type

Prdx – peroxiredoxin

Phb2 - Prohibitin 2

Pdp1 - Pyruvate dehydrogenase phosphatase catalytic subunit 1

Pomgnt2 - Protein O-Linked Mannose N-Acetylglucosaminyltransferase 2 (Beta 1,4-)

Pold4 - DNA polymerase subunit delta 4

P2rx3 - Purinergic receptor P2X 3

Pus10 - Pseudouridine synthase 10

Pkn1 - protein kinase N1

Pglyrp2 - peptidoglycan recognition protein 2

PRR - Pattern Recognition Receptor

PAMPs - Pathogen Associated Molecular Patterns

PB - Phosphate buffer

PDB - Protein Data Bank

PPI - protein–protein interaction

Pv – phosvitn

PCA - Principal component analysis

Ryk - related to receptor tyrosine kinase

Rpl22 - Ribosomal protein L22

Rtkn – Rhotekin

Rpp25l - Ribonuclease P/MRP Subunit P25 Like

RAS - recirculating water systems

Rag-1 - Recombination-activating gene 1

Ran - GTP-binding nuclear protein Ran

Ralgps1 - Ral GEF with PH Domain And SH3 Binding Motif 1

Rev1 - ADNA directed polymerase

Rac1a - Rac family small GTPase 1a

Snail1 - (a zinc finger protein expressed in the mesoderm in zebrafish embryos)

Sox19 - SRY (sex determining region Y)-box 19

Stat3 - Signal transducer and activator of transcription 3

Sodc - superoxide dismutase 1

Ski7 - superkiller 7

Slc29a1a - solute carrier family 29, member 1a

Smarca4 - Transcription activator BRG1

Smarcc1 - SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily C Member 1

Smarca5 - swi snf-related matrix-associated actin-dependent regulator of chromatin subfamily a member 5

Slc35e3 - solute carrier family 35 member E3

Shroom3 - shroom family member 3

Siglec15 - Sialic Acid Binding Ig Like Lectin 15

Snrk - SNF Related Kinase

Ssr1 - signal sequence receptor, alpha

Sf1 - splicing factor 1 isoform 1

Srsf9 - Serine/arginine-rich splicing factor 9

sncRNA - short non-coding RNAs

Sacs - Sacsin genes

SLAC - Single Likelihood Ancestor Counting

SA - *Sparus aurata*

SWATH-MS - Sequential Window data independent Acquisition of the Total High-resolution-Mass Spectra

Taram-a - TGFβ-related type I receptor

Ttn – titin

Trim24 - transcription intermediary factor 1α

Taf5l - TATA-box binding protein associated factor 5 like

Txndc9 - oredoxin domain containing 9

Tsga10 - testis specific 10

Top2b - dna topoisomerase 2-beta

Tcp1 - chaperonin containing

Tryp - trypsin

TQTs - total quantified transcripts

Tlr18 - toll-like receptor 18

Unc45b - Unc-45 Myosin Chaperone B

Unc50 - unc-50 inner nuclear membrane RNA binding protein

Vg1 - vegetalising factor-1

Vasa - RNA binding protein with an RNA-dependent helicase

Vtg – vitellogenins

Valopa - val-opsinA

Valopb - val-opsinB

Yeats4 - YEATS domain containing 4

Ybx1 - Y-box binding proteins

YPs - yolk proteins

YSL - yolk syncytial layer

Zorba - a zebrafish homologue of the orb germline gene

ZP – zona pellicida

Zpax - zona pellucida glycoprotein AX

Zpb - zona pellucida glycoprotein B

Zp3 - zona pellucida glycoprotein 3

Zp3.3 - zona pellucida glycoprotein 3.3

Zp1x2 - zona pellucida protein X2

Zp3-l1 - zona pellucida protein 3-L1

Zp3-l2 - zona pellucida protein 3-L2

Znf830 - Zinc Finger Protein 830

ZGA - zygote genome is activate

CHAPTER 1

General Introduction and Objectives

(Under revision for Reviews in Aquaculture. CRediT statement; LL- writing - original draft, DMP – supervision, writing - review & editing, JRC - supervision, writing - review & editing, LA - writing - review & editing)

1.1. Aquaculture and its current challenges

The UN sustainable development goals identify "zero hunger" as one of the 17 goals to be achieved by 2030 (sdgs.un.org/goals). However, this cannot come from terrestrial production systems that are reaching their maximum capacity for exploitation of space and water resources for food production. This means that alternative sources of good quality protein are required if the UN sustainable development goal, "zero hunger" is to be reached. Exploitation of the marine space that occupies 2/3^{rds} of the planet's surface is seen as one immediate solution towards "zero hunger". Aquaculture currently provides 52% of the fish consumed by humans and is valued at USD 250 billion (FAO, 2020) and has substantial space for growth. Currently the demand for fish products for human consumption is increasing and aquaculture production needs to expand to meet the global demand in line with the predicted growth in the world population.

Gains in aquaculture production can come from diversification of the species that are exploited, expansion of the space used for aquaculture but also improved management of existing production systems. Assessing quality and supplying an adequate quantity of fish seed is one of the most important factors for sustainable and profitable fish farming (Issa *et al.* 2022). Across the world, the major problems and challenges for fish nurseries and seed production are generally similar (Table 1.1) and they restrict the development and expansion of hatcheries. Limiting factors include high fry prices and feed costs, lack of knowledge and guidelines for sustainable management of production, broodstock genetics and disease. All the preceding factors negatively affect fish embryo and larval development and consequently the productivity, quality, and profitability of aquaculture. Other factors indirectly related to fish aquaculture that impact production include management problems and the lack of adequate credit lines (Yassien *et al.* 2022; Anastasiou *et al.* 2014).

Table 1.1. A summary of the main challenges facing the global aquaculture industry and hampering the development and expansion of fish aquaculture.

Country	Main species	Developmental stage	Major challenges	Impact	References
Bangladesh	Rui (*Labeo rohita*), Catla (*Gibelion catla*), Mrigel (*Cirrhinus cirrhosus*), Kalibaush (*Labeo calbasu*), Sarpunti (*Puntias sarana*), Pangus (*Pangasius hypophthalmus*), Tilapia (*Oreochromis mossambicus*), Bighead carp (*Hypophthalmichthys nobilis*), Silver carp (*Hypophthalmichthys molitrix*), Shing (*Heteropneustes fossilis*) and Magur (*Clarias batrachus*)	Fry and fingerlings	Lack of hatcheries, high feed costs, lack of stakeholder linkage, lack of capital for fish farming, lack of knowledge, guidelines, and consultancy for fish culture.	Mortality during transport: 7.2% (fry), 2.8% (fingerlings)	(Rana *et al.* 2022)
Egypt	Nile tilapia (*Oreochromis niloticus*) and African catfish (*Claris gariepinus*)	Fish fry (94%) and other developmental stages (6%)	High feed costs, lack of operating capital, low technology adoption, inadequate supply of fingerlings, high energy prices, high fry prices, lack of credit support, management problems, labour irregularity, high price of vaccination and medication.	98% of farmers face high fry price	(Yassien *et al.* 2022; Kaleem & Sabi 2021)
Indonesia	Catfish (patin), nile tilapia, goldfish, catfish (lele), pomfret fish	Broodstock fish	Decreasing fishery production.	n.a.	(Mustika 2022)
Spain	European sea bass	All stages	Disease transmission and imported	n.a.	(Muniesa *et al.*

	(*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*)		diseases, lack of communication between stakeholders, deficient coordination of health strategies, lack of seed, high competition with other EU producers.		2022; Cavallo *et al.* 2020)
Nigeria	Tilapia, carp, mudfish and catfish	Fry and fingerlings	Inadequate product development, lack of fingerlings, poor fish growth rate, unavailability and inadequate access to modern fish-farming technologies.	n.a.	(Issa *et al.* 2022)
Brazil	Mandis (*Iheringichthys labrosus* and *Pomadasys maculatus*), Traíra/Lobó (*Hoplias malabaricus*), and the silver croaker (*Plagioscion squamosissimus*)	Juveniles and adults	Lack of ecological studies and adequate scientific monitoring, lack of communication between research and industry, not economically sustainable, concentrated on a few species, low productivity.	n.a.	(Casimiro *et al.* 2022; Valenti *et al.* 2021)
Greece	European sea bass (*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*)	All stages	Low production volumes, financial difficulty, limited access to credit, harsh environmental conditions and occupational risk.	n.a.	(Anastasiou *et al.* 2014; Katselis *et al.* 2022)

n.a. – Not available; latin name of fish species were not shown in the studies of Indonesia and Nigeria.

The genetic background of the broodstock influences the characteristics and phenotype of their offspring. Since in the Mediterranean region, only 56% of the farmed European sea bass (*Dicentrarchus labrax*) originate from operational breeding programs (Janssen *et al.* 2017; Robledo *et al.* 2018), it means there is substantial genetic diversity and this increases the challenge of optimizing management conditions for performance (Chavanne *et al.* 2016; Pickering 1993; Pottinger 2008). The gilthead sea bream (*Sparus aurata*) was the second most produced fish species in the Mediterranean region in 2019 and more than 80% of the genetically improved fingerlings came from Greece. The movement of genetically selected larvae from the environment in which they were produced to an environment with different conditions, increases production risks due to potential genotype - environment interaction (GxE). Such GxE interactions may lead to reduced performance, which can limit substantially the effectiveness of breeding programs (Gulzari *et al.* 2022).

In addition to the challenges linked to production systems there are an increasing number of challenges arising from the environment. Extreme temperatures above the normal threshold for fish in some environments are increasingly frequent due to global warming and tend to lower the performance, health, and productivity of aquaculture (Mugwanya *et al.* 2022). Climate change induced temperature shifts in surface waters negatively impact the production of both marine and freshwater aquaculture species due to the reduced concentrations of dissolved oxygen, which increases fish metabolism and changes in the sea surface salinity due to either the rise of sea water levels in some regions and increased evaporation in others (Frost *et al.* 2012). Hence, achieving a full understanding of the biology and ecology of fish especially the effects of temperature is critical for improved production, management, and conservation of fishes (Hallerman *et al.* 2022). Associated anthropogenic-linked factors restricting aquaculture development are water pollution including non-biodegradable and persistent heavy metals and microplastics, which affect fish growth, reproduction, health and productivity as well as their safety for consumers (Bringer *et al.* 2021; Tsai *et al.* 2022; Taslima *et al.* 2022; Sussarellu *et al.* 2016).

Most commercial aquaculture fish species are produced using extensive, semi-intensive and intensive aquaculture systems depending on the external supply of feed (Figure 1.1). In extensive systems there is no external supply of feed and fish production depends entirely on natural processes for feed. In semi-intensive aquaculture systems some supplementary feed is required but in intensive systems there is a greater dependency on the use of external feeds and this allows a higher density of farmed organisms and favours higher production (Council 2021). Fed aquaculture production is progressively increasing and in 2020 represented 60 million

tonnes of the total worlds production while non-fed aquaculture is declining and represented 27.8 percent which corresponds to about 22 million tonnes of the total farmed aquatic animal production (FAO 2022).

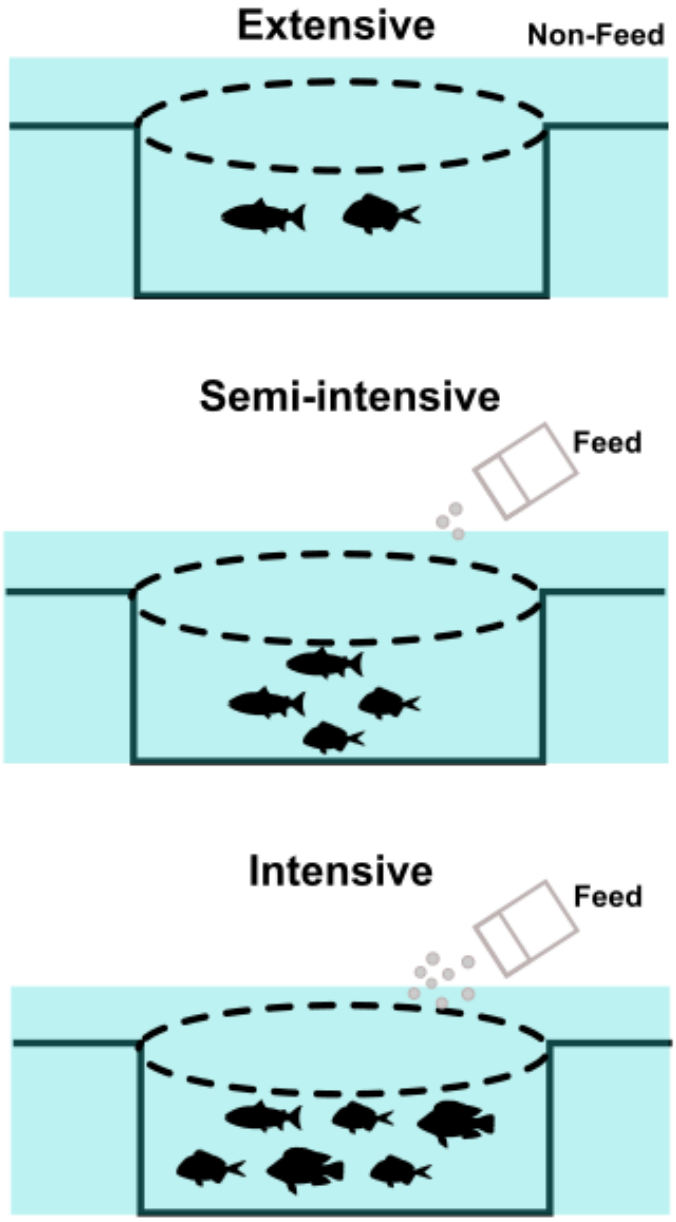


Figure 1.1. A representation of the three main aquaculture production systems for fish. In extensive systems there is no external food supply. In semi-intensive and intensive systems food is supplied, but the intensive production is almost entirely dependent on external feeds and a high stocking density of fish is practiced.

Intensive production using recirculating water systems (RAS) is the most common method for larval rearing. RAS differs substantially from the natural environment and requires strict control of live feed production and of physical and chemical parameters of the water as these

are all factors that influence larval and juvenile development, stress and physiology (Helvik *et al.* 2009). In intensive culture systems fish larvae are produced in indoor tanks and fed several times daily with live algae or algal pastes and cultivated rotifers and artemia (Helvik *et al.* 2009). The "live food" produced in hatcheries is generally of lower nutritional quality when compared to natural zooplankton and this can result in a reduced developmental performance, lower survival rates and unpredictable juvenile quality, which negatively affects management and sustainability (Helvik *et al.* 2009). For example, in post first feeding red snapper (*Lutjanus campechanus*) lower survival rates occurred with standard hatchery based live feed but this was relieved when copepod nauplii were supplied as initial prey (Shields *et al.* 2005; Rhodes & Phelps 2008; Watanabe *et al.* 2005; Saillant *et al.* 2013). In the turbot (*Psetta maxima*), egg ovulation is a major cause of poor-quality egg batches in captivity although the hatching rate of freshly ovulated eggs can be up to 97% by egg stripping and artificial dry fertilization (Rurangwa & Poelman 2011). Atlantic cod (*Gadus morhua*) larvae produced in intensive production systems have slower growth (Leifson 2003) and an increased rate of deformities (Imsland *et al.* 2006) when compared to larvae cultured in semi-extensive rearing systems (Hamre 2006).

Mass mortalities of fish larvae in hatcheries are also related with the rearing technology for intensive production that compromises water quality and generates an artificial environment that promotes proliferation of opportunistic bacteria, which negatively affect juvenile production (Skjermo & Vadstein 1999). RAS in hatcheries is aimed at reducing environmental impacts of aquaculture and prevents horizontal disease transfer between rearing systems and can ensure quality and continuous production of an adequate supply of fingerlings (Buric *et al.* 2014). RAS systems allow water reuse by managing waste and nutrient recycling, making intensive fish production environmentally sustainable. However, the frequency and volume of water exchange (Espinal & Matulić 2019) can compromise water quality and impact on the production and mortality of larvae and juveniles and thus brings new problems and challenges for the hatcheries (Nieto *et al.* 2010; Martins *et al.* 2009). Other impacts from RAS are a production that is largely dependent on the life stage, lack of trained personal, poor water quality in the system and poor system design and high initial investment (Table 1.2). Meanwhile, the management of the circuit can modify the dynamic microbial communities and bacterial activities, which are sensitive to variation in salinity, temperature, and pH and this can result in high mortality rates (Almeida *et al.* 2021; Bagarinao 1993).

Table 1.2. Some limitations of RAS system in the aquaculture sector.

Main issues	Reference
Largely dependent on the life stages	(Badiola *et al.* 2012; Preena *et al.* 2021)
Lack of trained personal	
Poor water quality in the system	
Poor system design	
Growth of opportunistic bacterial contamination	(Murray *et al.* 2014)
High initial investment	(Interdonato 2012; Clough *et al.* 2020)

The production of fish for aquaculture depends on many factors such as the genetic background of the broodstock, management regimes, good technical knowledge, stocking density, type of production system and the impact of climate change on the marine environment. Thus, monitoring fish egg, embryo and larval quality is essential to ensure a high-quality final product.

1.2.The importance of egg/embryo quality in fish aquaculture

The definition of egg/embryo quality is not consensual, and a diversity of parameters have been considered and their application may be linked to the final use of the eggs. However, from a more general perspective, egg quality is associated with fertilization rates and normal embryonic development (Bobe 2015b). From an aquaculture perspective the main concern of the hatchery is to increase broodstock fecundity, as well as to reduce the mortality of eggs/embryos at vulnerable stages including fertilization, hatching, and first feeding (Carnevali *et al.* 2001a). Predictable and routine production of good quality eggs/embryos remains a challenge for industry and can be a bottleneck for the management of production cycles in aquaculture. In the past decades myriad studies have been carried out to identify reference criteria for egg/embryo quality. Of note, Kjørsvik *et al*. (2003) after evaluating multiple parameters concluded that the use of a single or a few criteria to evaluate egg/embryo quality of all fish species used for aquaculture is not possible due to the high species variability for each criterion. For example, although the fertilization rate is regarded as a reference indicator for egg quality in salmonids, this parameter is not so well correlated with egg quality in marine fish species (Kjorsvik *et al.* 2003; Ienaga *et al.* 2021). Instead marine fish egg quality is better associated with cell symmetry at early cleavage stages (normal blastomeres) (Kjorsvik *et al.* 1990) and significant positive correlations exist between cell symmetry in the earliest cleavage stages and successful hatching rates and viability of yolk-sac larvae in the Atlantic cod, Atlantic

halibut (*Hippoglossus hippoglossus*), and turbot (*Scophthalmus maximus*) (Shields *et al.* 1997; Kjorsvik *et al.* 2003). The establishment of robust criteria to evaluate egg/embryo quality remains elusive despite the increasing importance of predicting egg/embryo quality as fish aquaculture expands worldwide to match the global ambition of increasing the supply of protein from Marine sources (Arreola *et al.* 2015).

Understanding the molecular basis of normal fish egg/embryo development is crucial to identify markers for egg/embryo quality monitoring in marine aquaculture. Most of the available studies of fish embryonic development have used the non-marine model teleost, the zebrafish (*Danio rerio*) (Haga *et al.* 2008; Yilmaz *et al.* 2018; Yilmaz *et al.* 2017). Vitellogenin is a dominant component of the egg yolk in vertebrates and so its potential importance for egg viability and quality has been assessed (Sullivan & Yilmaz 2018). In fish maternal vitellogenins (Vtg) are secreted by the liver into the bloodstream and when they reach the ovary are taken up by developing eggs and are integrated into the egg yolk (Hara *et al.* 2016; Murakami *et al.* 2019). Marine teleosts possess three isoforms of Vtg, VtgA, VtgB, and VtgC, which are the orthologues of Vtg type I, II and III, respectively in zebrafish (Rawat *et al.* 2013; Yilmaz *et al.* 2021). CRISPR/Cas9 (clusters of regularly interspaced short palindromic repeats/CRISPR-associated protein 9) knock-out assays in zebrafish revealed that Vtg type I and III play a crucial role in regulating normal egg ontogeny and quality (Yilmaz *et al.* 2021). As the egg matures, Vtgs are taken up from the blood by developing oocytes via clathrin mediated endocytosis and then *in ovo* proteolysis by cathepsin D (CatD) generates the primary yolk proteins in the early endosome (Carnevali *et al.* 1999; Carnevali *et al.* 2006; Mosconi *et al.* 2002; Romano *et al.* 2004). In marine fish that spawn pelagic eggs a second selective proteolysis of yolk proteins regulated by cathepsins B (CatB) and C (CatL) occurs to generate free amino acids (FAA) that induce oocyte hydration and promote egg buoyancy (Finn & Kristoffersen 2007; Matsubara & Koya 1997; Craik & Harvey 1987; Greeley *et al.* 1991; Thorsen & Fyhn 1991; Matsubara *et al.* 1999; Fabra *et al.* 2005).

Fertilization of the egg leads to zygote formation and embryogenesis begins and the main stages and processes are common across teleost fish although the timing of the main events is variable and depends on the species and type of egg (Figure 1.2). Teleost fish produce two main types of eggs: pelagic and demersal. Pelagic eggs are non-adhesive, and the eggs are buoyant due to the presence of an oil droplet, and they tend to be smaller in size than demersal eggs. Demersal eggs possess an adhesive membrane and have a thinner egg envelope. An early study describing the developmental morphology during the ontogeny of pelagic and demersal fish embryos identified differences in the egg chorion and yolk (Lonning *et al.* 1988).

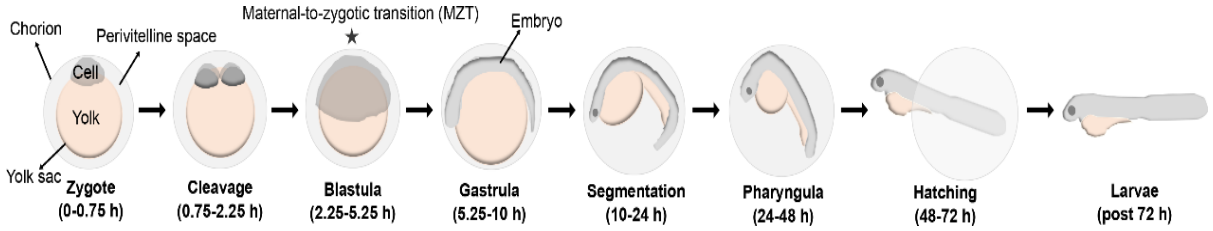


Figure 1.2. Principal embryonic developmental periods in teleost fish using the well-characterized non-marine zebrafish as the model. The embryonic development of zebrafish takes 48-72 hours post fertilization (hpf) and can be roughly divided into six principal periods: zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching. The Zygote stage represents the one cell stage after fertilization; the Cleavage stage is the period where rapid cell division occurs to generate the multicellular organisms; the Blastula stage is when the blastocoel is formed and it is during the mid-blastula stage when the maternal-to-zygote transition (MZT) occurs (denoted by a star); the Gastrula stage begins with the 512-cell stage (50%- epiboly stage) and is when the embryo is reorganized into a multilayered and multidimensional structure; Segmentation is when the somites are formed that give rise to the vertebrae and muscle of the trunk and tail; the Pharyngula stage is when the embryo develops features typical of vertebrates, such as the beginning of a spinal cord; the Hatching stage is when the embryo hatches out of its protective outer envelope. The figure was drawn based on (Kimmel *et al.* 1995; Van Eeden *et al.* 1998). Although the timing of each developmental stage is species-specific, the overview of the developmental processes in zebrafish is applicable across other fish species.

In deep-water elasmobranchs, like skates (*Bathyraja brachyurops*, *Bathyraja macloviana* and *Amblyraja doellojuradoi*) that produce demersal eggs, species-specific phenotypic characteristics related to egg pigmentation and spinulation were identified (Vazquez *et al.* 2020). As embryo develop and the zygote emerges the maternal-to-zygote transition (MZT), which is one of the most important events during early embryogenesis rapidly occurs, and is, characterized by the degradation of maternal messenger RNAs (mRNAs) and the synthesis of embryonic mRNAs. In zebrafish, MZT normally occurs at the mid-blastula stage (at approximately ten cell cycles, approximately 3.5 h post-fertilization, hpf) and this is when the zygote genome is activated (ZGA) and the heterochromatin structure is established (Kimmel *et al.* 1995; Aanes *et al.* 2011; Kane & Kimmel 1993; Lee *et al.* 2014b; Akdogan-Ozdilek *et al.* 2020; Laue *et al.* 2019). During MZT a large number of proteins are produced to support early embryogenesis (Bazzini *et al.* 2016) and this process is regulated by maternal Y-box binding proteins (Ybx1) that tightly control transcript stability and protein synthesis to avoid protein overload in the embryo (Sun *et al.* 2018; Yang *et al.* 2019) and to permit normal embryonic development.

Recently insulin-like growth factor 2 mRNA-binding proteins (Igf2bps) identified in developing zebrafish embryos were reported to be involved in the regulation of this process (Figure 1.3). The Igf2bps are a family of proteins that bind RNA in mammals and other tetrapods. These proteins regulate the cytoplasmic mRNA transition and degradation and are responsible for transcript transport along the microtubules and/or actin cytoskeleton in cells (Bell *et al.* 2013). In the zebrafish zygote, Igf2bp3 is essential for maternal RNA stability (Ren *et al.* 2020) and developmental abnormalities occurred in loss of function mutants (using CRISPR-Cas9). Notable changes in the Igf2bp3 mutants included a stretched yolk syncytial layer (YSL) and dysregulation of translation (Vong *et al.* 2021).

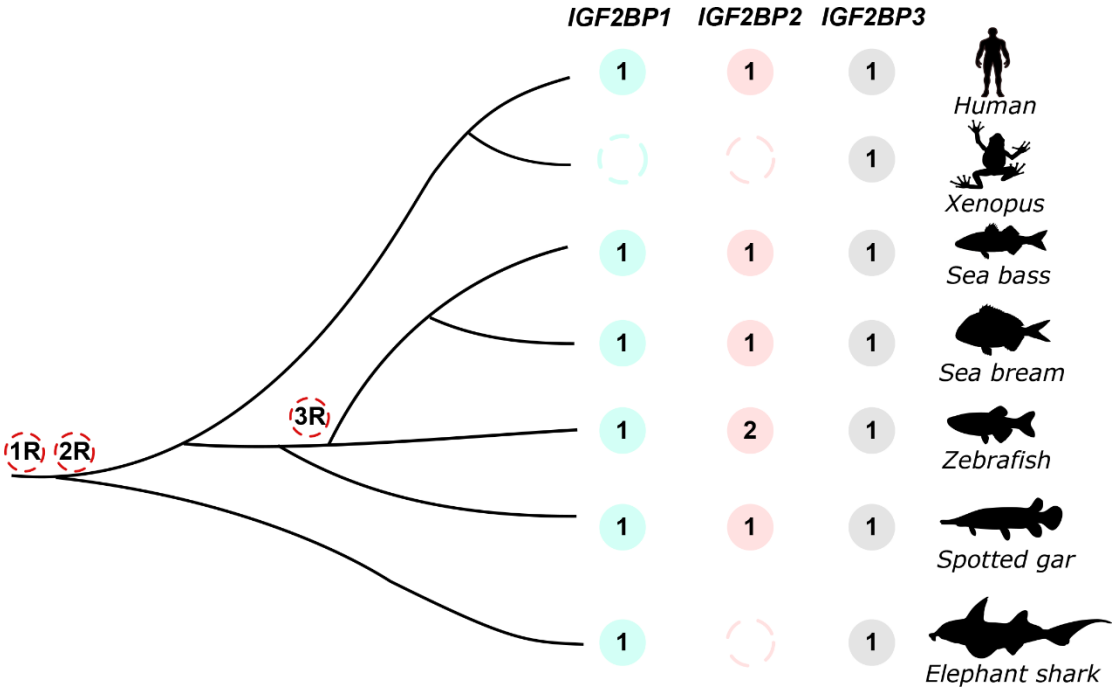


Figure 1.3. The vertebrate members of *IGF2BP* gene family. The numbers inside the circles indicate the number of genes identified in each species obtained by searching the species-specific genomes using the human predicted protein sequences in the ENSEMBL database (https://www.ensembl.org/). The genome duplication events that occurred earlier in the vertebrate radiation (1R and 2R) and the teleost specific genome duplication event (3R) are represented by circles. No sequence hits for *IGF2BP1* and *IGF2BP2* were retrieved from the Xenopus genome, and no orthologue of the human *IGF2BP2* gene was retrieved from the genome of the cartilaginous fish Elephant shark and these are represented by dashed circles. In zebrafish two *Igf2bp2* gene paralogues (*Igf2bp2a* and *Igf2bp2b*) were found, and they are localized in different chromosomes and are likely to have resulted from 3R. In other teleosts only one gene copy of *Igf2bp1, Igf2bp2* and *Igf2bp3* were retrieved suggesting that the duplicate *Igf2bp2* was eliminated from the genome. The dendrogram represents the evolutionary relatedness between the species represented and was drawn based on (Felix *et al.* 2015).

Other factors identified that influence fish egg quality, and embryonic and larval development as well as fish gametogenesis, spawning, and fertilization in an aquaculture

environment are the genetics and the physiological status of the broodstock (intrinsic factors) and the environment (extrinsic factors) (Bonnet *et al.* 2007a; Bobe 2015b). Genetics and physiological factors are commonly associated with heterogeneous egg quality and include, early broodstock spawning (Siitonen & Gall 1989), the females endocrine status (Brooks *et al.* 1997), and maternal dietary stress (Barnes *et al.* 2003). Environmental factors that are proposed to influence larval quality during aquaculture production are numerous, but strong experimental proof of causal links is frequently missing. Existing studies that report environmental effects on development include water salinity and temperature in the European smelt (*Osmerus eperlanus*) (Keller *et al.* 2020), temperature in Atlantic cod and European sprat (*Sprattus sprattus*) (Nissling 2004), photoperiod (Bobe 2015b) in catfish (*Heterobranchus bidorsalis* x. *Clarias gariepinus*) (Adebayo 2018) and Persian sturgeon (*Acipenser persicus*) (Kazemi *et al.* 2020), dissolved oxygen in water (Servili *et al.* 2020) in channel catfish (*Ictalurus punctatus*) (Torrans & Steeby 2008), and heavy metal (copper and cadmium) exposure in barbel (*Barbus barbus*) (Witeska *et al.* 2010).

Although biochemical and transcriptomic methods possess the advantage of high accuracy and sensitivity, they are inappropriate for routine use in hatchery monitoring since they are time-demanding, generally require a high level of technical competence and are costly (Ienaga *et al.* 2021). Approaches are needed to determine egg/embryo quality that are efficient, easy to use, give quick and robust results and are cost effective. In this review a comprehensive and updated consideration about the state of the art in relation to egg quality and embryonic development and the potential role of maternal factors is given building on ever increasing amounts of omics data. The species of focus are those that are used for aquaculture although it should be noted that many of the results with advanced genetic models derived with CRISPR/Cas9 are generally obtained in zebrafish. Considering the variability of fish genomes (eg. Gene copy number, gene homologues, functional homologues, etc) and the diversity of life history strategies of teleost fish caution is required when making between species inferences or generalizations.

1.3. **Egg**/embryo **composition and development**

1.3.1. Composition and egg/embryo quality

From a structural perspective, the yolk and eggshell proteins are major constituents of the egg in fish and other oviparous vertebrates (Arukwe & Goksøyr 2003). In finfish, ripe fertilized eggs are characterized by the presence of, ***a yolk*** surrounded by the vitelline membrane, the ***perivitelline space***, the egg membrane covers the egg/embryo and possess up to two to three layers (Table 1.3) the hard outer ***chorion or egg envelope***, and an ***outer gelatinous layer***. The

yolk consists of the yolk mass, which is rich in Vtg, yolk proteins (YPs) and lipids, and in fertilized eggs, cells undergoing cleavage (Woynarovich & Horváth 1980; Hiramatsu *et al.* 2015). The yolk mass plays a critical role in providing protein- and lipid-rich nutrients that support embryonic and larval development. The ***perivitelline space*** surrounds the yolk mass and consists of multilamellar envelopes and the perivitelline fluid containing dissolved proteins and has protective, nutritive and buoyancy functions (Laale 1980). In the Senegalese sole (*Solea senegalensis*) analysis of zygotes and larvae with a yolk sac revealed they contained an abundance of neutral lipids, which gradually decreased as development progressed (Vázquez *et al.* 1994). The egg envelope is composed of one to three layers depending on the species (Woynarovich & Horváth 1980) (Table 1.3).

Table 1.3. Number of fertilized egg envelopes in different fish species.

Species	Family	Type	Egg layer	References
Danio rerio	Cyprinidae	Demersal	3	(Joo & Kim 2013)
Cyprinus carpio	Cyprinidae	Demersal	3	(Wang *et al*. 2022)
Ctenopharyngodon idellus	Cyprinidae	Semi-pelagic	2	(Wang *et al*. 2022)
Ancistrus cirrhosus	Loricariidae	Demersal	3	(Kim 2020)
Corydoras adolfoi	Callichthyidae	Demersal	2	(Choi *et al*. 2019)
Corydoras sterbai	Callichthyidae	Demersal	2	(Choi *et al*. 2019)
Cichlasoma nigrofasciatum	Cichlidae	Demersal	2	(Deung *et al*. 1997)
Symphysodon aequifasciatus	Cichlidae	Demersal	2	(Deung *et al*. 1997)
Cichlasoma severum var.	Cichlidae	Demersal	2	(Deung *et al*. 1997)
Cichlasoma managuensis	Cichlidae	Demersal	2	(Kim *et al*. 2009)
Trichogaster trichopterus	Belontiidae	Pelagic	2	(Kim *et al*. 1999)
Trichogaster leeri	Belontiidae	Pelagic	2	(Kim *et al*. 1999)
Trichogaste rtrichopterus trichopterus	Belontiidae	Pelagic	2	(Kim *et al*. 1999)
Melanotaenia praecox	Melanotaeniidae	Demersal	2	(Sohn & Kim 2021)
Nothobranchius foerschi	Nothobranchiidae	Demersal	2	(Kwon *et al*. 2017)
Nothobranchius rachovii	Nothobranchiidae	Demersal	2	(Kwon *et al*. 2017)
Nothobranchius guentheri	Nothobranchiidae	Demersal	2	(Kwon *et al*. 2015)
Nothobranchius patrizii	Nothobranchiidae	Demersal	2	(Kwon *et al*. 2015)

From a biochemical perspective, fish embryos are composed of carbohydrates, lipids, and proteins (Huang *et al.* 2021b). Analysis of the composition of the unfertilized eggs in three tuna fish species, skipjack (*Katsuwonous pelamis*), tongol (*Thunnus tonggol*) and bonito (*Euthynnus affinis*), revealed they were composed mainly of water (72.2–73.0%), protein (18.2-20.2%), lipid (3.4–5.7%), and ash (1.8–2.1%) (Intarasirisawat *et al.* 2011; Yoon *et al.* 2018). The general biochemical composition of teleost fish eggs/embryos is similar, although species specific variations occur from both a qualitative (types of proteins, lipids, etc) and quantitative

perspective (Cejas *et al.* 2004; Sargent *et al.* 1989) (Figure 1.4). The substances found inside the yolk sac and in the vitelline fluid are of functional importance for normal embryonic development, although the requirements of each fish species may be variable (Grove & Wourms 1991). For example, during turbot development proteins and carbohydrates in the fertilized egg are preferentially catabolized (Planas *et al.* 1993), whereas in rainbow trout (*Oncorhynchus mykiss*) fertilized eggs, proteins, lipids, and carbohydrates are mobilized to meet energy needs (Boulekbache 1981).

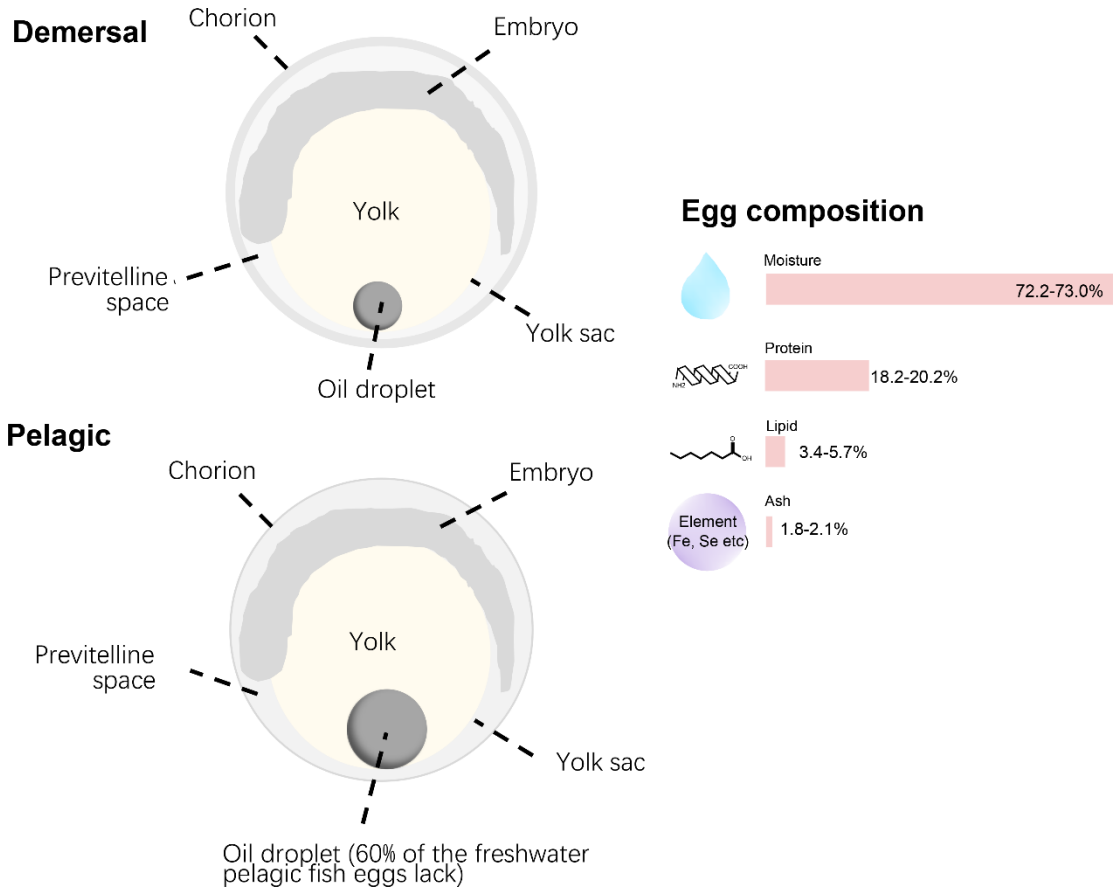


Figure 1.4. General representation of the organization and biochemical composition of a pelagic fish fertilized egg. The figure represents the fertilized egg during the Gastrula period (adapted from (Kimmel *et al.* 1995)) and the main characteristics of pelagic and demersal eggs were obtained from (Lonning *et al.* 1988; Baras *et al.* 2018). In general, teleost fish demersal eggs possess a thicker envelop and the oil droplet is smaller compared to pelagic eggs. The hard outer chorion layer or egg envelope is composed of one to three layers depending on the species. An oil droplet (ranging from one to many) is a characteristic of some fish eggs (60% of freshwater pelagic fish eggs lack an oil droplet (Baras *et al.* 2018).

Taking into consideration the critical importance of the composition of the egg for subsequent embryogenesis there have been numerous studies about this since the 1990's and they have attempted to link the composition with egg quality (Baki *et al.* 2021; Jia *et al.* 2014). To date, parameters regarded as putative indicators of egg/embryo quality (Gimenez *et al.* 2006) include their lipid content and composition (Sargent 1995; Bell & Sargent 2003; Tveiten *et al.* 2004), the amino acid content and composition (Rønnestad & Fyhn 1993) and vitamin composition and contents (Rønnestad *et al.* 1997; Maeland *et al.* 2003). However, despite the existence of studies investigating egg/embryo composition the link with embryo quality is still tenuous (Kjorsvik *et al.* 1990; Fraser *et al.* 1988; Furuita *et al.* 2000; Henrotte *et al.* 2008; Lavens *et al.* 1999).

1.3.2. Early fish ontogeny and embryo quality

The embryonic development of teleost fish occurs at variable rates according to the species under consideration (Figure 1.5). A comparative analysis of 9 species used for research and/or aquaculture clearly highlights the specific developmental characteristic of each species and the differences found between them and may explain why the identification of universal quality indicators for fish eggs/embryos is challenging (Supplementary Table 1.1 annex I, (Kimmel *et al.* 1995; Cucchi *et al.* 2012; Kamacı *et al.* 2005; Firat *et al.* 2005; Iwamatsu 2004; Hall *et al.* 2004; Velsen 1980; Sahin *et al.* 2008; Çiftci *et al.* 2002)). Based on a comparative analysis of the main developmental processes from the 4-cell stage to hatch across species it is possible to group them according to their similarity in development across time. For example, the Cypriniform, zebrafish and the Perciformes, gilthead sea bream and striped sea bream (*Lithognathus mormyrus*) have a short developmental timeline from fertilization to hatch. In contrast, the Salmoniformes such as, Sockeye Salmon (*Oncorhynchus nerka*) has a long developmental timeline and from fertilization to hatch it takes 1300 hours (Velsen 1980). The European sea bass (Eupercaria *incertae sedis*) and the Pleuronectiformes, turbot and European flounder (*Pleuronectes flesus luscus*), share a similar developmental timeline. Between pelagic and demersal eggs from teleosts no obvious differences were identified in the developmental timeline and overall developmental characteristics from fertilization to hatch. Although an inverse correlation is observed between the incubation temperature and the developmental timeline to hatch of teleost embryos (Supplementary Table 1.1 annex I). Information about the

normal developmental timeline for different fish species in aquaculture is valuable since it provides insight into normal development. If the developmental ontogeny changes under controlled abiotic conditions this may be an indicator of embryo quality and probable larval performance (Ma *et al.* 2019b; Finn *et al.* 2002).

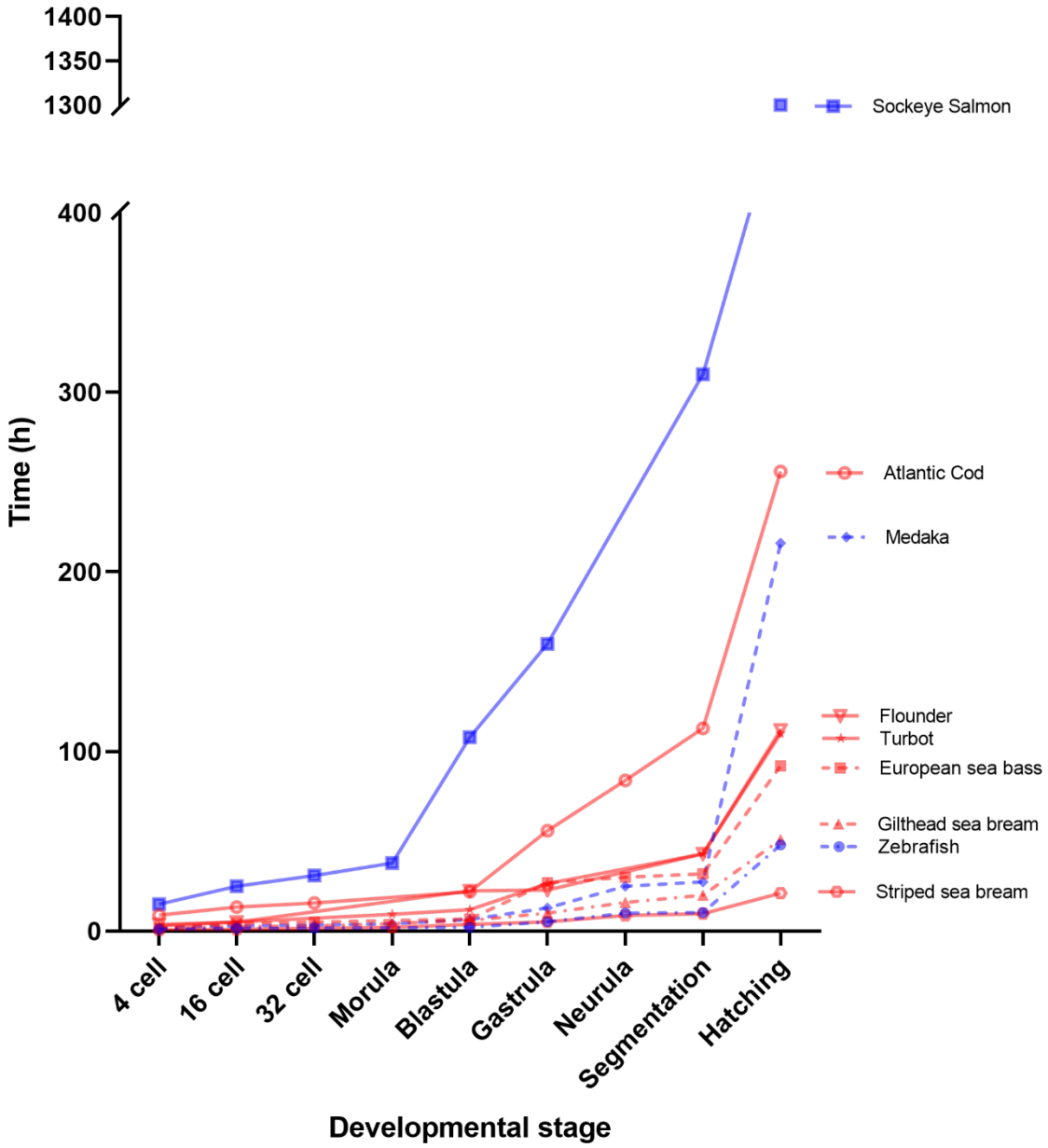


Figure 1.5. Developmental timeline showing the ontogeny of the main embryonic stages in several teleost species. The development of embryos from pelagic (in red) and demersal fish eggs (in blue) across time is indicated. A notable aspect is the extremely long time to hatch in the Sockeye salmon. The data sources for the figure are indicated in Supplementary Table 1.1 annex I.

1.4. Embryo quality characteristics

1.4.1. Egg/embryo morphology and quality

Studies of embryogenesis usually encompass zygote to hatch (Sutin & Tina 2020;

Fujimoto *et al.* 2004). Embryogenesis initiates when the zygote is formed and in fish during this process, high quality fish ovum are fertilized and if normal development proceeds a viable larva is formed (Cheung *et al.* 2018b). Nevertheless, morphological abnormalities can occur at each developmental stage in all developing fish species and is a cause of embryonic mortality and if the incidence is high the eggs/embryos are classified as poor quality.

Normally, malformations during embryogenesis are infrequent and occur in relatively few embryos and larvae although exposure to toxicants (pollution), and inadequate nutritional composition, physical damage, genetic disorders, or infections and disease reduce embryo viability (Wosnick *et al.* 2019). A low level of embryo mortality is normal because of the heterogeneity of embryo quality in any spawned batch of eggs and a number of studies have aimed to distinguish "background" mortality from mortality caused by external factors. The factors and characteristics associated with mortality in egg batches that have an overall low incidence of egg mortality under normal environmental conditions may yield potential markers for embryo quality assessment. For example, irregular first cleavage and gastrulae with loose cell aggregates were the main changes in fertilized eggs that failed to develop normally in batches of good quality flounder and cod embryos (Vonwesternhagen *et al.* 1988). Furthermore, in poor quality cod embryos, characteristic cellular abnormalities were observed as early as blastulae (Kjørsvik 1994). A change in embryogenesis is a common feature of poor-quality eggs and in the triploid catfish (*Clarias gariepinus*) abnormal embryos in a batch developing normally were characterized by early mitotic divisions in some of the progeny (Hassan *et al.* 2018). Similarly, in the triploid Black Sea turbot (*Psetta maxima*) abnormally developing embryos shared similar features to those described for triploid catfish (Aydın & Okumuş 2017). This abnormal development has been attributed to the deleterious "triploidization effect" resulting from the heat or cold shock applied to induce the triploid condition (Aydın & Okumuş 2017; Hassan *et al.* 2018).

In the sharpsnout sea bream (*Diplodus puntazzo*) and gilthead sea bream the shape of the lipid vesicle was proposed as a biomarker to assess embryo quality since abnormalities were correlated with larval survival rates (%) (Lahnsteiner *et al.* 2008). Thus, in these 2 Sparids changes in the shape of the lipid vesicle during embryogenesis, was proposed as a potential reference for embryo quality evaluation. Another study reported abnormal embryos in hybrid offspring of catfish, suggesting they might arise because of postzygotic isolation (Amini *et al.* 2007; Okomoda *et al.* 2017). It seems likely that the conditions used to establish polyploid and crossbreeding species means that the molecular and cellular processes leading to abnormal embryos are likely to be different from those causing poor quality in "nonmanipulated"

embryos. Therefore, caution is required when proposing or applying generalized quality markers to teleost embryos when they come from embryos that have undergone manipulations to generate polyploidy or species hybrids.

1.4.2. Physicochemical characteristics and embryo quality

1.4.2.1. Egg buoyancy

Although the most reliable measurement of embryo quality is still fertilization success and subsequent zygote development, such quality measures may not represent an optimal indicator for aquaculture production since they cannot be used as predictive indicators of embryo quality (Bobe & Labbe 2010; Bizuayehu *et al.* 2019). Objectively, the physical appearance of eggs/ embryos, and there buoyancy provide a good general reference parameter for fish embryo quality (Bobe & Labbe 2010). Egg buoyancy, is defined as the difference in the specific gravity of the environmental water (seawater or freshwater) and the egg (Sundby 1991). Water salinity is a determinant factor for egg buoyancy in marine fish eggs and influences their vertical distribution in the water column and consequently their dispersal from spawning areas (Sundby & Kristiansen 2015).

Egg buoyancy is a noticeable characteristic in reproductive strategies, especially those of pelagic fish, and is relevant for overall egg/embryo quality and embryonic viability (Palomino *et al.* 2021). Marine egg buoyancy is acquired by hydration during selective proteolysis of yolk proteins. In the case of freshwater fish eggs/embryos, the buoyancy is proposed to depend on the oil droplets. However, a puzzling aspect in relation to lipids and egg buoyancy is that a large proportion of freshwater pelagic fish eggs (60%) lack an oil droplet (Baras *et al.* 2018). Recent studies have assessed the relative contribution of the oil droplet and hydration for buoyancy in pelagic eggs of freshwater fish and revealed a negative correlation between oil droplets and the size of the perivitelline space and a positive correlation with parental care and habitat occupancy, suggesting alternative strategies exist in relation to egg buoyancy in freshwater teleost fish (Chen *et al.* 2021). Nonetheless evidence indicates that the efficiency of hydration is tightly correlated to buoyancy, and thereby influences the survival and development of teleost fish eggs and embryos (Adlandsvik *et al.* 2001; Seoka *et al.* 2003; Jung *et al.* 2012; Jung *et al.* 2014). Egg hydration occurs when the egg comes into contact with water during spawning and occurs via an aquaporin mediated mechanism and is an essential process that is still not fully understood (Cerda 2009; Cerda *et al.* 2013).

Under aquaculture conditions, the loss of buoyancy in early embryos is correlated with high mortality and limits the success of the hatchery stage of pelagic fish (Carnevali *et al.* 2001b;

Sawaguchi *et al.* 2006; Sundby & Kristiansen 2015). A study to establish the link between egg buoyancy and mortality in the yellow-tail kingfish (*Seriola lalandi*) studied markers of apoptosis (Palomino *et al.* 2021). The study revealed that poor buoyancy eggs/embryos of the yellow-tail kingfish had increased expression of pro-apoptotic genes (*bax-like*, *casp9*, *casp8*, and *casp3*), increased mortality and massive embryo cell death. The correlation between egg buoyancy and fertilized egg quality was first reported in the 80's for the European sea bass (Carrillo *et al.* 1989; Brooks *et al.* 1997) and subsequently for the Atlantic cod (Saborido-Rey *et al.* 2003) and the yellow-tail kingfish (Palomino *et al.* 2021). Part of the difficulty of using egg/embryo buoyancy as an indicator of quality is the heterogeneity that exists in the buoyancy of good quality embryos in batches produced by the same broodstock.

1.4.2.2. Egg stickiness

Stickiness is a crucial property for some teleost fish eggs/embryos and has been considered as a potential factor determining the quality of such eggs/embryos. The organization of the outer layers of fish eggs explains their stickiness, which usually develops when the egg contacts with water. Fish oocytes possess up to three layers, an outermost follicular layer, a median zona radiata, and an inner oolemma or oocyte plasma membrane (Shabanipour & Heidari 2017). The zona radiata, is synonymous with the terms, vitellin membrane, chorion, chorion vitellin membrane and zona pellucida (ZP) (the latter in mammals) (Hosokawa 1985). The zona radiata is composed of a complex extracellular matrix (Rizzo *et al.* 2002) of two layers, an inner protein and carbohydrate-rich layer (Murata *et al.* 1997; Scapigliati *et al.* 1999) and an outer layer that interacts with the environment and confers adhesiveness to the egg and is rich in mucopolysaccharides and glycoproteins (Bazzoli & Rizzo 1990; Rizzo & Bazzoli 1991; Bazzoli 1992; Wallace & Selman 1981). Compared to pelagic fish eggs, demersal eggs possess a chorionic membrane with a thicker and more complex structure (Stehr & Hawkes 1979), that can adhere to different substrata due to physicochemical changes in the outer layer of the zona radiata (Laale 1980; Riehl & Patzner 1998), when the eggs come into contact with water (Woynarovich & Horváth 1980; Yamagami *et al.* 1992). Understanding the importance and necessity of the egg outer adhesive layer for development and hatching success is relevant for aquaculture as egg adhesiveness makes their management more difficult. Studies of the rabbit fish (*Siganus randalli*) revealed that embryo development was more successful when they were attached to plastic plates (Collins & Nelson 1993; Komar *et al.* 2004), indicating that in this species egg stickiness is important for normal development. However, in the wolffish (*Anarhichas lupus*) embryo development was still successful after removal of their adhesive

outer layer when the eggs/embryos were incubated in upwelling incubators (Pavlov & Moksness 1995). Furthermore, in the common carp (*Cyprinus carpio*) removal of the adhesive outer egg layer increased the success of fertilization and hatching by diminishing egg agglomeration in egg incubators (Al-Bachry 2018) (Marimuthu 2019). In the winter flounder (*Pseudopleuronectes americanus*) a gene has been identified that shares high sequence identity with genes encoding a protein of the zona pellucida in the rabbit and mouse. The encoded protein contains several repeated sequence motifs, which are proposed to be associated with mechanical strength or to the adhesive properties of the egg membrane (Lyons *et al.* 1993). Since in hatcheries egg stickiness is an undesirable property as it makes management of eggs/embryos more difficult, studies about factors that determine adhesiveness are needed to establish strategies to minimize this characteristic. However, since the link between the adhesiveness of fish eggs and quality remains, unclear further studies are required to establish species-specific egg adhesiveness and how it is linked to quality before implementing strategies to control it.

1.4.3. High-throughput analytical technologies and embryo development

Proteomics and transcriptomics have recently been applied to decipher the molecular dynamics of fish embryogenesis and a myriad of transcripts and proteins have been identified and have the potential to be fundamental molecular tools for embryo quality assessment in aquaculture research (Table 1.4). Most available omics studies are centered on the embryonic development of the model species, zebrafish, a small freshwater species that inhabits warm water (average 28 ºC) and has a short lifespan. Studies of the zebrafish can provide important insights into transcripts and proteins of potential importance in teleost fish, but their habitat, lifecycle and evolutionary origin means that results cannot always be directly transferred to aquaculture species. In fact, the cypriniforms to which the zebrafish belong emerged around 193 million years ago and the approximately 4,200 species evolved primarily in freshwater (Tao *et al.* 2019). To date transcriptome and proteome studies are available for relatively few commercial aquaculture species, which makes identification and selection of putative quality biomarkers that are suitable for multiple species difficult. The diversity of fish species used for aquaculture worldwide (622 species in 2018) (FAO 2020), their taxonomic diversity and habitat specific adaptations means that caution is needed when making generalizations. Furthermore, the diversity of aquaculture species makes the outcome of knowledge transfer between species uncertain and explains in part the slow progress made in identifying universal embryo quality biomarkers.

Table 1.4. Available omics studies (proteome and transcriptome) developed for the study of fish embryogenesis.

Fish	Sample	Sequencing approach	Reference
PROTEOME			
Zebrafish (*Danio rerio*)	24 hpf	SDS-PAGE and HPLC−ESI−MS/MS (2D gel)	(Shaik *et al.* 2014)
	1 cell (0.5 hpf), 16 cell (1.5 hpf), 32 cell (1.75 hpf), oblong (3.75 hpf), Bud (10 hpf) stages	LC–MS-based shotgun proteomics analysis	(Purushothaman *et al.* 2019)
	Zygote prior first division	LC-MS/MS	(Yilmaz *et al.* 2021)
	Early blatstula (~256 cell) stage (~2–3 hps), at the shield to 75% epiboly stages (~8 hps), at the early pharyngula stage (~24 hps), and during the hatching period at 48 and 72 hps (long-pec to protruding-mouth stages)	Label-free protein quantification: SDS-PAGE and HPLC−ESI−MS/MS	(Yilmaz *et al.* 2017)
Hapuku (*Polyprion oxygeneios*)	8-cell stage	iTRAQ method	(Kohn *et al.* 2015)
Sterlet (*Acipenser ruthenus*)	Fertilized egg	Label-free protein quantification	(Niksirat *et al.* 2017)
Rainbow trout (*Oncorhynchus mykiss*)	50 min and 200 min after fertilization	label-free protein quantification: SDS-PAGE and LC−MS/MS	(Niksirat *et al.* 2020)
Longfin yellowtail almaco jack (*Seriola rivoliana*)	Fertilized egg	2D electrophoresis and MS	(Serrano *et al.* 2020)
Common carp (*Cyprinus carpio*) and grass carp (*Ctenopharyngodon idellus*)	1 hpf	Label-free protein quantification	(Wang *et al.* 2022)
TRANSCRIPTOME			
Zebrafish (*Danio rerio*)	unfertilized egg, 3, 4.5, 6, 7.7, 9, 10.7, 12, 15, 24, 30, and 48 hpf	Microarray (GenePix 4000B microarray scanner)	(Mathavan *et al.* 2005)

	1-cell (0.75 hpf), 16-cell (1.5 hpf), 512-cells (2.75 hpf) and 50% epiboly (5.25 hpf) stages	SOLiD System Sequencing platform version 3 plus (Applied Biosystems)	(Vesterlund *et al.* 2011)
	unfertilized eggs; 1-cell stage; 16/32-cell stage, 128/256-cell stage, 3.5-hpf (MBT, high-oblong); 5.3-hpf (post-MBT, 50% epiboly)	SOLiD3 (ABI) platform	(Aanes *et al.* 2011)
Rainbow trout (*Oncorhynchus mykiss*)	15 dpf (eyed and yolk vascularization); 19 dpf (caudal flexing; 28 dpf (beginning of hatching)	Microarray (Scanarray 4000 Microarray Scanner)	(Xu *et al.* 2011)
Atlantic Cod (*Gadus morhua*)	2 to 8 cells stage	454 GS-FLX pyrosequencer (Roche Applied Science)	(Lanes *et al.* 2013)
European sea bass (*Dicentrarchus labrax*)	3 hpf; 3 dpf; 4 dpf	Agilent 8x60K high-density oligonucleotide Microarray (GEO platform no. GPL22152)	(Zarski *et al.* 2017)
Killifish (*Austrofundulus limnaeus*)	1–2 cell stage	Illumina HiSeq 2000	(Romney & Podrabsky 2017)

hpf: hour post fertilization; hps: hours post spawning; dpf: days post fertilization

1.4.3.1. Gene transcripts and egg/embryo quality

Mathavan *et al*. (2005) was the first to report the transcriptome of pivotal developmental stages in zebrafish embryogenesis by microarray (Mathavan *et al.* 2005). Later, zebrafish embryonic development was further studied in four predominant stages using the SOLiD System Sequencing platform for transcriptomics (Vesterlund *et al.* 2011). Subsequently, Xu *et al*. (2011) used microarray-based transcriptomics (Scanarray 4000 Microarray Scanner; Perkin Elmer, Waltham, MA, USA) to compare fast and slow developing rainbow trout embryos (Xu *et al.* 2011) and reported that fast development of rainbow trout was associated with a significantly higher expression of genes associated with cell cycle, growth, muscle contraction, and protein synthesis. In the case of the zebrafish, it was reported that a major transition in gene regulation and transcriptional activity takes place between the 512-cell and 50% epiboly stages. It was noted that before the mid-blastula transition (MBT) and initiation of zygote transcription, maternal mRNAs were important (Figure 2). The functional importance of polyadenylation of maternal transcripts prior to MBT was revealed in zebrafish (Aanes *et al.* 2011), and a large number of novel transcribed regions were identified and contributed to improve the annotation rate of the zebrafish genome.

In the marine aquaculture species, the Atlantic cod, the embryo transcriptome revealed a diversity of differentially expressed transcripts with significant KEGG pathways including the immune and endocrine system (Lanes *et al.* 2013). In the meantime, many maternal mRNAs involved in fructose metabolism, fatty acid metabolism, glycerophospholipid metabolism, and oxidative phosphorylation were validated by qRT-PCR, and proposed as putative embryo quality biomarkers in cod or other fish species (Lanes *et al.* 2013). In sea bass embryos, six differentially expressed genes (*rnf213, irf7, usp5, mem-prot, plec* and *cenpf*) were identified from transcriptome analysis of low- and high-quality embryos and proposed as candidate quality markers in fish (Zarski *et al.* 2017).

A comparison of the transcriptome between diapause and escape bound embryos of the annual killifish (*Austrofundulus limnaeus*), revealed enrichment of a variety of molecular and metabolic pathways including glycolysis and the insulin/IGF signaling pathway (Romney & Podrabsky 2017). In summary, the available data from egg/embryo transcriptome studies of a limited number of teleost species indicate that in general high levels of gene transcription and a diversity of processes are active and modulate vital developmental processes and pathways in the embryo. However, transcriptome studies of fish embryonic development and developmental quality are infrequent for aquaculture species. More research is required that links transcriptome, function, and quality in different aquaculture species to establish if universal egg and embryo

quality markers exist and can be used by the industry.

Despite the scarcity of studies with a comprehensive scope looking at global changes in the transcriptome linked to quality, several candidate gene studies exist in a range of teleost species. Although not exhaustive a brief consideration will be provided of studies that use a candidate gene approach to evaluate embryonic development in different species. In the bullseye puffer fish (*Sphoeroides annulatus*) trypsin (*tryp*) mRNA suggested to be of maternal origin was expressed in developing embryos at 24 hpf, and detectable enzyme activity was found from day 2 post-hatch when the digestive system develops (Garcia-Gasca *et al.* 2006). At the same developmental stage, when the yolk syncytial layer forms, apolipoprotein E (*apoE*) mRNA expression was correlated with the synthesis of YSL lipoprotein in the turbot (*Scophthalmus maximus*) and this gene was proposed as a marker for endogenous lipid content during the endo-exotrophic period (Poupard *et al.* 2000). Based on its pattern of expression the cathepsin L gene isoform b (*ctslb*), was proposed as an excellent marker for mesoderm and axis induction in zebrafish during egg yolk formation (Vogel & Gerster 1997). In contrast, cathepsin isoform a (*ctsla*) was expressed in cells of the yolk syncytial layer in zebrafish and killifish (*Fundulus heteroclitus*) embryos and proposed as a putative protease involved in yolk processing (Tingaud-Sequeira & Cerda 2007). Bone morphogenetic protein gene (*bmp-2*) was strongly induced (150-fold) in gilthead sea bream embryos and proposed to have a role in embryogenesis. Embryo immune defense in zebrafish was inferred from the presence of embryo lectin and enhanced, phagocytosis of microbes by macrophages (Wang *et al.* 2016c). Although all the above-mentioned genes have been reported in a diversity of teleost fish eggs and embryos, their function still needs to be determined and substantial difference in species-specific expression patterns across development exists. There is still a vast number of genes that have a vital function in the development of teleost fish eggs and embryos that remain to be identified. The availability of more straight forward methods of *in vivo* gene editing such as CRISPR-Cas9 and their adaption for use in aquaculture species mean such approaches will in the future have a crucial contribution to establish the function of factors associated with the quality of fish eggs and embryos.

1.4.3.2. Non-coding RNAs and embryo quality

Non-coding RNAs (ncRNAs) are a specific category of RNA with non-protein coding potential that function by regulating gene expression and chromatin modifications (Alexander *et al.* 2010). There are three main categories of noncoding RNA, micro-RNAs (miRNAs), long non-coding RNAs (lncRNAs), and short non-coding RNAs (sncRNA). LncRNA is classified

as RNA of greater than 200 nucleotides and is transcribed by Polymerase II and often arises from gene splicing and modification by 5' capping and polyadenylation (Kaushik *et al.* 2013; Ulitsky & Bartel 2013). However, there are relatively few studies of lncRNA to date and their function is largely uncharacterized, especially in fish. The literature that exists for fish indicates that lncRNAs are functionally correlated with development and differentiation (Basu *et al.* 2016; Wang *et al.* 2016a). A nucleotide oligomerization domain 1 (*nod1*) antibacterial and antiviral-related lncRNA (*narl*) that regulates the immune response in adult teleost fish was recently identified and acts as a competing endogenous RNA (ceRNA) for miR-217-5p and modulates Nod1 protein abundance, (Zheng *et al.* 2021). Another study on rainbow trout reported a positive correlation between 229 differentially expressed lncRNAs and overlapping, neighboring or distantly located protein-coding genes, suggesting they might have multiple regulatory functions, especially in muscle quality traits (Ali *et al.* 2018). A response to lipid metabolism and skin color during embryogenesis was found between some lncRNA and mRNA (Xu *et al.* 2019a; Luo *et al.* 2019). In fish, lncRNAs during development are proposed to have an important biological role in zebrafish embryonic stages in spite of relatively low sequence conservation (Ulitsky *et al.* 2011), which has limited the elucidation of the function by sequence analysis. To understand the contribution of lncRNA to eggs/embryo and their quality it will be important to identify and characterize these molecules by identifying structural and regulatory elements across teleost fish since the existence and importance of secondary structure within lncRNAs means direct sequence comparisons are of limited use (Wang *et al.* 2018; Johnsson *et al.* 2014). During zygote embryonic genome activation in rainbow trout a high concentration of Dicer was linked with miRNA processing and degradation of maternal mRNA (Ramachandra *et al.* 2008), while high expression of Stat3 was proposed to induce the expression of *miRNA-21* and of other miRNAs. A role for miRNA in egg/embryo development is further favored by the observation that miRNA-430 promotes deadenylation (shortening of the poly(A) tail) and clearance of maternal transcripts, both factors that regulate morphogenesis during early embryogenesis of zebrafish (Giraldez *et al.* 2006). The potential involvement of ncRNA in egg/embryo quality is an enigma and much more work is needed to understand their function so that experiments can be designed with a view to establishing their contribution to oogenesis and development post-fertilization in fish.

1.4.3.3. Eggs protein content

Proteins are the second most abundant substance after water in fish eggs and they have a pivotal role as a source of cellular energy for embryo formation, as well as larval development

(Intarasirisawat *et al.* 2011; Akarte & Mudgal 2017; Lubzens *et al.* 2017). Characterization of the changing profile of proteins during fish embryonic development may yield embryo quality-indicators that can contribute to improved management with benefits for aquaculture. Several studies using high-throughput analytical methods to establish the embryo proteome exist in different aquaculture or model fish species (see section 1.4.3). Before the advent of high throughput approaches the function of abundant individual proteins in embryo were studied. Well characterized abundant proteins studied in fish eggs/embryos using classical biochemical methods, include vitellogenin (Vtg) and choriogenin (Chg), both of which are synthesized by the maternal liver (Hara *et al.* 2016). Chg forms the inner layer of the egg envelope and influences egg envelope disruption and 3 *chg* cDNAs (*chgH*, *chgL* and *chgH minor*) have been cloned and studied in medaka (*Oryzias latipes*) liver and share high sequence homology with the zona pellucida glycoprotein (Zp) (Hara *et al.* 2016). In addition, to studies of the possible function of this protein in the egg adhesive property in the winter flounder, studies of this gene family were performed as seven Zp proteins were identified in medaka and phylogeny revealed they are sequence homologues of Chgs (Kanamori *et al.* 2003). In the gilthead sea bream, four *zp* genes (*zp1a*, *zp1b*, *zp3*, and *zpX*) have been cloned and characterized, and all transcripts are expressed in the liver and are regulated by 17β-Estradiol and cortisol (Modig *et al.* 2006; DelGiacco *et al.* 1998). Existing studies provide evidence about the diversity and functional regulation of egg envelope genes in fish. However, despite the importance of Vtg, Chg and Zp pellucida proteins for the fish egg envelope structure there are relatively few studies that have considered or shown their correlation with egg and embryo quality.

1.4.3.4. Proteomic studies

Proteomic studies of fertilized eggs/embryo have been used to identify and quantify proteins that are synthesized or eliminated during embryonic development and to identify putative biomarkers of quality. Although not numerous there are some proteomics studies of aquaculture fish embryos, but most studies are on zebrafish (Table 1.4 and 1.5). One study integrated the 24 hpf embryo proteome and transcriptome in the zebrafish (Shaik *et al.* 2014) and identified a diversity of proteins, that exhibited a moderate correlation with the transcriptome in most cellular processes. Interactome analyses further revealed that transcriptional and post-transcriptional regulatory mechanisms are of importance in connecting mRNAs and proteins during embryonic development. Subsequently, in a similar study of the egg/embryo proteome in the sterlet (*Acipenser ruthenus*) (Niksirat *et al.* 2017) significant changes in the abundance of proteins occurred in fertilized compared to unfertilized eggs and

revealed a correlation between fertilization and the release of proteins that provide the necessary microenvironment for embryonic development. Yilmaz *et al*. (2017) performed a comparative study employing quantitative proteomics with zebrafish embryos of different quality (high- and low-quality). They reported significant differences in the highly enriched Biological Processes category in good- and poor-quality eggs/embryos and identified 17 protein candidates linked to developmental potential. The proteins that were up regulated in poor quality embryos were predicted to have an interaction with the Wnt signaling pathway, suggesting dysregulation of Wnt signaling in the etiology of poor-quality zebrafish embryos. A comparative analysis of zebrafish embryos with or without the yolk revealed that the most prominent enrichment in the embryo itself was related to cellular organization, cell cycle, and control of replication and translation, and mitochondrial functions (after the deyolking procedure) (Purushothaman *et al.* 2019). Furthermore, *vtg1* or *vtg3* knockout reduced the developmental competence of zebrafish embryos and the embryo proteome profile of the mutants was similar to poor quality embryos (Yilmaz *et al.* 2021).

Table 1.5. Proteomic studies in fertilized fish eggs and embryos.

Species	Sample	Biological process	KEGG pathway	Quality candidate	Bibliography
Zebrafish (*Danio rerio*)	24 hpf	***High abundant protein:** catabolic process, translation, generation of precursor metabolites and energy, mitochondrion organization, cellular homeostasis. ***Low abundant protein:** protein modification process, anatomical structure morphogenesis, cell differentiation, multicellular organismal development, cellular component organization.	n.a.	n.a.	(Shaik *et al.* 2014)
	Early blastula	**Good quality (+):** metabolic process, primary metabolic process, protein metabolic process, unclassified, cellular component organization, or biogenesis. P**oor quality (-):** unclassified, cellular component organization, cellular component organization or biogenesis, cell cycle, vesicle-mediated transport.	**74 interacting differentially expressed proteins:** Adherins junction, Tight junction, Regulation of actin cytoskeleton, Oocyte meiosis, Gap junction	*arf5*, *rpl22*, *arf4a*, *crp3*, *pfn2l*, *fell*, *pgm1*, *fel*	(Yilmaz *et al.* 2017)
	Non-deyolked and deyolked zygote	**Non-deyolked (=):** translation, protein folding, mitochondrion organization, generation of precursor metabolites and energy. **Deyolked (=):** respiratory electron transport chain, translation, tyrosine kinase signaling pathway negative regulation of apoptotic process, transmembrane receptor protein, cell-cell signaling.	**Non-deyolked cleavage (Unique):** metabolic pathways, ribosome, biosynthesis of secondary metabolites, proteasome, cell cycle. **Non-deyolked oblong and bud Stages (Unique):** metabolic pathways, ribosome, biosynthesis of secondary metabolites, thermo-genesis, carbon metabolism	n.a.	(Purushothaman *et al.* 2019)

	Zygote	**Vtg1-KO (+):** anatomical structure morphogenesis, cell division, cytokinesis, vesicle-mediated transport, exocytosis. **(-):** cellular response to chemical stimulus, response to organic cyclic compound, response to estradiol. **Vtg3-KO (+):** localization, establishment of localization, transport, vesicle-mediated transport, anatomical structure morphogenesis. **(-):** response to estradiol, response to organic cyclic compound, monosaccharide metabolic process, carbohydrate metabolic process.	***Vtg1-KO (+):** cysteine and methionine metabolism, cardiac muscle contraction, protein processing in endoplasmic reticulum, metabolic pathways, folate biosynthesis. * **(-):** Arginine and proline metabolism. ***Vtg3-KO (+):** Protein processing in endoplasmic reticulum, cysteine and methionine metabolism, pyruvate metabolism, arginine and proline metabolism, cardiac muscle contraction. * **(-):** carbon metabolism.	**Vtg1-KO:** 45 protein candidates **Vtg3-KO:** 18 protein candidates	(Yilmaz *et al.* 2021)
Hapuku (*Polyprion oxygeneios*)	8-cell stage	n.a.	n.a.	*vtg, hsp70*, *grp78*, *eef2*, *ran*, *ids*, *pgd*	(Kohn *et al.* 2015)
Sterlet *(Acipenser ruthenus)*	Oogamete and zygote	n.a.	n.a.	*zpax*, *zpb*, *zp3*, *hsp90 beta*, *eno1a*, *zp3.3*, *psma2-A*	(Niksirat *et al.* 2017)
Rainbow trout (*Oncorhynchus mykiss*)	Zygote	Biological regulation, localization, multicellular organismal process, cellular component, organization, contraction.	n.a.	n.a.	(Niksirat *et al.* 2020)
Longfin yellowtail (*Seriola rivoliana*)	Zygote	n.a.	n.a.	*prdx*, *sodc*	(Serrano *et al.* 2020)

Common carp (*Cyprinus carpio*) and grass carp (*Ctenopharyngodon idellus*)	Zygotic envelope	Cellular process, metabolic process, single-organism process, localization, biological regulation	n.a.	*ckap5*, *flna*, *zp1x2*, *zp3-l1*, *zp3-l2* (Wang *et al.* 2022)

Note: Top five BPs and pathways were listed on the basis of comprehensively assess the statistical p-value and/or the numbers of protein enriched in pathways. (+) enriched up-regulated proteins; (-) enriched down-regulated proteins; (=) enrichment of all proteins; "*" not statistically significant. "na" not available.
Transcript names: *arf5* (ADP-ribosylation factor 5), *rpl22* (Ribosomal protein L22), *arf4a* (ADP-ribosylation factor 4a), *crp2* (C-reactive protein 2), *pfn2l* (Profilin 2 like), *fell* (Fish egg lectin like isoform X1), *pgm1* (Phosphogluco-mutase 1), *fel* (Fish egg lectin like precursor), *hsp70* (heat-shock protein 70), *grp78* (glucose-regulated protein), *eef2* (elongation factor-2), *ran* (GTP-binding nuclear protein Ran), *ids* (iduronate 2-sulfatase), *pgd* (6-phosphogluconate dehydrogenase), *zpax* (Zona pellucida glycoprotein AX), *zpb* (zona pellucida glycoprotein B), *zp3* (zona pellucida glycoprotein 3), *hsp90 beta* (heat shock protein 90 beta), *eno1a* (Enolase A), zp3.3 (Zona pellucida glycoprotein 3.3), *psma2-A* (proteasome subunit alpha type), *prdx* (peroxiredoxin), *sodc* (superoxide dismutase 1), *ckap5* (cytoskeleton-associated protein 5), *flna* (filamin-A), *zp1x2* (zona pellucida protein X2), zp3-l1 (zona pellucida protein 3-L1), *zp3-l2* (zona pellucida protein 3-L2)

In relation to aquaculture species, proteomics has been used to identify potential quality biomarkers. For example, analysis of the proteome of early-stage embryos of the hapuku wreckfish (*Polyprion oxygeneios*), identified iduronate 2-sulfatase (I2s) and generic heat-shock proteins as potential biomarkers of embryo quality (Kohn *et al.* 2015). In the fertilized egg/embryo proteome of the longfin yellowtail (*Seriola rivoliana*) several proteins were identified and proposed to be indicative of egg and larval quality including Vtg proteins (A, B, C and Ab), β-actin, peroxiredoxin, superoxide dismutase 1, alpha subunit proteasome, and keratin II (Serrano *et al.* 2020). In the rainbow trout proteomic studies of oocyte and fertilized eggs and embryos at 50, 200 min and 15 h after fertilization identified proteins that are associated with the response to stimulus and the immune system (Niksirat *et al.* 2020). Recently, the proteome of the zygote envelop in two carp species was established and revealed that five proteins associated with adhesiveness and hardness of the fish egg envelop were more highly expressed in common carp that produces adhesive eggs than in the egg envelope of grass carp (*Ctenopharyngodon idellus*) that produces semi-buoyant eggs (Wang *et al.* 2022). This was the first study to uncover the ultrastructure and protein composition of the egg envelop between two species that have eggs of different stickiness. Overall, the embryo proteome has been established for relatively few fish species. Our analysis of the studies that exist revealed that significant changes in differentially expressed proteins between stages or embryo quality were linked to metabolism, translation and cellular organization in fish embryos and preliminary evidence suggests that endocrine factors presumably of maternal origin influence development (Table 1.5).

1.5. Gene knockout and gene mutation studies (phenotype alteration)

Most gene editing assays have been performed with zebrafish and relatively few have focused on gene knockout (KO) studies linked to egg and embryo quality. KO of specific maternal vitellogenins, *vtg1* and *vtg3* genes, evoked substantial mortality in zebrafish embryogenesis (Yilmaz *et al.* 2021). This corroborates the outcome of earlier biochemical studies of fish eggs when the importance of Vtg was first appreciated (Rizzo & Bazzoli 1991; Matsubara *et al.* 1999). Recently, in medaka knock-out of the gene for the major vitellogenin receptor (*vtgr*) that has eight ligand-binding repeats reduced larval survival and the resulting phenotypes included disrupted yolk composition and diminished embryo quality further supporting the vital function of Vtg in fish embryos (Namgung *et al.* 2021). Other KOs with a phenotype in early development included the starmaker gene homozygous mutant ($stm^{-/-}$) in zebrafish that had abnormal otoliths in embryos and abnormal fiber-supported knob-like

structures in the fertilization envelope of eggs (Pachoensuk *et al.* 2021). Zebrafish KO of val-opsinA (*valopa*) and val-opsinB (*valopb*) genes had no or only partial chorion elevation, and the eggs and embryos died before hatch (Hang *et al.* 2016). Okada *et al*. (2020) found that *neu1*-KO zebrafish exhibited slightly abnormal embryogenesis with the accumulation of pleural effusion, but no embryonic lethality was observed (Okada *et al.* 2020). Forkhead box 1 (*foxr1*) is one of several transcription factors that regulate gonadogenesis and embryogenesis. KO of the *foxr1* gene in zebrafish caused significantly lower embryonic survival rates and the embryos either failed to undergo cell division or underwent abnormal cell division that culminated in growth arrest at around the mid-blastula transition and early death (Cheung *et al.* 2018c).

The miRNA430 family is one of the most highly expressed miRNAs during fish development and regulates expression of a series of genes during embryogenesis (Mishima *et al.* 2006; Mei *et al.* 2014) and zebrafish miR-430 mutants display defects during gastrulation, brain morphogenesis, somitogenesis and heart development (Giraldez *et al.* 2006) (Giraldez *et al.* 2005). In general, the results of gene knock out studies in fish are confined to zebrafish and medaka, although gene-edited Atlantic salmon (*Salmo salar*) have been generated to manipulate reproduction (Edvardsen *et al.* 2014). Gene knockout experiments on other aquaculture fish species are still rare due to the technical challenge of obtaining fertilized eggs and optimizing gene knock out procedures. To take advantage of this promising technology, more omics studies are required to pre-screen for potential gene candidates for knock out approaches initially in zebrafish although ultimately studies in important aquaculture species will be required.

1.6. Maternally inherited molecules and embryo quality

The development of the fish oocyte will not be considered in depth as it is not the main objective of this review and other extensive reviews exists (Reading *et al.* 2018; Lubzens *et al.* 2017). Nonetheless, gametogenesis almost certainly plays a role in subsequent egg and embryo quality and embryogenesis and studies have revealed that egg protein composition is indispensable for maintenance of gametogenesis. During gametogenesis, maternal proteins accumulate in eggs and are exploited at the onset of fertilization and later in embryogenesis (Cerdà *et al.* 2007). The contribution of maternal factors to embryo development now includes a catalogue of factors, including transcripts that regulate cell division and determine oocyte polarity, embryo patterning, and the transition from maternal to zygotic gene expression (Figure 1.6) (Lubzens *et al.* 2017). The success of embryogenesis is marked by successful hatching and is dependent on maternally-inherited substances deposited in the oocyte throughout oogenesis and these factors are presumably determinant for egg and embryo quality (Cheung *et al.* 2018b).

It is notable that the oocyte is transcriptionally inactive following gametogenesis and is transcriptionally repressed until ZGA, which normally occurs at the MBT in most vertebrates (Lyman & Pelegri 2007). For this reason, since maternal factors dominate embryo development up until MBT much attention has been given to characterizing the maternal contribution to eggs and embryos and, maternal RNAs that are the dominant genes expressed before ZGA (Zhao *et al.* 2017).

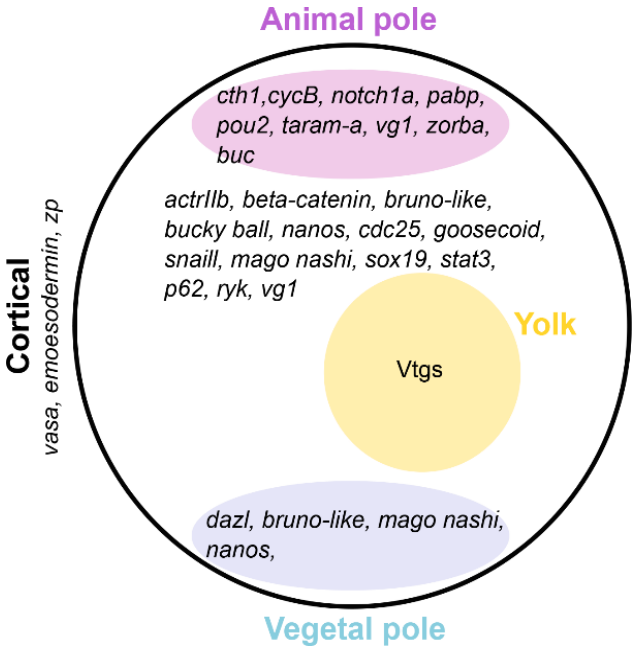


Figure 1.6. Maternally inherited transcripts and their distribution in the zebrafish embryo. Animal pole, vegetal pole, yolk and cortical (egg envelop) regions are indicated in different colors to show the localization of transcripts for each region. Data was obtained from (Howley & Ho 2000; Lyman & Pelegri 2007; Abrams & Mullins 2009). Transcript names are: *cth1* (encodes a protein with two CCCH zinc fingers), *cycB* (cyclin B), *notch1a* (encodes a member of the Notch family), *pabp* (Poly(A)-binding protein), *pou2A* (gene encoding POU domain), *taram-a* (TGFβ-related type I receptor), *vg1* (vegetalising factor-1), *zorba* (a zebrafish homologue of the orb germline gene), *buc* (bucky ball), *actRIIB* (activin receptor type-2B), *bruno-like* (zebrafish mutant involved in animal-vegetal polarity), *nanos* (Nanos2, nanos2, Nanos Homolog 2 (Drosophila)), *cdc25* (cell division cycle 25), *goosecoid* (a homeobox protein, Wnt anatagonist), *snail1* (a zinc finger protein expressed in the mesoderm in zebrafish embryos), *mago-nashi* (a conserved protein of unknown function with expression in zebrafish blastula), *sox19* (SRY (sex determining region Y)-box 19), *stat3* (Signal transducer and activator of transcription 3), *p62* (Nucleoporin p62), *ryk* (related to receptor tyrosine kinase), *dazl* (DAZL, deleted in azoospermia-like), *Vasa* (RNA binding protein with an RNA-dependent helicase), *eomesodermin* (a zebrafish gene with mesoderm inducing activity), *zp* (Zona pellucida), Vtgs (Vitellogenins).

Maternal mRNAs are stabilized for storage by specific post-transcriptional processing that generates a short poly (A) tail (approximately 15–40 nucleotides) (Ma *et al.* 2019b). In vertebrates during embryogenesis activation of maternal mRNA involves cytoplasmic polyadenylation and extension of the length of the poly(A) tail to 80 or more nucleotides

(Bachvarova 1992; Villalba *et al.* 2011; Gohin *et al.* 2014; Cabada *et al.* 1977; Richter 1999). Studies of fish fertilized eggs and embryos seem to be concordant with other vertebrates and presumptive maternal mRNAs were identified with a poly(A) tail of differing lengths (Ma *et al.* 2019b). In zebrafish a study that aimed to characterize the transcriptome dynamics from egg to early gastrulation stages revealed that maternally inherited transcripts are activated at different developmental stages and that this process depends on the length and rate of polyadenylation (Aanes *et al.* 2011). In the Atlantic cod maternal transcripts in eggs had long 3' UTRs (mean 187.1 and 208.8 bp) with alternative polyadenylation motifs that produced a higher number of 3' UTR isoforms when compared to zygote transcripts (Kleppe *et al.* 2012). Studies on zebrafish fertilized eggs revealed that the presence of a longer 3'UTR contributed to protect maternal mRNA from poly(A) tail-shortening (deadenylation), a process that involves the CCR4-NOT (carbon catabolite repression 4 - negative on TATA-less) complex in fertilized eggs and embryos, as well as codon usage in the mRNA sequence (Mishima & Tomari 2016). Recently an epigenomic study of early embryos revealed that in zebrafish maternal transcript stability during MZT was dependent on 5'-mRNA methylation (5-methylcytosine, m^5C) and that Ybx1 protein plays a key role in this process and binds to m5C by recruiting poly(A) binding protein cytoplasmic 1a (Pabpc1a) to methylated transcripts to prevent decay (Yang *et al.* 2019).

To further understand how maternal factors and more specifically how diverse forms of RNA influences embryogenesis, RNAseq has been used to compare high- and low-quality fish eggs and embryos. Previous studies on rainbow trout embryos revealed many ribosomal transcripts (40% of total) were DE between high- and low-quality embyos. This led to the suggestion that poor quality embryos have an inadequate number of ribosomes so that maternal mRNA translation is insufficient to support optimal embryogenesis and leads to mortality and poor-quality larvae (Ma *et al.* 2019b). In the Japanese eel (*Anguilla japonica*) comparative analysis of the transcriptome of good- and poor-quality embryos, identified 10 DEs that were proposed to be candidate maternal transcripts that explained the different quality of the embryos (Izumi *et al.* 2019). In Atlantic salmon embryos maternal transcripts were analyzed and Hepcidin-1 (*hamp1*) was identified as a potential marker for embryo quality (Bizuayehu *et al.* 2019). The existing studies of candidate maternal mRNA transcripts in aquaculture species are promising and provide insight into important candidate genes linked to the quality of fish embryogenesis. However, larger scale studies are required on a greater diversity of species to validate and demonstrate the practical applicability of using maternal mRNA transcripts for embryo quality monitoring in fish aquaculture.

1.7. Egg/embryo quality and immunity

The importance of eggs for species survival means that they have acquired very specific characteristics to ensure survival and fitness. This raises intriguing questions about the contribution and relative importance of maternal immune-related factors to the egg and developing embryo. One intrinsic form of protection typical of innate immunity is barrier function and the egg and developing embryo has specific mechanical barriers, like the impermeable egg envelope that defends against pathogens and external aggression (Li & Leatherland 2012). Additional protection also seems to be conferred by humoral factors, like enzymes that are typically part of the innate immune response in vertebrates. To date, a large number of genes have been identified in fish embryos that are proposed to contribute to the intrinsic pathogen response during fish embryogenesis (van der Vaart *et al.* 2012). At the initial stage of embryogenesis, as early as one day post fertilization, some zebrafish embryo cells display phagocytic activity in response to microbial infections (Herbomel *et al.* 1999). As zygote development proceeds, maternally derived proteins and mRNAs are produced with a broad functional scope (Shaik *et al.* 2014) although their importance in immune surveillance and protection is largely unstudied. The complexity of studying embryo immunity is linked to the significant change in the molecular background that occurs at ZGA when maternal RNA is degraded and substituted by embryonic RNA. Furthermore, studies directed at systematically cataloguing RNA in eggs and embryos and determining the relationship between immune-related molecules and embryonic quality are non-existent in fish. The highly dynamic nature of development means that each stage is characterized by a different developmental status and set of expressed genes and proteins and more studies are required to establish their role, if any, in fish egg and embryo immunity. The recent recognition of the contribution of the microbiome to immunity in mammals (Liu *et al.* 2019) and the presence of a specific teleost fish egg and embryo microbiome (Najafpour *et al.* 2021) highlights the need for more studies of immune protection focused on the fish embryogenesis and its associated microbiota. This is particularly important in aquaculture where large volumes of fertilized eggs are handled and maintained in a confined space making eggs and embryos particularly vulnerable to pathogens.

Studies of fish embryo immunity are mainly based on knowledge derived from immunity in juvenile and adult fish and other vertebrates and tend to focus on species-stage specific molecules that are central regulators of the immune response. Interferon gamma (IFN-γ) is one such molecule and was shown to have increased expression and to protect zebrafish embryos against bacterial infections (Sieger *et al.* 2009) and in this way may influence embryo quality. Lysozymes are an ancient group of antimicrobial enzymes of the innate immune system, and

copy number is greatly expanded in teleosts. Enzyme activity during embryo development is species-specific and changes in *lyzg1*, *lyzg2* and *lalba* expression were found between embryos and developing larvae with a different growth performance and from different gilthead sea bream brood stock (Li *et al.* 2021).

Hatching enzymes, are crucial proteolytic enzymes, that are secreted by the embryo and digest the egg envelope (Sano *et al.* 2014). These enzymes were recently suggested to be involved in immunity since they also contribute to defense when fish are infected with parasites (Saleh *et al.* 2019). We recently identified multiple forms of hatching enzyme in sea bass and sea bream embryos and proteomic analysis revealed significant changes in the enzyme protein content during hatching (unpublished data). This suggests specialization of hatching enzymes may have occurred and we hypothesize that some isoforms may confer protection against pathogens when they are present in high concentrations in early vulnerable larval stages of fish. Further work is required to assess the relative importance of hatching enzyme as a protective factor in the pre-hatching embryo and the post-hatch larvae when they are exposed to a microorganism rich environment.

Elavl1a, is a novel molecule, which functions as a pattern recognition receptor (PRR) to identify pathogen associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS) and lipoteichoic acid (LTA), but it also represses growth of Gram-positive and Gram-negative bacteria. Elavl1a was recently proposed as a maternally derived immune factor, which protects zebrafish embryos from bacterial infection (Ni *et al.* 2021). Based on the outcome of previous studies, we propose that Elavl1a may have a vital function in the defense of the egg and early embryo against many kinds of pathogens in early stages of fish development. If the immune function of Elavl1a in a range of aquaculture fish is demonstrated, it may be a good marker of embryonic quality in fish.

Several studies have used markers to improve knowledge about immune system ontogeny during fish development. Recombination-activating gene 1 (*Rag-1*) and Sacsin genes (*Sacs*) were used as markers of the physiological maturity of the immune system and nervous system diseases, respectively in the gadoid haddock (*Melanogrammus aeglefinus*), goldfish (*Carassius auratus*), grass carp, and Nile tilapia (*Oreochromis niloticus*) (Corripio-Miyar *et al.* 2007; Fan *et al.* 2009; Zhang *et al.* 2009; Elkatatny *et al.* 2020). Furthermore, these two genes have different expression patterns during embryogenesis and early post-embryonic development in the loach (*Misgurnus anguillicaudatus*) (Lee *et al.* 2015). Cytokines, crucial immune related proteins, have been suggested as potential markers for embryo quality and to regulate embryonic development in zebrafish and expression of Interleukin-1 beta (*il1b*) and tumor

necrosis factor b (*tnfb*) were detected in the zygote (0.2 hpf) and blastula (4.3 hpf) stages before emergence of the fish immune system (24 hpf) (Ito *et al.* 2008). The identified cytokines in zebrafish embryos may be useful indicators of embryonic immune status and the developmental stage in species used for aquaculture although experimental validation is required. A study in whole carp embryos revealed that specific transcripts of Igs (Immunoglobulins) occurred in the mucosa-associated lymphoid tissue (MALT) at 4 days post fertilization (dpf), although IgM and IgZ proteins were only found 4–6 weeks later (Sahoo *et al.* 2021). This may be a promising area for future in depth studies of immunity in early developmental stages and their link with quality.

1.8. A reference criterion for evaluating "embryo quality"

Due to very limited studies and unavailable or inconsistent industry standards, to assess parameters for evaluation of embryo quality we searched for them by comparing published studies (Table 1.6) with a view to identifying common criteria for aquaculture. The majority of published studies directed at identifying embryo quality indicators are with zebrafish and although studies of aquaculture species exist, they are not numerous, and few have been replicated. Furthermore, the studies on aquaculture species are not very uniform and have used a diversity of criterion to assess embryonic quality and are at an experimental scale although some industry-based studies or research and industry scale comparative studies exist. Most of the published studies have adopted methods that combine embryonic morphology and developmental performance to pre-screen embryo quality and before biochemical or molecular analysis (Table 1.4). Despite the limitations caused by study variability and with a view to defining criterion for embryo quality classification we integrated the available results and synthesize below a working basis for embryo quality assessment:

a. Observe the fertilized egg morphology including transparency, shape of the yolk sac, and fat droplet (if the eggs are pelagic).
b. Check the buoyancy of pelagic fish eggs. Generally, high buoyancy is indicative of good quality, abnormal buoyancy, or heterogeneous buoyancy of the same egg batch is indicative of variable embryo quality.
c. Determine the developmental performance, by assessing fertilization success, eyeing rate, hatching rate, and the survival rate. By building up a database of these parameters for each aquaculture species, deviations from the median between egg batches will be indicative of quality variation.

d. Molecular markers such as gene transcripts or proteins exist for some species but are not adequately validated for routine embryo quality screening by industry.

Table 1.6. List of potential indicators of teleost fish embryo quality. The study species, origin of the fertilized eggs, phenotype estimation used for quality assessment and bioindicators proposed are provided.

Species	Type	Morphological, physical and performance assessment	Molecular markers	References
Zebrafish *(Danio rerio*)	L	Good: clear separation of yolk and cell; normal egg activation and chorion swelling; non-opaque appearance Poor: opaque, deformed, and no clear cell or chorion elevation visible	*ski7*	(Cabrera-Quio *et al.* 2021)
	L	Good: high survival rate (> 93%) at 48 hpf Poor: low survival rate (< 38%) at 48 hpf	*otulina, slc29a1a*	(Cheung *et al.* 2019)
Japanese medaka *(Oryzias latipes*)	L	Poor: abnormal fertilized, fat droplets are not assembled correctly in the vegetal pole, eggs not viable	n.a.	(Ishigaki *et al.* 2014)
Atlantic cod (*Gadus morhua*)	L & Ind	Poor: cellular malformations in early blastulae	n.a.	(Kjørsvik 1994)
European sea bass (*Dicentrarchus labrax*)	L & Ind	Good: pelagic Poor: sinking	*ctsd*	(Carnevali *et al.* 2001a)
Gilthead sea bream (*Sparus aurata*)	L & Ind	Good: pelagic Poor: sinking	*ctsd; ctsl*	(Carnevali *et al.* 2001b)
Gilthead sea bream (*Sparus aurata*) / Sharpsnout sea bream (*Diplodus puntazzoI*)	L & Ind	Poor: spherical lipid vesicles or extremely ellipsoid lipid vesicles	n.a.	(Lahnsteiner & Patarnello 2005)
Atlantic halibut *(Hippoglossus hippoglossus*)	L & Ind	Good: fertilization success $\geqslant 90 \pm 2\%$ and hatching success $\geqslant 85 \pm 2\%$	*irf7; mhc2a*	(Mommens *et al.* 2014)
Rainbowtrout *(Oncorhynchus mykiss*)	L & Ind	Percentage of normal alevins monitored at yolk-sac resorption (YSR)	*phb2*	(Bonnet *et al.* 2007b)
	L & Ind	Good: eyeing rates over 80%. Medium: eyeing rates 30–50%. Poor: eyeing rates < 5%.	50 DEGs	(Ma *et al.* 2019b)
Atlantic salmon *(Salmo salar)*	L & Ind	Good: higher percentages eyed embryos and survival at yolk-sac resorption Poor: lower percentages eyed embryos and survival at yolk-sac resorption	*hamp1*	(Bizuayehu *et al.* 2019)
Japanese eel (*Anguilla japonica*)	L & Ind	Good: fertilization rate > 80% and hatching rate >80%. Poor: fertilization rate >80% and hatching rate $\leqslant$20%.	*dnajb4, gnpat, card14, pdp1, fcgbp, ttn, b4galnt1, acsl6, rtkn, trim24*	(Izumi *et al.* 2019)
Pikeperch *(Sander lucioperca*)	L & Ind	Good: fertilization rate and survival rate > 80%, hatching rate > 75%. Poor: fertilization rate > 80%, survival rate < 40%, hatching rate < 35%.	*smarca4*, *h3f3a*	(Zarski *et al.* 2021)

L – Laboratory, Ind – industry. n.a.. – Not applicable, DEGs – Differentially expressed genes; transcript name: *ski7* (superkiller 7), *otulina* (OTU deubiquitinase with linear linkage specificity a) *slc29a1a* (solute carrier family 29, member 1a), *ctsd* (Cathepsin D), *ctsl* (Cathepsin L), *irf7* (Interferon regulatory factor 7), *mhc2a* (mhc class 2A chain), *phb2* (Prohibitin 2), *hamp1* (Hepcidin-1) *dnajb4* (dnaJ heat shock protein family (Hsp40) member B4), *gnpat* (glyceronephosphate O-acyltransferase), *card14* (caspase recruitment domain family member-14), *pdp1* (Pyruvate dehydrogenase phosphatase catalytic subunit 1), *fcgbp* (Fc fragment of IgG-binding protein), *ttn* (titin), *b4galnt1* (beta-1,4-N-acetyl-galactosaminyltransferase 1), *acsl6* (long-chain-fatty-acid–ligase 6-like), *rtkn* (Rhotekin), *trim24* (transcription intermediary factor 1α), *smarca4* (Transcription activator BRG1), *h3f3a* (Histone H3.3).

1.9. Prospects and suggestions

To date, fish ontogeny and its link with developmental quality and viability has been continuously studied and there is a better understanding of this compared to a decade ago. However, abnormal phenotypes or symptoms are very heterogeneous and frequently species specific. This means it is important to investigate and build species-specific databases integrating, phenotype, physical and molecular characteristics for use in quality assessment in the future. More than that, relevant studies of the main aquaculture fish species or representatives of a genus or family are still limited, which is a drawback since criterion for quality may not be transferable between genus or even related species. Furthermore, this means that genus or species-specific candidate biomarkers validated for quality monitoring of eggs and embryos may not exist and those developed for one genus/species may not have sufficient specificity for another genus/species. In addition, a key point to consider for quality screening of fish eggs and embryos is linked to logistics and the feasibility and effectiveness of applying potential molecular indicators to aquaculture production. Bearing this in mind a shortlist is proposed for development of an integrated tool for monitoring egg/embryo quality:

a. Build digital photo atlas's of normal and abnormal morphological alterations for commercially relevant fish species.
b. Apply omics methods such as transcriptomic, proteomic, and whole-genome association analysis to establish potential embryo quality biomarkers.
c. Further develop knowledge about fish gametogenesis, by harnessing omics and PCR approaches to enrich knowledge about molecules in oocyte and ova, to acquire an accurate picture of the maternal contribution to fish embryos.
d. Test and validate potential quality biomarkers by incorporating multiple commercial populations of the same species from different production sites and/or execute experiments relevant for functional validation, thereby improving the accuracy of the biomarkers.
e. Prioritize technological support for small and medium-sized aquaculture companies to improve management, secure welfare, minimize waste and in this way financial losses and sustainability.

1.10. Thesis objectives

The main goal of this PhD thesis is to investigate immune molecules that are potentially of functional importance in maintaining fish embryonic development. Fish are vertebrates and so possess both innate and adaptive immune defense systems. In fish, the innate immune system has been widely studied and a myriad of immune-related molecules have been identified and characterized in different species by immune challenge trials. However, in fish adaptive immunity is suggested to be less well developed compared to other vertebrates. Understanding how immunity is established and developed during fish embryogenesis may contribute to enhance comprehension of embryo quality, which is important for understanding fish developmental biology and improving aquaculture production.

The experimental fish used in the current study were chosen from the Mediterranean region, a biodiversity hotspot with high species endemism and an important source of food for approximately 500 million people who inhabit this area. In the current thesis, *Diplodus sargus*, *Argyrosomus regius*, *Sparus aurata*, and *Dicentrarchus labrax* were selected to study immunity and developmental quality during embryogenesis. It should be noted that due to the COVID-19 outbreak, I was retained in China and so continued my thesis work by focusing on complement C5 especially its molecular evolution and predicted function in an important aquaculture species the Cyprinidae fish *Ctenopharyngodon idella*. A range of approaches were deployed in my thesis including molecular evolution, comparative proteomics and transcriptomics with the aim of gaining insight into the molecular mechanism in fish early development linked to immunity and quality.

The following four specific objectives were pursued to achieve the aim:

1) Characterize the molecular evolution of the lysozyme gene family and its divergent function in fish early ontogeny to establish their contribution to fish innate immunity. This section includes phylogenetic trees constructed with amino acid sequences retrieved from the genomes of different fish including gilthead sea bream and European sea bass. Chromosomal synteny analysis unveiled specific duplication events and possible regulatory mechanisms. Quantitative gene expression and enzymatic assays were employed to identify and characterize the functional pattern of lysozyme in different developmental stages and their potential link with egg and embryo quality in sea bream and sea bass after fertilization.
2) Study the molecular evolution of complement C5 and its predicted function in fish using bioinformatic analysis. The phylogeny and sequence characteristics of complement C5 were

analyzed. Codon usage bias analysis was used by comparing C5 sequences from different species and selective pressure homologous modeling and molecular docking approaches were applied to simulate the interaction between fish C5a and the C5a anaphylatoxin receptor and to calculate the potential critical interaction site.

3) Investigate the embryonic developmental proteome of Eupercaria species (white sea bream, meagre, and gilthead sea bream). Total protein was extracted from embryos before hatch and at hatch and visual inspection of proteome profile comparisons were stablished via SDS-PAGE. SWATH-MS (Sequential Window data independent Acquisition of the Total High-resolution-Mass Spectra)-Based quantitative proteomics technology was used to map the embryo proteome in different developmental stages among the three fish species. Proteins that were significantly modified were identified and categorized via Gene Ontology-biological process and pathways. Hatching enzyme was found to be significantly modified between developmental stages and quantitative gene expression was subsequently employed to characterize the functional pattern of this enzyme between different developmental stages.
4) The gene expression profile in sea bream early embryonic development and its functional potential was established using an RNAseq approach. Differentially expressed genes were identified in embryonic samples across different stages. Gene Ontology-biological process and pathways were categorized and analyzed. The regulatory epitranscriptome represented by non-coding RNA was identified and potential gene targets were predicted. The embryonic RNAseq dataset was validated by quantitative PCR using a series of candidate genes identified in the dataset. The *Sec14l2* gene previously reported to be important in the development of the vascular system in fish was quantified and its functional potential in embryonic development and quality was assessed.

CHAPTER 2

Fish lysozyme gene family evolution and divergent function in early development

Lisen Li+, João C. R. Cardoso+*, Rute C. Félix, Ana Patrícia Mateus, Adelino V. M. Canário and Deborah M. Power*

+ *Contributed equally to the study*

Comparative Endocrinology and Integrative Biology, Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

Manuscript published in Developmental & Comparative Immunology, 2021, 114: 103772
https://doi.org/10.1016/j.dci.2020.103772

CRediT statement; LL- Investigation and methodology, data analysis, writing - original draft, DMP – conceptualization, supervision, data analysis, writing - review & editing, JRC - supervision, data analysis, writing - review & editing, RCF – methodology & data analysis, writing - review & editing, APM - methodology & data analysis, writing - review & editing, AVMC - data analysis, writing - review & editing.

Acknowledgments

The authors thank Yi-Feng Li and Jin-Long Yang (Shanghai Ocean University, China) for access to the hard-shelled mussel transcriptome and Zachary Delacqua for performing the mollusc database searches. This work was financed by the European Union's Horizon 2020 research and innovation programme under Grant Agreement no 727610 (PerformFISH) and Portuguese national funds from FCT - Foundation for Science and Technology through project UIDB/04326/2020. JCRC was supported by FCT (UIDB/04326/2020). R.F was funded by an auxiliary research contract with the Universidade do Algarve.

2.1. Abstract

Lysozymes are an ancient group of antimicrobial enzymes of the innate immune system. Here we provide a comparative analysis of the evolution and function of lysozymes during early development in fish, the most speciose vertebrate group. In fishes, lineage and species-specific evolution of both C-type (chicken or conventional) and G-type (goose type) genes occurred. Phylogenetic analysis revealed that the teleost lysozyme G-type members group with the tetrapod homologues but the teleost C-type form three different clusters with the tetrapods. Most of the teleost C-type cluster with tetrapod Lyz but there are some that group with mammalian Lyzl1/2 and LALBA. This suggests that early in gnathostome evolution these genes already existed and that *lyzl1/2* and *lalba* genes are present in fish and tetrapods. Gene synteny analysis to confirm sequence orthologies failed to identify conserved genome regions between teleosts and other vertebrate lysozyme gene regions suggesting that in the ancestral bony fish genome *lyz*, *lyzl1/2*, *lalba* and *lyg* precursor genes were transposed to different chromosome regions. The homologue of the mammalian lactalbumin (*LALBA*) gene was identified for the first time in teleosts and was expressed in skin and during embryo and larval development. Lysozyme activity was detected in teleost eggs and embryos and varied between species and in the gilthead sea bream *lyg* and *lalba* transcript abundance differed in embryos and larvae from different brood stock suggesting differences exist in maternal innate immune protection.

Keywords: lysozymes; evolution; fish; innate immunity; ontogeny

2.2. Introduction

Lysozymes are a group of non-specific innate immune molecules involved in host protection against bacterial invasion. They play an important role in protective immunity of fish, which have a less diverse specific immune response and are more reliant on innate immunity (Jimenez-Cantizano *et al.* 2008). The activity of lysozymes is increased by infection but also stress and nutrition amongst other factors (Castanho *et al.* 2017; Saurabh & Sahoo 2008; Cecchini *et al.* 2000; Magnadottir 2010).

Lysozymes hydrolyse the peptidoglycan layer of the bacterial cell wall by cleaving beta-1,4 glycosidic bonds between N-acetylmuramic acid and N-acetylglucosamine. Lysozymes act directly on the outer peptidoglycan layer in Gram-positive bacteria but cannot degrade the inner peptidoglycan layer in Gram-negative bacteria unless it is exposed by other innate immune processes (Saurabh & Sahoo 2008; Harikrishnan *et al.* 2011). The early appearance of bacteria during evolution and the conserved organization and composition of their exterior wall, together with their ubiquitous and abundant presence, probably explains why lysozymes are found from plants to animals and why, despite limited amino acid sequence similarity, lysozymes have retained a common activity and overall conserved protein structure (Grutter *et al.* 1983; Gavilanes *et al.* 1984; Weaver *et al.* 1984).

In vertebrates, two major types of lysozymes exist. One type was isolated from chicken eggs and so it is designated chicken-type or C-type lysozyme, although it has subsequently been found in a diversity of vertebrates (Irwin *et al.* 2011). The other type of lysozyme has been identified in numerous vertebrate species but since it was first isolated from goose eggs it was named goose-type or G-type lysozyme (Irwin 2014). Invertebrates express a specific lysozyme, the invertebrate-type or type-I, that was first described in the common European starfish (*Asterias rubens*) (Callewaert & Michiels 2010; Bachali *et al.* 2004; Bachali *et al.* 2002). Outside the animal kingdom other lysozyme forms have been reported in plants (Jollès 1996), bacteria (Jollès 1996) and bacteriophage (Liu *et al.* 2006).

In vertebrates, LYZ is the most common lysozyme C-type found and it has a widespread tissue distribution and has been identified in several body fluids. This enzyme is part of a larger family which includes other sequence related enzymes that have been described mostly in mammals (Nitta & Sugai 1989): i) Lactalbumin (LALBA), that has no bacteriolytic activity and is only found in mammals (Irwin *et al.* 2011), ii) a calcium-binding lysozyme which has anti-bacterial activity and has been described in a few mammals (Lysc1) and birds (Calyz) (Prager & Jolles 1996; Nitta *et al.* 1988) and, iii) LYZ-like sequences specific to mammals that were

initially isolated from the human testis (LYZL2, LYZL4, LYZL6 and Sperm acrosome associate (SPACA) 3) (Mandal *et al.* 2003; Chiu *et al.* 2004; Irwin *et al.* 2011) and others retrieved from the human (LYZL1, SPACA5 and SPACA5B) and platypus genomes (Lyzl8) (Irwin *et al.* 2011).

In humans, lysozyme G-type (LYG*)* members also have a widespread tissue distribution, but its function is not very well characterised. A tandem gene duplication event in humans originated the *LYG1* and *LYG2* genes and multiple *LYG* genes have been identified in other vertebrates as a result of a large gene family expansion prior to the divergence of birds, reptiles and mammals (Irwin 2014).

Teleosts are the most numerous and species-rich group of vertebrates and comprehension of their immune response is still rudimentary relative to tetrapods and is focused on relatively few species. In fish, lysozymes are most abundant on external surfaces in contact with the water such as the gills and the skin but they are also present in the kidney (Saurabh & Sahoo 2008). The antibacterial activity of lysozymes has been described in several teleost species, such as the Japanese flounder (*Paralichthys olivaceus*, Lyz) (Minagawa *et al.* 2001), Asian seabass (*Lates calcarifer*, Lyz and Lyg) (Fu *et al.* 2013), European seabass (*Dicentrarchus labrax*, Lyg) (Buonocore *et al.* 2014), the orange-spotted grouper (*Epinephelus coioides*, Lyg) (Wei *et al.* 2014), the brill (*Scophthalmus rhombus*, Lyg and Lyc) (Jimenez-Cantizano *et al.* 2008; Krogdahl *et al.* 2005) and the Atlantic cod (*Gadus morhua*) (Seppola *et al.* 2016). A single *lyz* gene has been described in fish with the exception of cod where the gene is absent but in contrast multiple *lyg* genes occur, some of which are believed to be intracellular (Yin *et al.* 2003; Zheng *et al.* 2007; Irwin & Gong 2003; Irwin 2014; Seppola *et al.* 2016).

The activity of lysozymes in fish eggs and embryos has been used as a proxy to assess embryo and larval fitness during early developmental stages. For example, in Coho salmon (*Oncorhynchus kisutch*) and other salmonid eggs and embryos lysozyme activity was high and a single enzyme was isolated (Yousif *et al.* 1991; Yousif *et al.* 1994). In the embryos and larvae of sea bass (*Dicentrarchus labrax*) derived from brood stock fed programmed spawning diets enriched in vitamin C there was improved enzyme activity (Cecchini *et al.* 2000) and in zebrafish (*Danio rerio*), egg lysozymes and complement are important bacteriolytic factors (Wang & Zhang 2010). However, lysozymes activity seems to be species-dependent and this is presumably associated with the diversity of gene family members, which is presently unknown for most fish species.

The evolution of lysozymes has mostly been studied in mammals in which lysozyme C-type gene diversity is proposed to have been fuelled by a gene expansion event in the ancestral

amniote genome and lysozyme G-type evolution occurred by independent gene duplications (Irwin *et al.* 2011; Irwin 2014). However, in other vertebrates and invertebrates where lysozyme homologues exist, they are much less studied, despite their importance for immunity due to the reduced or absent adaptive immune response. Herein, we provide a general overview of the evolution of the lysozyme gene family by performing a comparative analysis between vertebrates and invertebrates using a combination of phylogenetic and genomic neighbourhood analyses. We explore in depth, data from fish for which limited information is currently available and include information from species that radiated earlier during vertebrate evolution and possess less modified genomes such as the freshwater ray-finned fish the spotted gar (*Lepisosteus oculatus*), the lobe-finned fish the coelacanth (*Latimeria chalumnae*) and cartilaginous fish. To understand how the vertebrate lysozyme gene families evolved available data from the protostome molluscs that have a more similar gene content to vertebrate genomes than the model protostomes, nematode and arthropods, was explored (Simakov *et al.* 2013). In protostomes lysozyme I-type exists but in the molluscs homologues of the vertebrate C-type and G-type have also been characterized (Wang *et al.* 2013a; Itoh & Takahashi 2009; Ding *et al.* 2011) while in the nematodes only lysozyme I-type exist and in arthropods the putative lysozyme G-type is absent (Daffre *et al.* 1994; Callewaert & Michiels 2010; Kaizu *et al.* 2011; Ye *et al.* 2009; Supungul *et al.* 2010; Van Herreweghe & Michiels 2012; Zhao *et al.* 2007; Itoh & Takahashi 2009; Wang *et al.* 2013a; Ding *et al.* 2011). To characterize lysozyme function in early development of fish we assessed enzyme activity during embryo development in three aquaculture species, white sea bream (*Diplodus sargus*), meagre (*Argyrosomus regius*) and gilthead sea bream (*Sparus aurata*). Transcript expression was characterized in embryos and early larval stages from different brood stock of gilthead sea bream to assess the intrinsic variation in early innate immune protection conferred by lysozymes.

2.3. Materials and methods

2.3.1. Database searches

2.3.1.1. Deuterostome databases

The deduced protein sequences of the nine human lysozyme C-types (*LYZ*, ENSG00000090382; *LYZL1*, ENSG00000120563; *LYZL2*, ENSG00000151033; *LYZL4*, ENSG00000157093; *LYZL6*, ENSG00000275722; *SPACA3*, ENSG00000141316; *SPACA5*, ENSG00000171489; *SPACA5B*, ENSG00000171478, *LALBA*, ENSG00000167531) and the unique avian calcium-binding *CaLYZ* (duck, ENSAPLG00000018472) and the human

lysozyme G-types (*LYG1*, ENSG00000144214 and *LYG2*, ENSG00000185674) were used to search for homologues in fish genomes using sequence similarity searches and gene annotations. Twelve ray-finned fish genomes were initially explored including the teleost Tetraodon (*Tetraodon nigroviridis*), Takifugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*), Amazon molly (*Poecilia_formosa*), Tongue sole (*Cynoglossus semilaevis*), Nile tilapia (*Oreochromis niloticus*), Japanese medaka HdrR (*Oryzias latipes*), platyfish (*Xiphophorus maculatus*), Atlantic cod (*Gadus morhua*), zebrafish (*Danio rerio*) and cavefish (*Astyanax mexicanus*) and the freshwater ray-finned fish the spotted gar (*Lepisosteus oculatus*), which diverged prior to the teleost whole genome duplication (Braasch *et al.* 2016), all available from ENSEMBL (www.ensembl.org, 2020). Searches were also carried out with the deduced lysozyme protein sequences from the stickleback (ENSGACG00000018290 and ENSGACG00000019327) in the European seabass (*Dicentrarchus labrax*) (http://seabass.mpipz.de/ 2018) (Tine *et al.* 2014) genome and in five other assembled teleost genomes in NCBI (https://www.ncbi.nlm.nih.gov/), the Antarctic black rockcod (*Notothenia coriiceps*) (Shin *et al.* 2014), the Atlantic salmon (*Salmon salar*) (Lien *et al.* 2016), the Japanese flounder (*Paralichthys olivaceus*) (Shao *et al.* 2017) and the Japanese (*Anguilla japonica*) and European (*Anguilla anguilla*) eel that are suggested to have diverged immediately after the teleost genome duplication and prior to the diversification of the teleost lineages.

Searches were also made for putative fish *lyzs* in the genomes of a fish basal to the tetrapod radiation, the lobe-finned fish, the coelacanth (*Latimeria chalumnae,* http://ensembl.org/, 2018) (Amemiya *et al.* 2013) and in cartilaginous fishes, the Elephant shark (*Callorhinchus milii*) (Venkatesh *et al.* 2014) and the whale shark (*Rhincodon typus*) (Read *et al.* 2017), available from the NCBI, 2018 and the little-skate (*Leucoraja erinacea*, http://skatebase.org), which diverged prior to the radiation of the bony vertebrates. Data was also retrieved from the Sea lamprey genome (*Petromyzon marinus*, http://ensembl.org/, 2020), a fish that is a representative of the most ancient lineage of extant vertebrates and diverged immediately after the origin of the vertebrates (Smith *et al.* 2013).

To better understand lysozyme gene evolution in fish, gene homologues from the house mouse (*Mus musculus*), the Opossum (*Monodelphis domestica*), the chicken (*Gallus gallus*), the duck (*Anas platyrhynchos*), two reptiles, the Anole lizard (*Anolis carolinensis*) and Chinese softshell turtle (*Pelodiscus sinensis*) and from the amphibian Xenopus (*Xenopus tropicalis*) were retrieved and included in the analysis (http://ensembl.org/, 2020). A basal invertebrate deuterostome genome of the cephalochordate Amphioxus (*Branchiostoma floridae*)

(https://genome.jgi.doe.gov/) was also included in the analysis.

To increase the number of representative teleost species included in the analysis the general NCBI teleost specific database (taxid: 32443) for proteins, gilthead sea bream (*Sparus aurata*) expression data (http://146.193.226.37/sequenceserver/) and transcriptomes for Antarctic black rockcod (*Notothenia coriiceps*) (unpublished data) and marbled rockcod (*Notothenia rossii*) (mixed tissues, http://146.193.226.37/sequenceserver/) were queried. The predicted lysozyme transcripts that were retrieved (e-value similarity score, $e < {}^{-20}$) were translated into their corresponding amino acid sequences and their identity was confirmed by searching against the NCBI human protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi, 2018) using the Blastp algorithm. The NCBI amphibian specific database (taxid:8292) was also searched to increase the number of amphibian species represented in the analysis because their sequences are suggested to have evolved differently from other tetrapods (Irwin 2014).

2.3.1.2. Mollusca databases

To better understand gene evolution of the vertebrate C-type and G-type lysozyme they were also procured in molluscs and in evolutionary related invertebrate species. Searches were performed in NCBI (https://www.ncbi.nlm.nih.gov, 2018) and Ensembl genomes (http://ensemblgenomes.org, 2018) and also in two mussel mantle transcriptomes for the Mediterranean mussel (*Mytilus galloprovinciallis*) (Bjarnmark *et al.* 2016) and for the hard-shelled mussel (*Mytilus coruscus*) using the deduced protein sequences of the bay scallop (*Argopecten irradians*) LYG (AAX09979.1) and the abalone (*Haliotis discus discus*) LYZ (AGQ50330.1) as bait. To complete the characterization of the lysozyme repertoire in Mollusca, the NCBI database (Mollusca taxid: 6474) was searched for the putative lysozyme I-type with the blue mussel (*Mytilus edulis*) homologue (AAN16207.1). Searches in NCBI were performed against the non-redundant protein sequences and Mollusca (taxid: 6447) sub-dataset using the blastp algorithm. Sequences were retrieved with a cut-off e-value $< e^{-20}$ and their identity were further confirmed by searching against the human genome. Searches in the Ensembl genomes for Mollusca lysozyme genes were directed at Pacific oyster (*Crassostrea gigas*), gastropod Owl limpet (*Lottia gigantea*) and cephalopod Octopus (*Octopus bimaculoides*) and to increase the data included in the analysis for lophotrochozoans, the lamp shell brachiopod (*Lingula anatina*) and annelid (*Capitela teleta*) genomes were also searched. The deduced proteins of the genes identified were compared with human sequences to confirm their identity. Other lysozyme sequences were also retrieved by consulting available literature (Irwin 2014; Irwin *et*

al. 2011).

2.3.2. Multiple sequence alignments and annotations

The deduced amino acid sequence of the retrieved genes were aligned using the MUSCLE algorithm available from AliView software (Edgar 2004). The percent amino acid sequence similarity/identity in selected species was calculated using the GeneDoc program (http://www.nrbsc.org/gfx/genedoc/) and specific amino acid motifs that have previously been shown to be involved in enzyme function and structure were annotated. The existence of a signal peptide was predicted using the software SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/).

2.3.3. Phylogenetic analysis

Phylogenetic trees for lysozyme members were built using both Maximum-Likelihood (ML) and Bayesian Inference (BI). Maximum-Likelihood trees were performed with the ATGC platform and run in PhyML 3.0 (Guindon *et al.* 2010) using the SMS automatic model selection, for protein evolutionary analysis according to AIC (Akaike Information Criterion) (Lefort *et al.* 2017). The alignments of the deduced amino acid sequences of vertebrate and invertebrate sequences were manually edited to remove gaps and poorly aligned regions (Supplementary Data). The edited alignments were used to build the phylogenetic trees. ML trees were constructed using a WAG matrix (best model according to SMS) and 100 bootstrap replicates. BI analysis was constructed in the CIPRES Science Gateway V. 3.3 (Miller *et al.* 2010) using the MrBayes method (Ronquist *et al.* 2012) run on XSEDE with the WAG matrix and 1,000,000 generation sampling and probability values to support tree branching. Both ML and BI trees were displayed in FigTree software (http://tree.bio.ed.ac.uk/software/figtree/) and were rooted using the invertebrate (mollusc and cephalochordate) clade. Trees were edited with Inkscape (https://inkscape.org/).

2.3.4. Gene organization and gene environment

To identify structural changes in the lysozyme genes that were associated with family evolution, the exon/intron organization of stickleback and seabass *lyz* and *lyg* genes were characterized and compared with the human. The predicted human and stickleback gene structures were retrieved from Ensembl (www.ensembl.org) and seabass from http://seabass.mpipz.de.

The gene environment of tetrapod (human and chicken) and teleost homologue genome

regions for the two lysozyme types were compared. The neighbouring genes of human lysozyme C-type genes (*LYZ* on chromosome 12, *LYZL1* and *LYZL2* on chromosome 10, *LYZL4* on chromosome 3, *LYZL6* on chromosome 17, *SPACA3* on chromosome 17, *SPACA5* and *SPACA5B* on chromosome X) and *LYG1* and *2* (chromosome 2) were characterized and used to search for homologues in chicken, spotted gar and in three teleosts (medaka, stickleback and zebrafish). The genes flanking human *LALBA* (chromosome 12) were procured in seabass and tilapia (the two teleost species with annotated assembled genomes and where a putative gene homologue was found) and also in zebrafish where the *lalba* gene seems to be absent. The neighbouring gene environment were obtained from Genomicus genome region annotation (http://www.genomicus.biologie.ens.fr) and from the BioMart tool (http://www.ensembl.org/biomart/) and their existence was confirmed by sequence searches in the genome databases of each species available from Ensembl. The seabass genome environment was obtained from genome annotations available from http://seabass.mpipz.de. Due to the overall lack of conserved genome synteny between human and ray-finned fishes, the genes in the neighbourhood of stickleback and zebrafish *lyz* and *lyg*, the spotted gar *lyz* and seabass *lalba* were also characterised and homologues were procured in human and in other teleosts.

2.3.5. Ethics Statement

Samples of gilthead sea bream eggs, embryos and larvae were collected as part of the routine sampling at hatcheries to verify production performance and was carried out in compliance with 2009/58/EC (protection of animals kept for farming). The collaborating commercial hatcheries had GLOBAL G.A.P. (Good Agricultural Practice) Certification, which includes demonstration of compliance with good animal welfare practices.

2.3.6. RNA extraction and cDNA synthesis

Expression analysis of teleost lysozyme genes were characterized during gilthead sea bream early development. Samples were obtained from brood stock of several different commercial hatcheries and were collected at well recognized and routinely sampled developmental stages. Samples for RNA extraction included embryos (n = 6 with 10 embryos per sample) and a series of consecutive days post-hatching (dph) stages: 4-8 dph (n = 3 with 20 individuals per sample) when 100% of the yolk sac was resorbed; 19-23 dph (n = 6, 3 individuals per sample) when flexion occurred; 36-43 dph (n = 6, 1 individual per sample) the

end of larval rearing and at 50-52 dph (n = 6, 1 individual per sample), mid-metamorphosis. All samples were collected in RNAlater (Sigma) and stored at -20ºC until RNA extraction.

Total RNA (tRNA) was extracted using an E.Z.N.A kit (VWR, USA) according to the manufacturer's instructions. Tissues were defrosted in lysis buffer and homogenized by mechanical disruption with two iron beads (5 mm) using a Tissue lyser II Qiagen and 4 cycles of 30 seconds at room temperature. DNase I digestion was performed directly on the columns for RNA isolation following the manufacturer's instructions. For cDNA synthesis, 500 ng of DNAse treated tRNA was used and reactions were performed with RevertAid-RT (Thermo Fisher, USA) for a 20 µL final volume with 100 pmol random hexamers (Jena Bioscience, Germany), 1 mM dNTPs (Nzytech, Portugal), 200 U of enzyme and 20 U RNase Inhibitor (Nzytech, Portugal). Reaction conditions were 25°C, 10 min; 42 °C, 60 min; 70 °C, 10 min. The quality of cDNA was initially assessed by amplification of ribosomal subunit *18s* (Table 2.1) using the following cycle: 95°C, 3 min; 25 cycles x (95°C, 20 sec; 62 °C, 20 sec; 72 °C, 20 sec); 72 °C, 5 min.

Table 2.1. Primers used in the real-time PCR amplification reactions.

Primer name	Sequence (5'-3')	T (ºC)	Efficiency (%)	R^2
Lysozyme C-type				
lalba-F	GCAGGAGGAAGTGAAGCAGA	62	100.2	0.99
lalba-R	GCTATGCTGAACAAGCCCGC			
Lysozyme G-type				
lyg1-F	GGAGGTGGACACACTAAACGG	64	97.9	0.99
lyg1-R	AACATCAACACCTGCAACGGTC			
lyg2-F	GGAGGTGGACACACTAAACGG	62	100.5	0.99
lyg2-R	GCCGTCGACATTTTCATATGAA			
lyg3-F	CTGGCTGGAGCAAGGAGCAACA	*	-	-
lyg3-R	TAGTCTCCACCATAGGTTTTGG			
lyg4-F	CTGGCTGGAGCAAGGAGCAACA	*	-	-
lyg4-R	CCCTCCTGGAAAGACAAATACA			
Reference genes				
18s-F	TGACGGAAGGGCACCACCAG	60	99.4	0.99
18s-R	AATCGCTCCACCAACTAAGAACGG			
ef1a-F	TCAAGGGATGGAAGGTTGAG	62	92.8	0.99
ef1a-R	AGTTCCAATACCGCCGAT			

* no amplification was obtained

2.3.7. Transcript expression

2.3.7.1. In silico expression

Expression of the teleost lysozyme C-type and G-type transcripts was characterized by exploring the tissue origin of expressed sequence tags (ESTs) deposited in the NCBI database. ESTs were retrieved from the zebrafish, stickleback and medaka and the tissue source was

annotated. ESTs for the human, chicken and *Xenopus* were also identified for comparison. Teleost skin transcriptomes available for Senegalese sole (*Solea senegalensis*, PRJEB29449 (Pinto *et al.* 2019)), seabass (*Dicentrarchus labrax*, GFJW00000000 (Pinto *et al.* 2017) and from the black rockcod (*Notothenia coriiceps*) (unpublished) were analysed to assess teleost lysozyme distribution in an organ that is in direct contact with the environment.

2.3.7.2. Quantitative expression

Real-time quantitative PCR (RT-qPCR) was used to determine changes in lysozyme expression in gilthead sea bream embryos and subsequent developmental stages. Specific primers were designed for the gilthead sea bream *lalba* and four *lyg* transcripts identified in this study, but amplification products were only obtained for *lyg1*, *lyg2* and *lalba* (Table 1). Reactions were performed in duplicate (< 5% variation between replicates) using a BioRad CFX Connect Real Time System and SsoFast EvaGreen supermix (Bio-Rad, Portugal). The final reaction volume was 10 µl with 200 nM of both primers and 2 µl of template cDNA (diluted 1:5) in low volume 96-well microplates (Axygen, The Netherlands). Optimized cycling conditions were 95 ºC, 30 sec followed by 45 cycles of 95 ºC, 5 sec; 58 ºC, 10 sec. Melting curves were performed to detect non-specific products and primer dimers. Control reactions were included to confirm the absence of contaminating genomic DNA. For this study the gilthead sea bream elongation factor 1-alpha (*ef1α*) and the 18 ribosomal subunit (*18s*) were selected as reference genes as they showed stable expression levels in the larval stages and embryo samples analysed. Expression data was normalized against the geometric mean of the two reference genes. PCR efficiencies and R^2 (coefficient of determination) were established with standard curves prepared in duplicate from a 10-fold serial dilution series of the purified PCR product of each of the target genes.

To associate lysozyme gene expression with gilthead sea bream embryo/larval quality, samples from two production batches from two broodstocks that had a divergent growth performance by mid-metamorphosis were compared (see above). Divergent growth was established by measuring the length of larvae reared under similar conditions at mid-metamorphosis.

2.3.8. Enzyme activity

Lysozyme activity was measured in protein extracts of gilthead sea bream (*Sparus aurata*)

embryos and compared to extracts of embryos from two other teleosts, white sea bream (*Diplodus sargus*) and meagre (*Argyrosomus regius*). Total protein was extracted from 10 mg of three embryo development stages: before hatching, hatching and after hatching. Samples from white sea bream and meagre were collected from two egg batches from two brood stocks and five aliquots of each batch were analysed per developmental stage. For gilthead sea bream, five aliquots of a single batch of eggs from a single brood stock was assayed. For white sea bream samples, before hatching, corresponded to a mixture of embryos in the blastula and gastrula stages, for meagre and gilthead sea bream embryos were in the neurula stage. Protein extracts were performed in 4 volumes of 1x PBS pH = 6.2 (0.25 mg of tissue /ul PBS) and centrifuged at 10,000 rpm at 4 ºC for 10 min. The protein content of the embryo extracts was quantified with the Bradford method using a BSA standard set (Quick StartTM Bradford Protein Assay, BioRad, USA). Enzymatic assays were performed using the method of Parry et al (Parry *et al.* 1965) with the modifications of Ellis (Ellis 1990). The assay is based on the ability of the enzyme to lyse Gram-positive bacteria *Micrococcus luteus* (*M. luteus*). Assays were performed in duplicate with 20µl of the total protein extract diluted 1:2 in 0.05M Phosphate buffer (PB, pH = 6.2) and 130 µL of an *M. luteus* (0.6 mg/mL) suspension prepared in the same buffer. Enzyme activity was monitored immediately after the addition of *M. luteus* using a Microplate Biotek Synergy 4 reader (450 nm, kinetics read every 2 min for a total of 15 min). The decrease in *M. luteus* turbidity was determined using a standard curve that was constructed with increasing concentrations of chicken egg white lysozyme (C-type) standard (Sigma L6867, from 100 to 450 µM) and read at the same time as the samples. Standard curves were plotted and the slope with the highest coefficient of determination ($R^2 \geq 0.98$) were selected to calculate Lyz activity (U). Enzyme activity (U. mg^{-1}) was normalized by the sample dry weight and one unit was defined as a reduction in absorbance of 0.001 per min per mg of embryo extract.

2.3.9. Statistical analysis

For q-PCR analysis significant differences between lysozyme transcript expression levels in the embryo samples were detected using an unpaired t-test (two-tailed). For transcript expression and enzyme activity in the larval stages a two-way ANOVA followed by a Holm-Sidak multiple comparisons post-hoc test was used. Samples with no detected enzyme activity were removed from the analysis. All data was presented as the mean ± SEM and statistical analysis was performed using SPSS 25.0 and $p < 0.05$ was considered significantly different.

2.4. Results

2.4.1. Lysozymes in fish and other animals

Homologues of human lysozyme C-type and G-type genes and transcripts were identified in fish and other vertebrates. Putative homologues of vertebrate C-type and G-type and of the invertebrate I-type were also identified in the cephalochordate and in molluscs (Figure 2.1 and Supplementary Table 2.1-2.2 annex I). Based upon phylogeny, sequence comparisons and gene neighbourhood analysis (see below) a revised nomenclature to designate the fish Lysozyme C-types and G-types is proposed (Table 2.2). The nomenclature adopted is based on the established human official gene nomenclature (HGNC) and zebrafish nomenclature convention for fish (https://wiki.zfin.org/). Fish and other vertebrate species-specific duplicates with homologues found in human were numbered and for the genes that were likely to result from the teleost genome duplication event a and b were added after the gene symbol. For other teleost genes only the gene symbol was used. For fish lysozyme C-types, when identity with the tetrapod homologues could not be assigned with certainty (due to species-specific sequence divergence) they were named Lysozyme-like (Lyzl) and when multiple gene copies were retrieved in species they were assigned letters (which do not coincide with the established Lyzl groups in tetrapods). For the cephalochordate and mollusc gene homologues the same nomenclature adopted for vertebrates (non-human) was followed and they were assigned to Lyzl or Lygl. As far as we are aware no nomenclature has been established for the invertebrate lysozyme type (I-type) and so in concordance with the naming of other vertebrate types we named them LYI. The abbreviations used in this study and the corresponding accession numbers are provided in Supplementary Table 2.1 annex I.

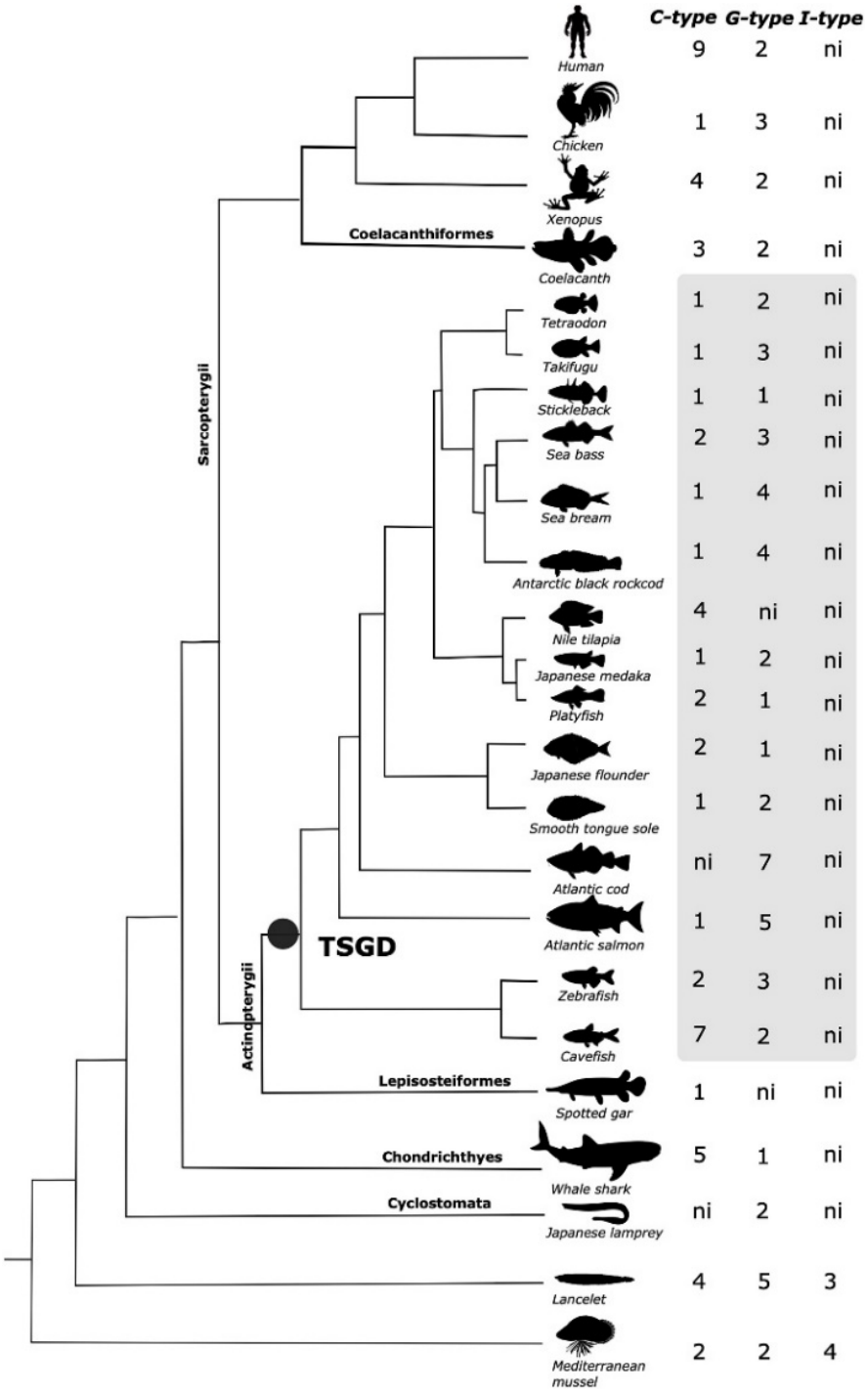


Figure 2.1. Number of lysozymes in fish. The dendogram describes the number of genes and transcripts for putative lysozyme C-type and G-type retrieved from fish (Cyclostomes, Chondrichthyes, Actinopterygii and Sarcopterygii). To better understand gene family evolution the human, chicken and frog (*Xenopus tropicalis*) lysozyme genes as well as the putative homologues from the cephalochordate (lancelet, *Branchiostoma floridae*) and mollusc (Mediterranean mussel, *Mytilus galloprovincialis*) are also indicated. For comparative purposes the number of lysozyme I-type that are specifically found in invertebrates is also shown for the cephalochordate and mollusc. The figure was drawn considering the relative evolutionary relationship between the species. The fish sequences are boxed in grey and the teleost specific genome duplication event is indicated (TSGD). Accession numbers are available in Supplementary Table 2.1 annex I. * obtained from transcriptome data. ni- not identified.

Table 2.2. Adopted nomenclature for the metazoan lysozymes.

Lysozyme C-type		**Lysozyme G-type**	
Gene/ transcript	Protein	Gene/transcript	Protein
Primate (Human)			
LYZ, LYZL1, LYZL2, LYZL4, LYZL6, *SPACA3*, *SPACA5*, *SPACA5B, LALBA*	LYZ, LYZL1, LYZL2, LYZL4, LYZL6 SPACA3, SPACA5, SPACA5B, LALBA	*LYG*	LYG
Ray-finned fish			
Lyz, Lyzl1/2, lalba	Lyz, Lyzl1/2, Lalba	*Lyg (a or b)*	Lyg (a or b)
Cartilagenous fish/Agnatha			
Lyzl	Lyzl	*lyg*	Lyg
Cephalochordate and Mollusca			
Lyzl	Lyzl	*Lyg*	Lyg

2.4.1.1. C-type Lysozymes

In human, nine lysozyme C-type members exist but in fish gene number is lower and is distinct across the species analysed (Figure 2.1). In teleosts, a single lysozyme C-type gene copy was retrieved from the genomes of Tetraodon (*Tetraodon nigroviridis*), Takifugu (*Takifugu rubripes*) and stickleback (*Gasterosteus aculeatus*) but multiple genes were found in other species. For example, two genes were retrieved from the seabass (*Dicentrarchus labrax*) genome and four and five genes were retrieved from the tilapia and cavefish genomes, respectively (Supplementary Table 2.1 annex I). However, no lysozyme C-type gene was identified in the Atlantic cod (*Gadus morhua*) genome (Figure 2.1). In other fishes, such as the ray-finned spotted gar a single lysozyme C-type transcript was retrieved but, in the coelacanth (*Latimeria chalumnae*) and in the cartilaginous fish, the elephant shark (*Callorhinchus milii*), three genes were found, respectively. In the genome of another cartilaginous fish, the little skate (*Leucoraja erinacea*), a putative lysozyme C-type gene was identified but in the whale shark (*Rhincodon typus*) five genes were found (Supplementary Table 2.1 annex I). Searches in the lamprey (*Petromyzon marinus*) genome failed to retrieve a lysozyme C-type gene and mining of the invertebrate deuterostome genome of the cephalochordate (*Branchiostoma floridae*) identified five putative homologue members of the lysozyme C-type family (Figure 2.1). In the transcriptome data for gilthead sea bream (*Sparus aurata*) and for two nototheniids, the Antarctic fish black rockcod (*Notothenia coriiceps*) and marbled rockcod (*Notothenia rossii*), a single gene transcript for lysozyme C-type was identified.

In molluscs, gene or transcript homologues that are putative members of the lysozyme C-type family in human were only identified in some of the gastropods and bivalve species analysed (Supplementary Table 2.2 annex I). In bivalves three transcripts were retrieved from the Mediterranean mussel (*Mytilus galloprovinciallis*) and gene transcripts for a single gene were found in the hard-shelled mussel (*Mytilus coruscus*) and in the Venus clam (*Cyclina sinensis*) (Supplementary Table 2.1 annex I) (Gerdol & Venier 2015). In gastropods the two lysozyme C-type sequences previously identified in the variously colored abalone (*Haliotis diversicolor*) (Bachali *et al.* 2002) and the single lysozyme C-type gene in the owl limpet (*Lottia gigantea*) were retrieved. No putative genes were retrieved from the annelid or brachiopod.

2.4.1.2. G-type Lysozymes

In human, two lysozyme G-type genes (LYG1 and LYG2) have been described and three genes exist in chicken. In fish, two lysozyme G-type genes were isolated from the coelacanth and in the teleosts gene number was variable (Figure 2.1). While some species have multiple gene copies, in others, the tilapia, spotted gar and the cartilaginous elephant shark (*Callorhinchus milii*) no genes were found. However, in another cartilaginous fish, the whale shark (*Rhincodon typus*), a single gene was retrieved and in the little skate two genes were isolated. In lamprey, two lysozyme G-type genes were also found (Figure 2.1, Supplementary Table 2.1 annex I). Searches in the genome and transcriptome data of the sea bream, identified three lysozyme G-type transcripts and a fourth transcript was obtained from the NCBI database. For the black rockcod (*Notothenia coriiceps*) four genes were identified in the genome and three transcripts were retrieved from the species transcriptome. In the marbled rockcod (*Notothenia rossii*) two transcripts were identified. In the cod genome a large expansion of the lysozyme G-type members occurred and in addition to the four genes previously described (Seppola *et al.* 2016), in this study three additional genes were identified. Five putative lysozyme G-type genes were retrieved from the amphioxus genome (Supplementary Table 2.1 annex I).

Within the molluscs a larger number of species possessed lysozyme G-type genes relative to those containing lysozyme C-type (Supplementary Table 2.1 annex I). Searches identified putative genes and transcripts in five bivalves (bay scallop, *Argopecten irradians*; Pacific oyster, *Crassostrea gigas*; Japanese scallops, *Mizuhopecten yessoensis*; Zhikong scallops, *Chlamys farreri*; and the Mediterranean mussel) and in seven gastropods (the disk abalone, *Haliotis discus discus*; Betuline cone, *Conus betulinus*; variously coloured abalone; owl limpet;

freshwater snail, *Biomphalaria glabrata*; European physa, *Physella acuta*; and California sea hare, *Aplysia californica*) and the number was variable among species.

2.4.1.3. I-type Lysozymes

Lysozymes I-type was absent from vertebrates, but it was the most common lysozyme-type present in invertebrates. Lysozymes I-type was identified in species of the three major representative classes of molluscs (gastropods, bivalves and cephalopods) as well as in annelids, in a brachiopod and in invertebrate deuterostome (echinoderm and cephalochordate) genomes (Supplementary Table 2.2 annex I).

2.4.2. Phylogeny of the fish lysozymes

Phylogenetic analysis of fish Lysozymes C-type (Figure 2.2A, Supplementary Figure 2.1A annex I) and Lysozymes G-type (Figure 2.2B, Supplementary Figure 2.1B annex I), including previously characterised vertebrate proteins and the deduced protein homologues in invertebrates, confirmed their identity. Both BI and ML algorithms produced trees with similar branching topologies and suggested that the vertebrate and invertebrate sequences shared a common origin and that metazoan Lysozyme C-types and G-types resulted from lineage-specific and recent species-specific gene duplication events. Most of the phylogenetic tree's branches had poor statistical support suggesting that the metazoan lysozymes evolved under distinct selective pressures and gene paralogues and orthologues across species had variable mutation rates (Figure 2.2 and Supplementary Figures 2.1 annex I). Inclusion of sequences from the lamprey (a basal vertebrate) and from fish representatives that diverged earlier in the vertebrate radiation and thus are likely to possess genomes that are less modified or rearranged than other vertebrates (e.g. the cartilaginous fishes, the spotted gar and coelacanth) did not improve the resolution of the phylogenetic trees, which suggests that even in these fish species lysozyme family genes rapidly evolved. Moreover, when lysozyme sequences from amphibian species were included in the phylogenetic analysis, they did not cluster with other tetrapods, suggesting their evolution diverged from other vertebrates. For this reason, only lysozyme sequences from a single amphibian (*Xenopus tropicalis*) were included in the analysis. Similarly, since no fish gene homologues to bird Calyzl were identified this sequence was removed from the final phylogenetic tree.

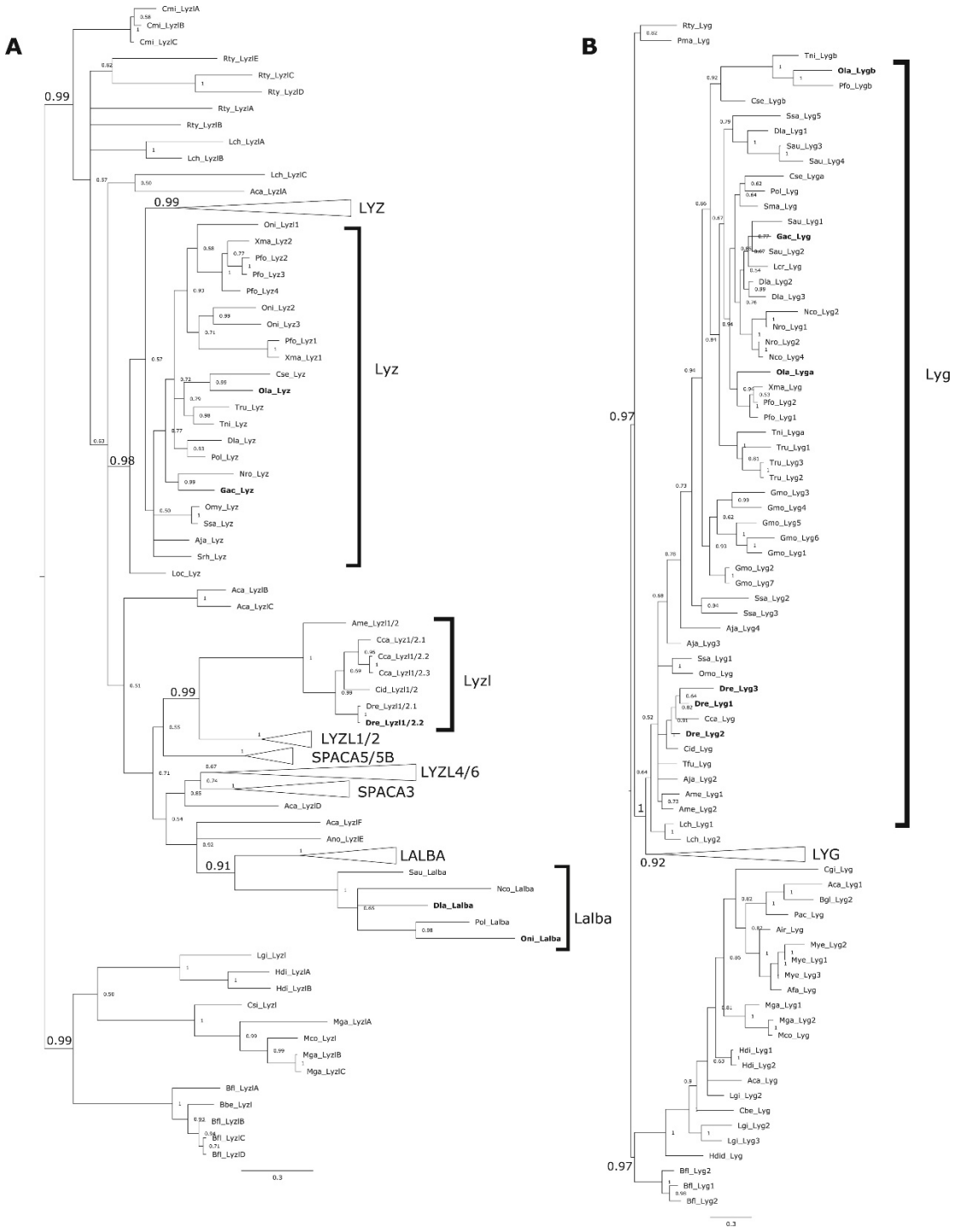


Figure 2.2. Phylogenetic tree of the fish Lyz (A) and Lyg (B) and other metazoan LYZ families. The phylogenetic trees were constructed using the Bayesian inference (BI) method and only branch support values (posterior probability values) for the main branches are shown. To facilitate interpretation of both trees the teleost clusters are annotated. Both lysozyme C-type and G-type trees were rooted using the invertebrate (cephalochordate and mollusc) sequence cluster. In the tree nodes that contain teleost sequences the posterior probability values and the ML bootstrap branch support values are indicated and the font size is larger. To facilitate interpretation the tetrapod (mammalian, bird, reptile and amphibia) branch was collapsed. The teleost lysozyme C-type (A) are in three main clusters with the Human LYZ, LYZL1/2 and LALBA. The fish Lyg (B) forms a single cluster with the tetrapod Lygs. In both trees the species that were used for gene environment characterization (Figure 2.6, 2.7 and 2.8) are highlighted in bold. Similar trees were generated with the ML method and are available as Supplementary Figure 2.1 annex I. Abbreviations and accession numbers are available as Supplementary Table 2.1 annex I. * represents transcript sequences.

In the lysozyme C-type tree, the cephalochordate and mollusc members (Figure 2.2A) form a cluster basal to the vertebrates. The teleost sequences group with the vertebrate members in three main clusters: a) one containing tetrapod LYZ and the majority of the teleost sequences, b) another containing mammalian LYZL1/LYZL2 and the zebrafish and other cypriniform sequences and c) mammalian LALBA with some teleost sequences (Figure 2.2A, Supplementary Figure 2.1A, annex I). The clustering of the phylogenetic tree suggests that before the tetrapods and teleosts diverged genes encoding the three members of the lysozyme C-type existed. Based upon sequence clustering the teleost sequences that group with the human and other tetrapod LYZ were named Lyz, the zebrafish and other cypriniform sequences were named Lyzl1/2 and the teleost sequences from the Japanese flounder, sea bass, black rockcod and tilapia that cluster with the mammalian LALBA were named Lalba. The spotted gar lysozyme C-type is incomplete and lacks part of the C-terminal end and this probably explains why it radiates prior to the Lyz cluster. The tetrapod LYZL4/6 and SPAC3 sequences form a single cluster and the SPACA5 and SPACA5B also cluster together and gene homologues in fish were not found. The multiple family members retrieved from the cartilaginous fish the elephant shark and the whale shark radiate basal to gnathostome sequences and they seem to be the result of species-specific duplications that occurred after their divergence.

For the metazoan LYG phylogenetic tree all the teleost sequences group with the tetrapod Lygs (Figure 2.2B). Clustering of the multiple cod *lyg*s suggest that they are the result of species-specific duplications as has previously been suggested (Seppola *et al.* 2016) and this is also the case for most of the teleost *lygs* and is corroborated by gene mapping that indicates that most teleost *lygs* probably arose by tandem gene duplications (Supplementary Table 2.1 annex I). The clustering of the teleost Lyg sequences in the phylogenetic tree revealed that most of the teleosts lost one of the duplicate forms that arose during the teleost specific genome duplication with the exception of tetraodon, medaka, amazon molly and smooth tongue sole.

2.4.3. Sequence comparisons of the fish lysozymes

The amino acid sequence conservation across the deduced fish Lyz, Lyzl1/2 and Lalba (Figure 2.3) and Lyg (Figure 2.4) proteins is low and comparison of the same type of Lyz from different fish species also revealed that sequence homology was variable. However, many of the amino acid changes identified in the sequences resulted in substitutions with amino acids that share similar physicochemical properties suggesting they will have relatively little effect on the function and structure of the enzyme. Furthermore, key amino acids known to determine

function and structure of LYZ were almost totally conserved. The conserved motifs in the deduced Lyz and Lyg proteins from fish were different and a series of conserved cysteine residues were present across vertebrate and invertebrate LYZ and absent from LYG.

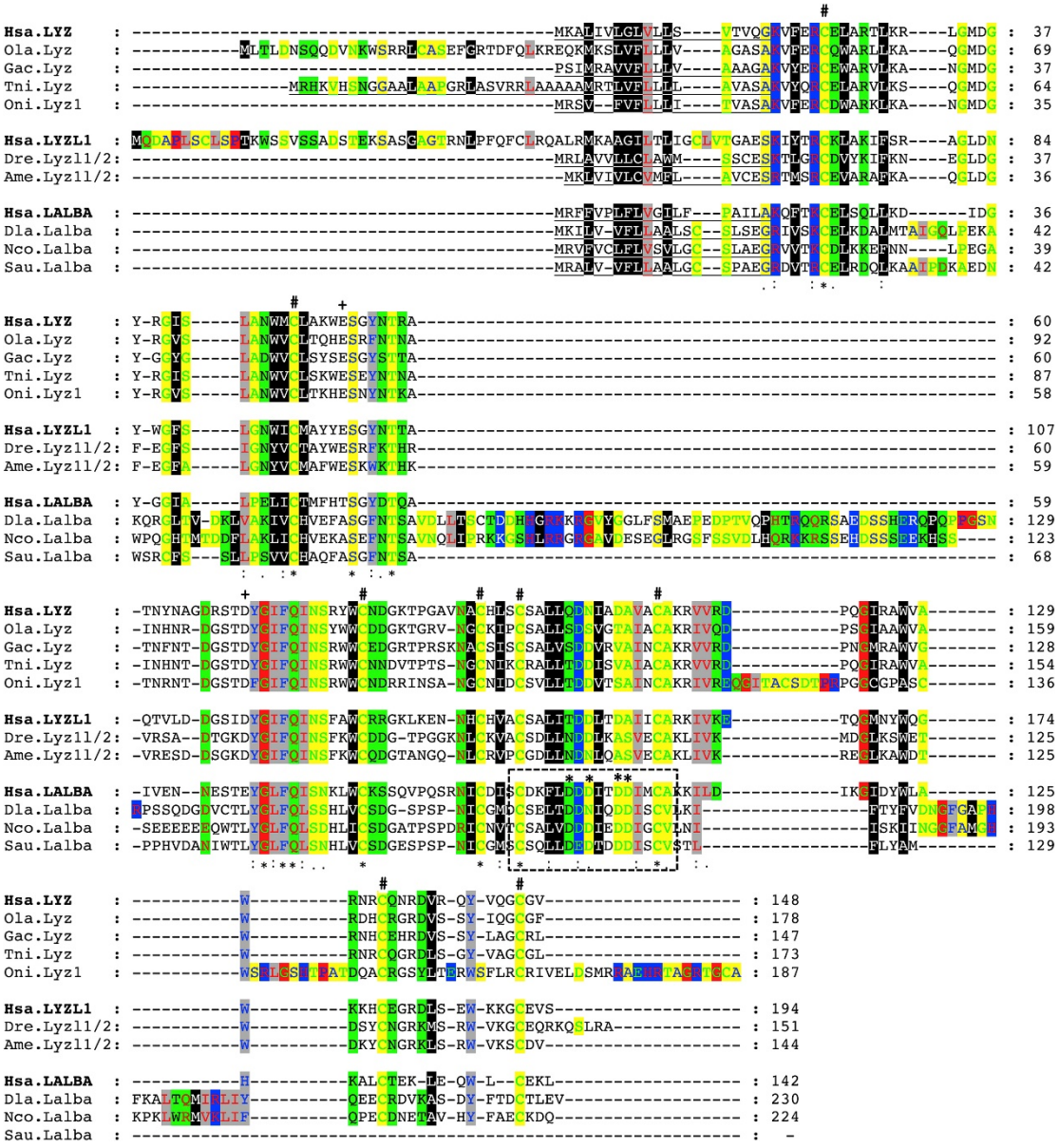


Figure 2.3. Multiple sequence alignment of the teleost Lyz, Lyzl1/2 and Lalba with the human homologues. The predicted signal peptide for each sequence is underlined, the conserved cysteine residues involved in the Lyz structural conformation are indicted with "#" and the amino acids putatively involved in the catalytic mechanisms are indicated with "+". The conserved motif characteristic of the mammalian LALBA is annotated by a dashed black outlined box (Qasba & Kumar 1997) and the calcium binding residues within this motif are marked with "*". Alignment columns were highlighted using GeneDoc according to physiochemical properties (Text color/Shade color): Proline Blue/Red; Glycine Green/Red; Tiny Blue/Yellow; Small Green/Yellow; Positive Red/Blue; Negative Green/Blue; Charged White/Blue; Amphoteric Red/Green; Polar Black/Green; Aliphatic Red/Gray; Aromatic Blue/Gray; Hydrophobic White/Black). Gaps in the sequence alignment are indicated by dashes. Complete residue conservation is annotated with a "*" and highlighted in bold. Partially conserved residues are denoted by "." and the position of the consensus amino acids present in the greatest number of sequences are indicated with ":".

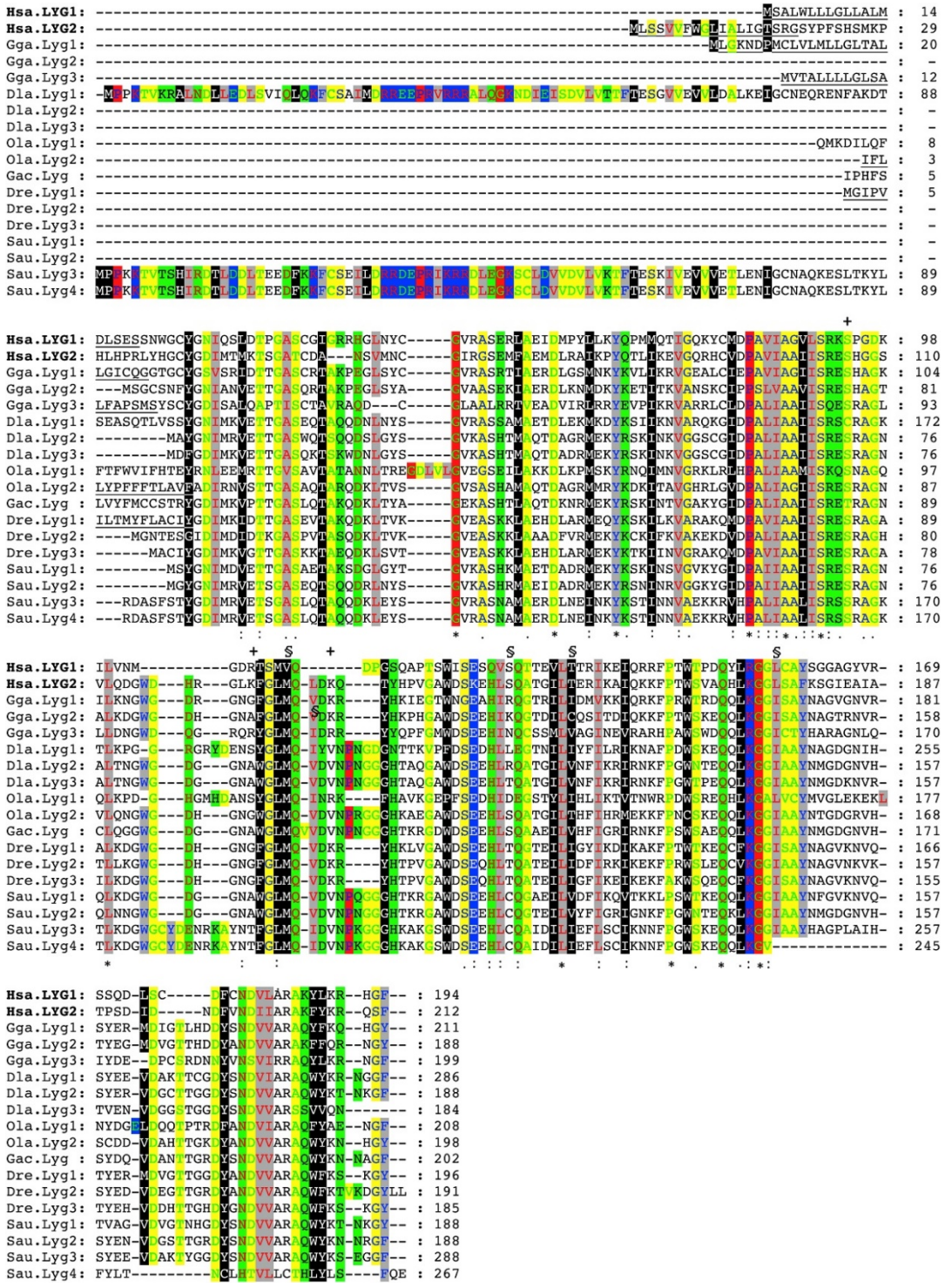


Figure 2.4. Multiple sequence alignment of the teleost Lyg with the human homologues. The three catalytic residues are annotated with a "+" and the amino acids potentially involved in substrate binding are annotated with "§". Alignment columns were highlighted using GeneDoc according to physiochemical properties (Text color/Shade color): Proline Blue/Red; Glycine Green/Red; Tiny Blue/Yellow; Small Green/Yellow; Positive Red/Blue; Negative Green/Blue; Charged White/Blue; Amphoteric Red/Green; Polar Black/Green; Aliphatic Red/Gray; Aromatic Blue/Gray; Hydrophobic White/Black). Gaps in the sequence alignment are indicated by dashes. The predicted putative signal peptides are underlined. Complete residue conservation is annotated with a "*" and highlighted in bold. Partially conserved residues are denoted by "." and the position of the consensus amino acids present in the greatest number of sequences are indicated with ":".

2.4.3.1. C-type Lysozyme

The stickleback and sea bass Lyz share 72 and 74% aa similarity with human LYZ but the latter only shares 32% amino acid similarity with the sea bass lalba, which was more similar (52% aa similarity) to human LALBA. The zebrafish lysozyme members are more identical in sequence with human LYZL1 and LYLZ2 (64% aa identity) than with the human LYZ (57%) and the stickleback (52%) Lyz. Eight conserved cysteine residues are present in the teleost and other metazoan Lyz amino acid sequences and these are suggested to be responsible for the enzymes conserved structural conformation (Figure 2.3). In other regions of the protein several amino acids have been substituted by others with similar physicochemical proprieties suggesting that despite changes in sequence the function and structure is likely to have been maintained. Moreover, the two LYZ catalytic residues Glu (E) and Asp (D) located in the middle region of the protein were also conserved in the teleost Lyz and Lyzl1/2 but not in the Lalba where Glu (E) was mutated to a Thr (T) or Ala (A) and Asp (D) to a Leu (L). Comparison of teleost Lalba with human LALBA revealed conservation of four calcium binding residues (D, Asp) that are characteristic mammalian lactalbumin motifs and several other conserved blocks containing amino acids that share similar physiochemical properties (Figure 2.3). Several other sequence motifs are conserved between the teleost and human Lyz, Lyzl1/2 and Lalba but the functional significance of this is currently unknown. Within the putative teleost Lalba the characteristic catalytic residues of Lysozyme C-type were mutated suggesting that in common with human LALBA they may lack the characteristic enzyme function. All teleost Lyz, Lyzl1/2 and Lalba sequences had a predicted signal peptide suggesting they are secreted, which agrees with their extracellular localization.

2.4.3.2. G-type Lysozyme

The medaka Lyg proteins share 54% sequence similarity while the three zebrafish proteins share 73-84% similarity, revealing that sequence variation between the teleost paralogues is distinct. Similarly, human LYG1 shares 56% sequence similarity with its duplicate LYG2 and sequence similarity between the chicken homologues varied between 56 to 70%. Sequence alignment of the fish sequences with other metazoan sequences revealed conservation of amino acid residues within the mid-region of Lyg1 that are involved in enzyme structure and function (e.g. Glu (E) and two Asp (D), Figure 2.4) and amino acid substitutions in this region were generally for residues that share similar physicochemical characteristics and are unlikely to affect enzyme activity. Similarly, the amino acid residues proposed to bind to the substrate

(three Leu (L) and a Gly (G)) were highly conserved across species (Figure 2.4). The presence of a signal peptide in some, but not all Lyg sequences suggests differences may occur in the cellular location of some forms.

2.4.4. Gene organization of the teleost lysozymes

The gene organization of the teleost stickleback and sea bass *lyz, lyzl1/2, lalba* and *lyg* genes were compared with the human genes. The single stickleback *lyz* and zebrafish *lyzl1/2* share similar gene organization to the human homologues and both genes are composed of 4 exons (Supplementary Figure 2.2 annex I). However, the sea bass *lalba* gene structure differs from the human *LALBA* and it possesses one extra exon (total 5 exons). The sea bass *lalba* exon 2 is larger than the homologue exon in the human gene (Supplementary Figure 2.2 annex I). The two human *LYG* genes (*LYG1* and *LYG2)* are composed of 5 exons and those of the stickleback *lyg* have a similar organization.

2.4.5. Neighbouring gene environment

The human *LYZ* gene maps to chromosome 12 and the chicken gene to chromosome 1 and both share a similar gene environment (Figure 2.5). Analysis of the *lyz* gene environment in the spotted gar, medaka, stickleback and zebrafish (Figure 2.5, Supplementary Figure 2.3 annex I) revealed no synteny with the LYZ genome region in human or chicken. However, the spotted gar *lyz* that maps to LG11 shares a similar gene environment with the human LYZL4 on chromosome 3 and a homologue genome region to human LYZ was found on LG8 (Figure 2.5). Moreover, comparison of the *lyz* flanking regions between the spotted gar, medaka, stickleback and zebrafish reveal poor overall conservation. The exception was the medaka and stickleback *lyz* (chromosome 12 and group XIV, respectively) where gene synteny was maintained (Figure 2.5). Homologues of the zebrafish *lyzl1/2* flanking genes are found in medaka but they map to different genome regions (Supplementary Figure 2.3A annex I). In chicken, the gene homologues that flank zebrafish *lyzl1/2* are on chromosome 1 but are not in proximity with the chicken *Lyz* gene (Supplementary Figure 2.3A annex I). The zebrafish *lyzl1/2* genome region shares a similar gene repertoire with human *SPACA5* and *SPACA5B* on chromosome X and two genes are shared with human *LYZL1/2* genes on chromosome 10 but they are very distant. This indicates that the lysozyme C-type from cypriniform fish had a different origin from the *Lyz* gene that persisted in other teleosts and corroborates the clustering of the teleost sequences in the phylogenetic tree (Figure 2.2A). The comparative analysis of the linkage maps in teleosts

reveals the complexity of the evolution of the *lyz* genes and although linkage is found between zebrafish and human it is absent for the other teleosts analysed (Supplementary Table 2.1 annex I).

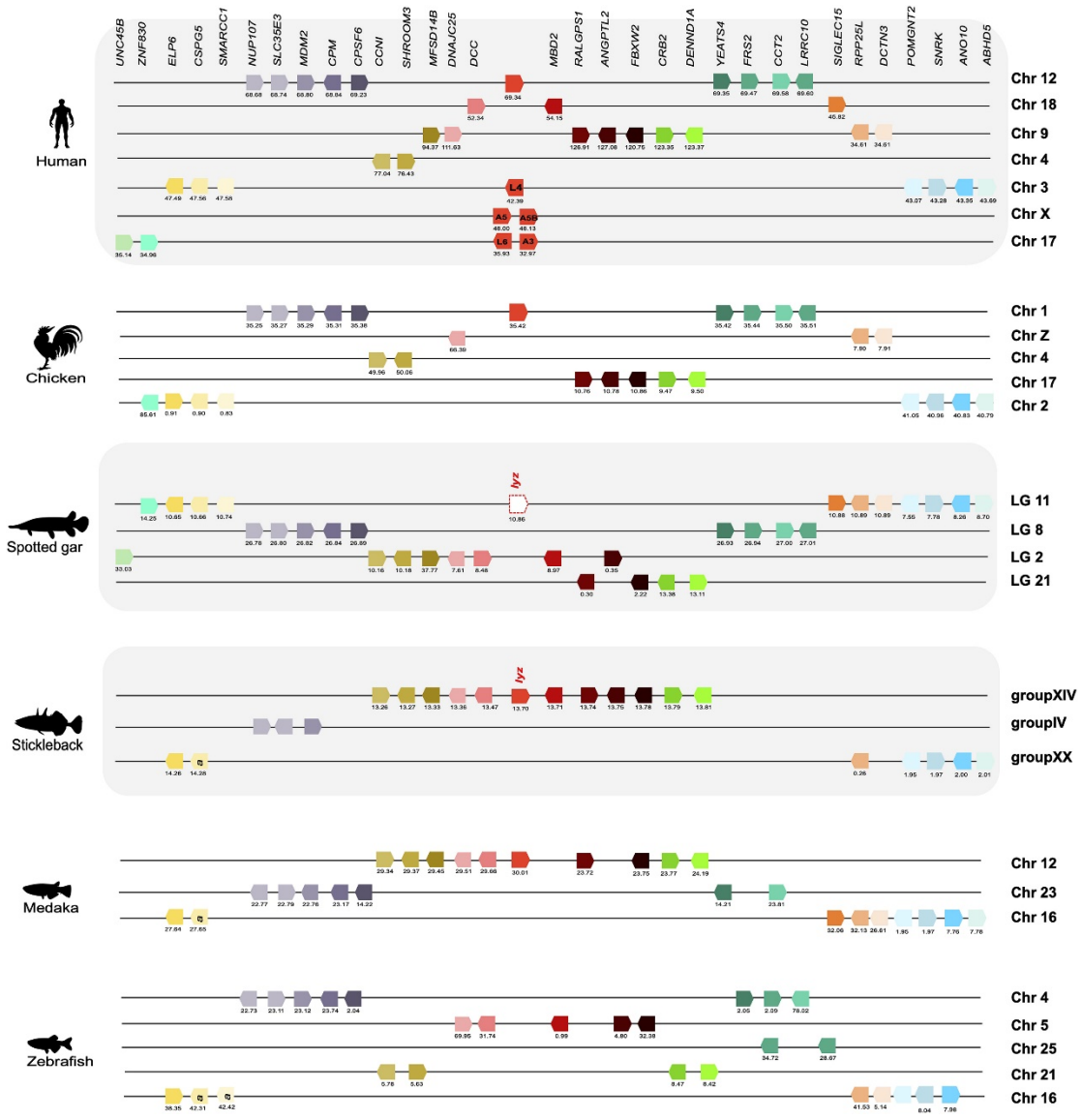


Figure 2.5. Lysozyme C-type neighbouring gene environment in ray-finned fish and tetrapods. The lysozyme C-type gene members are represented by a solid red arrow and the neighbouring gene families are represented by different coloured blocks. The direction of the arrowheads represents transcript orientation predicted in the analysed genomes. The human lysozyme C-type members represented are: Lyz on chromosome 12, lysozyme like 4 (L4) on chromosome 3, Lysozyme-like 6 (L6) and SPACA3 on chromosome 17 and SPACA5 (A5) and SPACA5B (A5B) on chromosome X. For simplicity the genome region of the human LYZL1/2 on chromosome 10 is not represented in the figure because with the exception of zebrafish no gene synteny was found with teleosts (shown in Supplementary Figure 2.3 annex I). Genes are mapped based on their actual positions predicted in their genome assemblies. Only genes that were conserved across the homologous genome regions retrieved from the species analysed are represented. The human, spotted gar and stickleback are shaded in grey as their genome was used as the reference to search for homologues in other species. Only chromosome regions that share at least two genes are represented. Genes represented are: Unc-45 Myosin Chaperone B (*UNC45B*), Zinc Finger Protein 830 (*ZNF830*),

Elongator Acetyltransferase Complex Subunit 6 (*ELP6*), Chondroitin Sulfate Proteoglycan 5 (*CSPG5*), SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily C Member 1 (*SMARCC1*), nucleoporin 107 (*NUP107*), solute carrier family 35 member E3 (*SLC35E3*), MDM2 proto-oncogene (*MDM2*), carboxypeptidase M (*CPM*), cleavage and polyadenylation specific factor 6 (*CPSF6*), cyclin I (*CCNI*), shroom family member 3 (*SHROOM3*), Major facilitator superfamily domain containing 14B (*MFSD14B*), DnaJ heat shock protein family (Hsp40) member C25 (*DNAJC25*), DCC netrin 1 receptor (*DCC*), ethyl-CpG binding domain protein 2 (*MBD2*), Ral GEF With PH Domain And SH3 Binding Motif 1 (*RALGPS1*), angiopoietin-like 2 (*ANGPTL2*), F-box and WD repeat domain containing 2 (*FBXW2*), crumbs family member 2 (*CRB2*), DENN domain containing 1A (*DENND1A*), YEATS domain containing 4 (*YEATS4*), fibroblast growth factor receptor substrate 2 (*FRS2*), chaperonin containing TCP1 subunit 2 (*CCT2*), leucine rich repeat containing 10 (*LRRC10*), Sialic Acid Binding Ig Like Lectin 15 (*SIGLEC15*), Ribonuclease P/MRP Subunit P25 Like (*RPP25L*), Dynactin Subunit 3 (*DCTN3*), Protein O-Linked Mannose N-Acetylglucosaminyltransferase 2 (Beta 1,4-) (*POMGNT2*), SNF Related Kinase (*SNRK*), Anoctamin 10 (*ANO10*), Abhydrolase Domain Containing 5 (*ABHD5*). Comparisons using the zebrafish genome as the reference are available as Supplementary Figure 2.3 annex I.

For *LALBA*, the human gene maps to chromosome 12 and no gene synteny was found with the sea bass (LG7) and Nile tilapia (LG6) *lalba* genome regions. Searches of the sea bass *lalba* neighbouring genes did not find homologues in human chromosome 12, although homologues were found in the Nile tilapia *lalba* genome region (Figure 2.6).

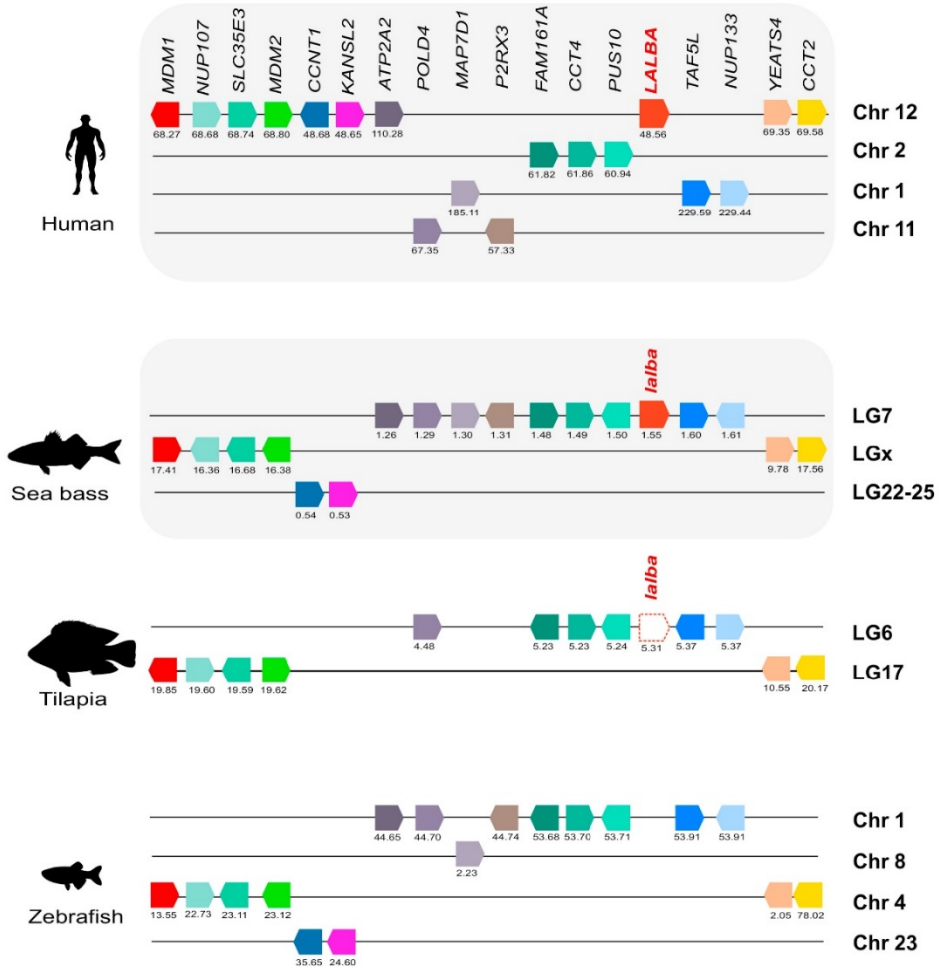


Figure 2.6. Teleost and human *LALBA* neighbouring gene environment. The human and teleost *LALBA* genes are represented by solid red arrows and the neighbouring gene families by different coloured blocks. The direction of the arrowheads represents the transcript orientation predicted in the species genomes. Genes are mapped based on their actual positions predicted in their genome assemblies. Only genes that were conserved across the homologous genome regions retrieved from the species analysed are represented. The species that were taken as the reference for the searches (human and sea bass) are shaded in grey. Only chromosome regions that share at least two genes are represented and include: Mouse double minute 1 homolog (*MDM1*), nucleoporin 107 (*NUP107*), solute carrier family 35 member E3 (*SLC35E3*), MDM2 proto-oncogene (*MDM2*), Cyclin T1 (*CCNT1*), KAT8 regulatory NSL complex subunit 2 (KANSL2), ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2 (*ATP2A2*), DNA polymerase subunit delta 4 (*POLD4*), map7 domain containing protein 1 (*MA7D1*), Purinergic receptor P2X 3 (*P2RX3*), ATPase H+ transporting V0 subunit e2 (*atp6v0e2*), FAM161 centrosomal protein A (*FAM161A*), Chaperonin containing TCP1 subunit 4 (*CCT4*), Pseudouridine synthase 10 (*PUS10*), interferon gamma receptor alpha chain (*CRFB17*), TATA-box binding protein associated factor 5 like (*TAF5L*), Nucleoporin 133 (*NUP133*), YEATS domain containing 4 (YEATS4), Fibroblast growth factor receptor substrate 2 (*FRS2*), Chaperonin containing TCP1 subunit 2 (*CCT2*).

LYG gene evolution was also complex and no gene synteny was found between linked genes in tetrapod *LYG* and teleost *lyg*. Although, the flanking genes were present in conserved blocks of synteny in other genome regions in teleosts (Figure 2.7). The human, chicken and zebrafish paralogues map in tandem suggesting that they were the result of independent species-specific recent gene duplications. No *lyg* gene was predicted in the spotted gar genome, although a similar genome region to that flanking *lyg* in the stickleback and medaka was found, suggesting that the gene was most likely deleted from the spotted gar genome. In medaka the two identified *lyg* genes map to different chromosomes: 1) the *lyg* gene on chromosome 6 is flanked by similar genes to the homologue region in stickleback and 2) the gene environment flanking *lyg* on chromosome 18 shares no resemblance to the other fish or tetrapod regions (Figure 2.7). Mapping of the two medaka *lygs* to different chromosomes suggest that they probably emerged from the teleost genome duplication event like in the tetraodon, amazon molly and smooth tongue sole paralogues, with which they cluster in the phylogenetic tree, (Figure 2.2B). This is supported by the identification of duplicates of the neighbouring genes (eg: *cnga1*, *aimp1* and *sgms2* (data not shown) in the two genome regions. No homologue genome region is found between zebrafish and *lyg* locus in other teleosts (Supplementary Figure 2.3B annex I). In summary, our evolutionary analysis confirmed that lysozymes are an ancient family. The evolutionary and gene synteny analysis suggests that lysozyme-type family members in vertebrates evolved under different selective pressures and that considerable genome rearrangements occurred in each lineage.

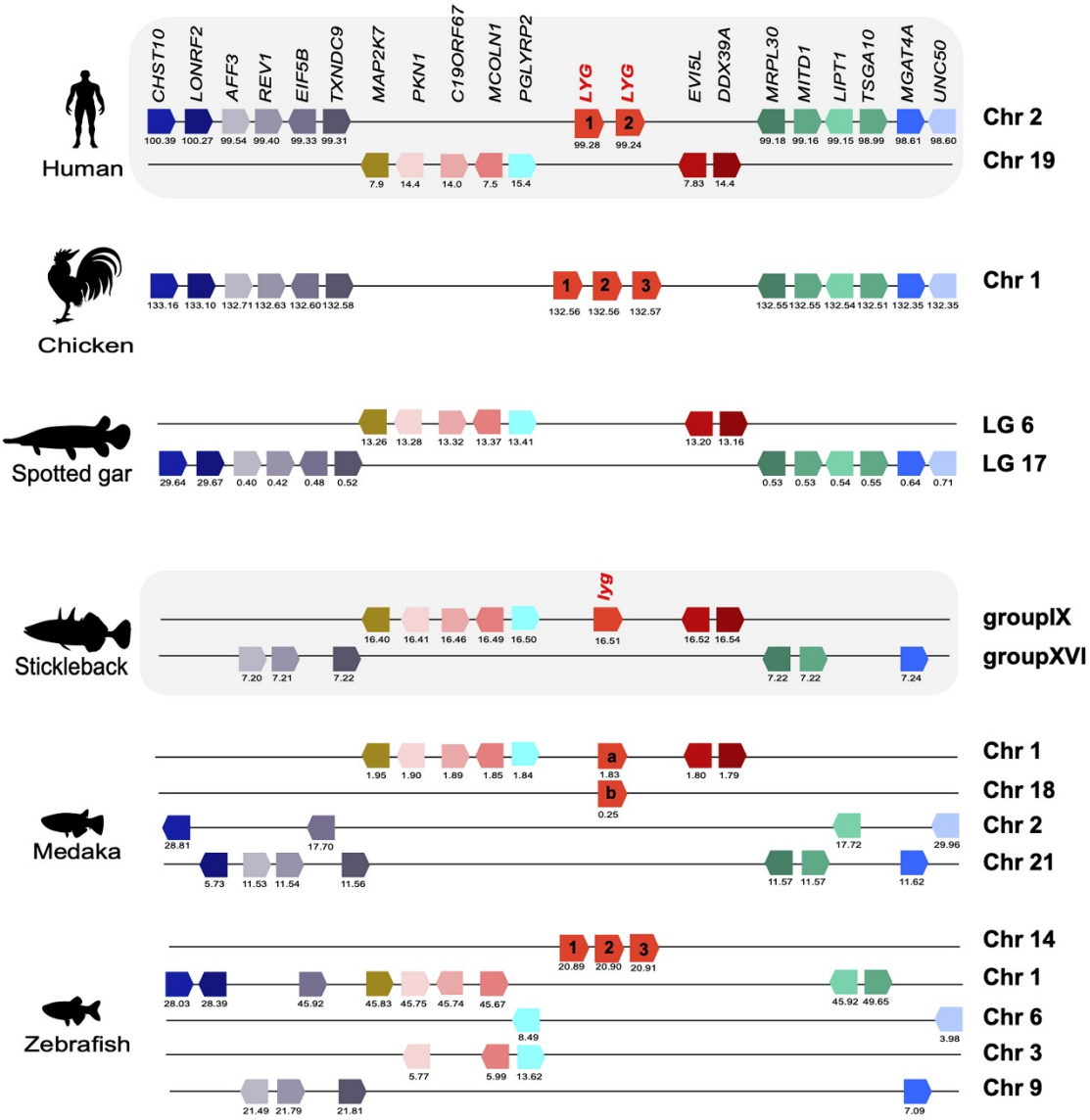


Figure 2.7. Lysozyme G-type neighbouring gene environment in ray-finned fish and tetrapods. The *LYG* genes are represented by a solid red arrow and the numbers inside them indicate the family member. The neighbouring gene families are represented by different coloured blocks or coloured outlines. The direction of the arrowheads represents the transcript orientation predicted in the genomes analysed. Genes are mapped based on their actual positions predicted in their genome assemblies. Only genes that are conserved across the different homologous genome regions are represented. Only chromosome regions that share at least two genes are represented and include: carbohydrate sulfotransferase 10 (*CHST10*), LON peptidase N-terminal domain and ring finger 2 (*LONRF2*), F4/FMR2 family member 3 (*AFF3*), ADNA directed polymerase (*REV1*), eukaryotic translation initiation factor 5B (*EIF5B*), oredoxin domain containing 9 (*TXNDC9*), thimitogen-activated protein kinase 7 (*MAP2K7*), protein kinase N1 (*PKN1*), chromosome 19 open reading frame 67 (*C19ORF67*), mucolipin 1 (*MCOLN1*), peptidoglycan recognition protein 2 (*PGLYRP2*), ecotropic viral integration site 5 like (*EVI5L*), ExD-box helicase 39A (*DDX39A*), mitochondrial ribosomal protein L30 (*MRPL30*), microtubule interacting and trafficking domain containing 1 (*MITD1*), lipoyltransferase 1 (*LIPT1*), testis specific 10 (*TSGA10*), KIAA1211 (*KIAA1211*), mannosyl (alpha-1,3-)-glycoprotein beta - 1,4 – N – acetylglucosaminyltransferase A (*MGAT4A*), unc-50 inner nuclear membrane RNA binding protein (*UNC50*) and cytochrome c oxidase assembly factor 5 (*COA5*).Comparisons using the zebrafish genome as the reference are available as Supplementary Figure 2.3 annex I.

2.4.6. Tissue distribution

Expression of lysozymes in teleost tissues was characterised to assess their potential function. This was done by a) characterizing digital data for teleost transcripts (Table 2.3), b) analysis of transcript abundance in the skin of teleosts that occupy different ecological niche (e.g. sea bass, black rockcod and sole) (Table 2.4) and by c) quantifying gene expression levels in eggs and during early development of sea bream larvae from different brood stocks.

Table 2.3. Tissue origin of *LYZ* and *LYG* transcripts in vertebrates.

Organism	transcript	EST tissue origin
Lysozyme C-type		
Human	*LYZ*	eye; colon; trachea; normal nasopharynx; thymus; cerebellum; small intestine; stomach; lung
Chicken	*Lyz*	embryo; brain; ovary; skin; bone marrow; muscle; spleen; thymus; testis; kidney; liver; gut
Xenopus	*Lyz*	stomach; testes; whole body; embryo; liver
Zebrafish	*lyzl1/2*	testis; gut; brain; whole body; embryo; kidney; eye
Japanese medaka	*lyz*	Liver
Stickleback	*lyz*	liver; gill
Lysozyme G-type		
Human	*LYG1*	connective tissue; ovary; kidney; mixed; embryonic tissue; eye; uterus
	LYG2	eye; skin; testis
Chicken	*Lyg (1, 2* and *3)*	liver; embryo; brain; ovary; skin; bone marrow; muscle; spleen; thymus; testis; gut kidney
Xenopus	*Lyg*	spleen; whole body; lung; liver
Zebrafish	*lyg1*	brain; reproductive system; olfactory rosettes; gills; skin; eye; whole body
	lyg2	muscle; gills
	lyg3	gill; embryo; brain; testis; gonad; gut; skin; fin; eye
Japanese medaka	*lyg (1* and *2)*	liver; fin; whole embryo
Stickleback	*lyg*	Gill

Table 2.4. Lysozyme transcripts retrieved from teleost skin transcriptomes and their relative abundance indicated in RPKM and CPM.

	Sea bass (RPKM)	Black rockcod (logCPM)	Senegalese sole (CPM)
C- type			
lyz	*ni*	*ni*	*lyz1* (20)
lalba	*ni*	*lalba* (5.54)	*lalba* (540)
G- type			
lyg	*lyg1* (6.04)	*lyg1* (1.59)	*lyg1* (140)
	lyg3 (81.09)	*lyg2* (1.83)	*lyg2* (183)

ni. Not identified

2.4.7. Digital gene expression

Digital gene expression revealed that teleost *lyz* and *lyg* transcripts have a widespread tissue distribution and the expression pattern of the teleost paralogues suggests that gene function may be different and may vary across species (Table 2.3). For example, the zebrafish *lyg* duplicates were found in different tissues suggesting that they may have acquired different functional roles. However, in the medaka *lyg* duplicates were co-expressed in tissues suggesting that enzyme distribution and function may be species specific. No digital expression data was available for the teleost *lalba* transcript.

The expression of lysozymes was analysed in the transcriptome of an important innate immune barrier, teleost skin (Table 2.4). In the seabass skin transcriptome, *lyg1* and *lyg3* were expressed and no transcript for *lyz* was identified. In seabass, *lyg3* is relatively more abundant than *lyg1* suggesting that they may have different roles in skin. In contrast, in the skin transcriptome of the black rockcod and Senegalese sole, transcripts for *lyz, lalba* and *lyg* were detected. In the black rockcod, the two *lyg* transcripts had a similar abundance and are relatively less expressed than *lalba*. In the Senegalese sole skin transcriptome *lyz, lalba* and two *lygs* were identified. The human *LALBA* homologue (Table 2.4) in the sole skin transcriptome was the most abundant lysozyme transcript and *lyg1* was more abundant than *lyg*2 and *lyz* was the least expressed.

2.4.7.1. Expression in gilthead sea bream embryos and larvae

Expression analysis by qPCR during early development in sea bream revealed that *lyg1*, *lyg2* and *lalba* were expressed in embryos and early larval stages but *lyg3* and *lyg4* transcripts were not detected (Figure 2.8 and Supplementary Figure 2.4 annex I). Changes in lysozyme expression were also detected between embryos and developing larvae from different sea bream brood stock [that originated juveniles with a good (G) or poor (P)] growth performance. Transcripts of *lyg1*, *lyg2* and *lalba* were up regulated in larvae that originated juveniles with a poor growth performance compared to those with a good growth performance (Figure 2.8).

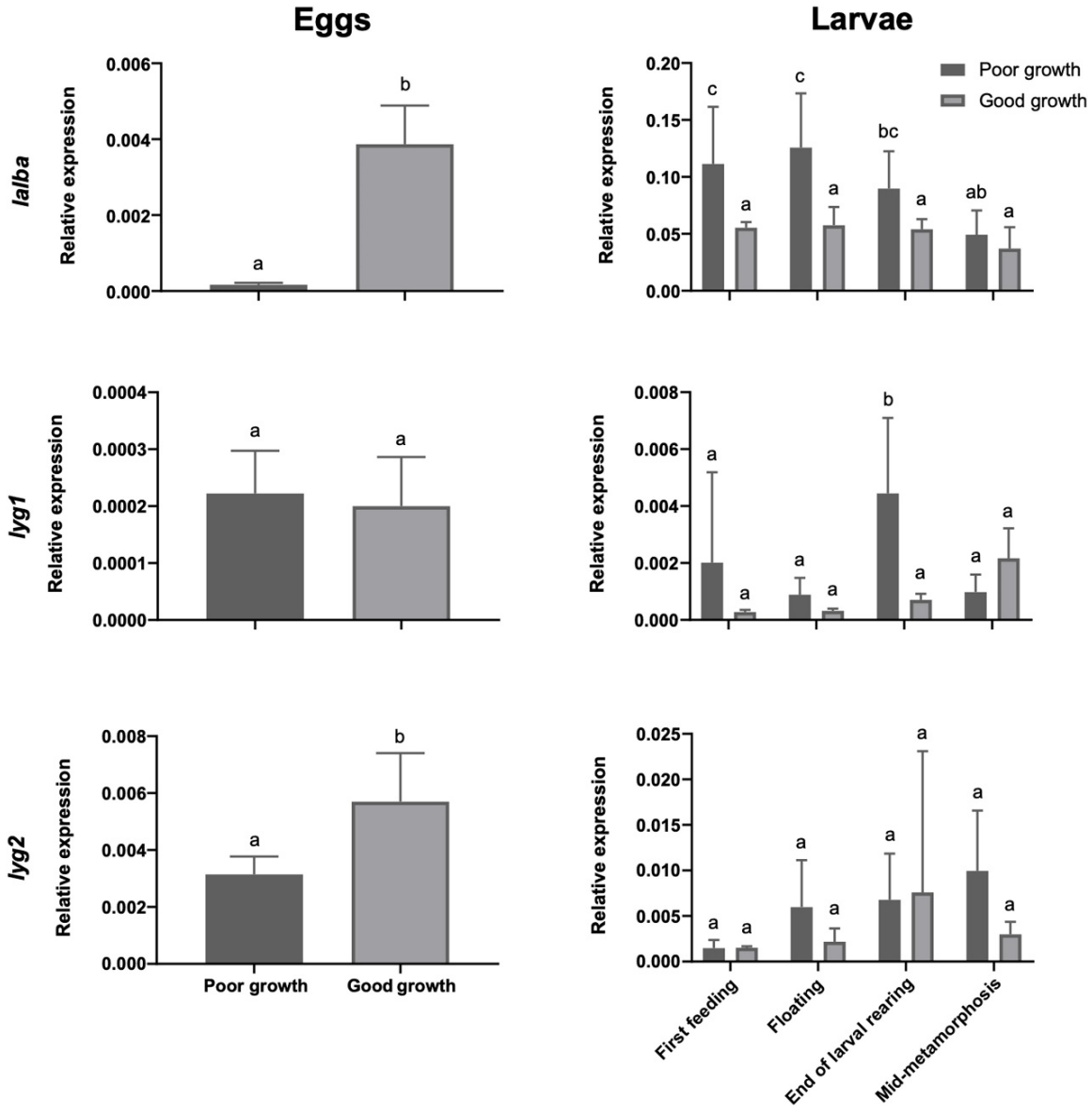


Figure 2.8. Relative expression of *lalba* and two *lyg (lyg1*, *lyg2*) during gilthead sea bream development. Embryos and larvae were from two brood stock that showed a divergent growth performance by mid-metamorphosis (classified as good and poor) were compared (good growth performance - light grey bars and poor growth performance - dark grey bars). Data corresponds to the mean ± SEM of three to six different

samples and gene expression levels were normalized using the geometric mean of two reference genes (*18s* and *ef1-α*). SPSS 25.0 software was used to assess the significance of differences between the experimental groups using an unpaired student t-test (two-tailed) for the embryos and two-way ANOVA for the different larval stages. Bars with different letters are significantly different ($p < 0.05$).

In gilthead sea bream embryos, a different gene expression pattern was observed and *lyg2* and *lalba* were significantly up regulated in the embryos and larvae of G juveniles compared to embryos that originated P juveniles. In developing larvae of G juveniles, no significant differences in expression were detected for the lysozyme genes (Supplementary Figure 2.4 annex I). In the developing larvae of P juveniles *lyg1* was significantly up regulated at the end of larval rearing compared to earlier larval stages (Supplementary Figure 2.4 annex I). *Lalba* transcripts were significantly downregulated in mid-metamorphosis larvae compared to earlier larval stages (Supplementary Figure 2.4 annex I).

Comparison of larvae at a similar developmental stage from the different brood stocks revealed *lyg1* expression was significantly up regulated at the end of larval rearing in the P juveniles compared to the larvae of the G juveniles (Figure 2.8). The *lalba* expression was significantly up regulated up until the end of larval rearing in larvae originating P juveniles compared to larvae originating G juveniles (Figure 2.8).

2.4.8. Total lysozyme activity

Measurement of lysozyme activity during gilthead sea bream, white sea bream and meagre early development revealed that enzyme activity is present before hatch and is maintained in early developmental stages of all species and is highly variable between individuals of the same species (Figure 2.9). Before hatch, no significant differences were found in lysozyme activity between species. At hatch, lysozyme activity in the embryos of gilthead sea bream was significantly higher than after hatching. Lysozyme activity was significantly lower in embryos after hatching in gilthead sea bream compared to the white sea bream and meagre.

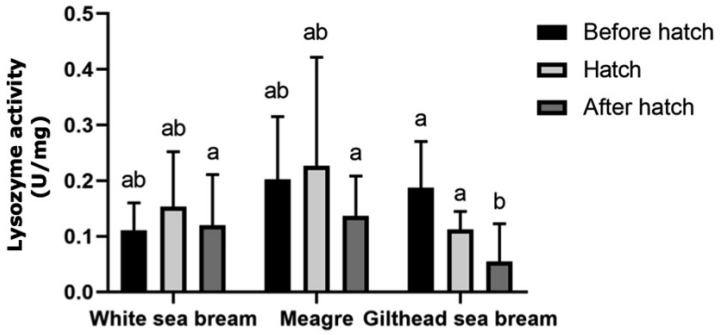


Figure 2.9. Lysozyme activity during early developmental of different teleost species. Enzyme activity was measured directly using total embryo protein extracts from white sea bream (*Diplodus sargus*), meagre (*Argyrosomus regius*) and gilthead sea bream (*Sparus aurata*) from three different developmental stages: before hatching, at hatching and after hatching. Enzyme activity was measured using its ability to lyse Gram-positive bacteria *Micrococcus luteus*. Significant differences were assessed using a two-way ANOVA followed by a Tukey's multiple comparisons tests. Samples with no detected enzymatic activity were removed from the analysis. Bars with different letters are significantly different ($p < 0.05$).

2.5. Discussion

The present study revealed that the ancestral genes for vertebrate LYZ and LYG emerged prior to the protostome and deuterostome divergence but that gene evolution of the two lysozyme -types was very dynamic. Despite the relatively low levels of sequence conservation between the lysozyme -types in vertebrates the residues associated with protein structure and catalytic function have been conserved. Expression analysis and enzymatic activity assays revealed that in the teleosts lysozymes are expressed early in development and the pattern of enzyme activity during development is species-specific. A surprising result was the identification in teleost embryos and larvae of a putative functional sequence homologue of mammalian *LALBA*.

2.5.1. Diversity of lysozymes in fish probably results from distinct evolutionary events

Homologues of mammalian lysozymes C-type and G-type were found in several teleosts and other fish genomes and in common with mammals, gene family evolution was affected by lineage-specific and species-specific events (Irwin 2014). In fish as in other vertebrates, members of this gene family have rapidly evolved as revealed by changes in their sequence and the lack of conservation of their gene environment. However, amino acids essential for enzyme function and structure remain conserved. In fish the two catalytic residues Glu (E) and Asp (D) characteristic of human LYZ have persisted, along with the four calcium binding residues (D, Asp) characteristic of the mammalian LALBA and the residues Glu (E) and Asp (D) that are involved in LYG structure and function (Brew *et al.* 1967; Qasba & Kumar 1997). Moreover,

during evolution many of the amino acid substitutions result in residues with the same physicochemical properties indicating there was pressure to maintain the proteins chemical characteristics. In the majority of fishes, members of lysozyme C-type and G-type were found but fewer lysozyme C-type genes exist in relation to the G-type genes and this is because in most teleosts increased *lyg* gene number resulted from recent tandem gene duplication events. In teleosts, orthologues of three mammalian lysozyme C-types namely, *Lyz*, *Lyzl1/2* and *Lalba* were found based on sequence conservation but the others mammalian forms are likely to be absent from fish. The lack of gene synteny found across the teleost and tetrapod (human and chicken) genome regions is intriguing and suggests that members of this gene family map to highly modified chromosomes. In the ancestral bony fish genome considerable rearrangements and chromosome fusion events are suggested to have occurred after the divergence of the tetrapod lineage (Nakatani *et al.* 2007). It is likely that during these events the teleost ancestral genes were transposed to another genome/chromosome region as homologous genome regions that lack *lyz*, *lalba* and *lyg* gene but that possess a similar neighboring gene repertoire to those flanking lysozyme gene family members in the human and chicken are found in teleost genomes. The differences observed between the teleosts are likely to be explained by the different rates of chromosomal rearrangements in each species (Kasahara *et al.* 2007; Ravi & Venkatesh 2008). The exception was the zebrafish *lyzl1/2* genes which share a similar gene environment with the human *SPACA5* and S*PACA5B* and with *LYZL1/2* genome regions suggesting that the cypriniform and the human genes may have a common genome origin. Zebrafish lack the homologue of the vertebrate *Lyz* and no *lyzl1/2* genes were found in other teleosts except cypriniforms. This suggests that prior to the teleost and tetrapod divergence gene precursors for both gnathostome lyz and *lyzl1/2* existed in the genome and that both copies were retained in humans, but one copy was lost in teleosts. The human Lyzl1 and Lyzl2 resulted from a recent gene duplication event and map in close proximity. Lyzl1 and Lyzl2 gene transcripts are abundant in human testis and in zebrafish gene transcript were found in testis and in other tissues (Table 2.3) suggesting that they may have common functions in reproduction. Conservation of the gene environment between spotted gar *lyz* on LG11 with the human *LYZL4* on chromosome 3 is intriguing and suggests the genes may have shared a common origin, the absence of the *lyzl4* gene in teleosts suggests it was eliminated from the genome during the teleost radiation.

A group of teleosts Lyg are likely to have emerged from the teleost specific genome duplication event. The existence of a teleost specific cluster was previously reported (Irwin

2014). The identified genes were designated LygF2 members and the other teleost sequences LygF1 because they were proposed to have arisen from an ancient fish-specific (F) gene duplication. Herein, we have renamed teleost Lyg genes following the nomenclature established for genes duplicated in the teleost-specific tetraploidization event, namely Lyga (LygF1) and Lygb (LygF2). While most teleost have members of both lysozyme families in the cod only *lyg* are present and members of the lysozyme C-type family are absent. In contrast, in the Nile tilapia and in the spotted gar only lysozyme C-type members are present. Analysis of the spotted gar genome suggests that this gene was deleted from LG17, which has a highly similar gene content to the homologue genome region on human chromosome 2 that hosts *LYZ*. Interestingly, in the cod a gene family expansion of Lyg has occurred and in addition to the four genes previously identified (LygF1a-F1d renamed Lyg1-4) (Seppola *et al.* 2016) we identified a further three genes, suggesting this may be a compensatory mechanism for the lack of lysozyme C-type.

A further gene of the lysozyme family has been described in mammals, LALBA, which shares high sequence similarity with other lysozyme C-types but has a different biochemical function (Qasba & Kumar 1997). The identification in the present study of an LALBA homologue in the gilthead sea bream (*Sparus aurata*) and some but not all of the teleost species analysed raises questions about how common it is in fish and its biological function. In mammals LALBA is similar in sequence (40%) and structure to LYZ but lacks bacteriolytic activity (McKenzie & White 1991; Nitta & Sugai 1989). In humans, this protein is secreted by the mammary gland and regulates lactose synthesis however, its presence in the gut has been linked to digestive functions (Pettersson-Kastberg *et al.* 2009). The teleost LALBA homologue shares conserved sequence motifs with the human form, such as the calcium binding residues and it also lacks the catalytic activity ascribed to LYZ (Irwin *et al.* 2011). Interestingly, LALBA in fish is highly abundant in the skin transcriptome and in embryos and larvae and deciphering its function in the future will be of considerable interest.

In summary, our extended analysis of the fish lysozymes suggests that a large diversity of proteins are encoded in fish genomes and the existence of a variable gene number, overall poor sequence conservation and gene synteny suggests they have evolved under lineage and species-specific pressures. Identification in ray-finned fish of sequence homologues of several members of the mammalian lysozyme C-type family (*lyz*, *lyzl1/2*, *lalba* and potentially *lyzl4*) (Irwin *et al.* 2011) in the present study substantiates a previous study suggesting they evolved much earlier in the gnathostome evolution and prior to the ray-finned and lobe-finned fish divergence

(Dautigny *et al.* 1991).

2.5.2. Lysozymes emerged prior to the protostome-deuterostome divergence

In molluscs, representatives of the two vertebrate lysozymes (C-type and G-type) and the specific invertebrate LYI have been reported in several species (McHenery *et al.* 1979; McDade & Tripp 1967). For example, homologues of human LYZ was first reported in abalone (*Haliotis discus hannai*) (Ding *et al.* 2011) and LYG in the Zhikong scallop (*Chlamys farreri*) where it was proposed to be a multifunctional lysozyme (Zhao *et al.* 2007; Van Herreweghe & Michiels 2012). The invertebrate specific LYI was first isolated and characterised from the Japanese clam (*Ruditapes philippinarum*) and shown to be important for defence against bacteria but was also suggested to have digestive functions (Ito *et al.* 1999; Pipe 1990; Foley & Cheng 1977). Overall, our phylogenetic analysis suggested that the mollusc and vertebrate C-type and G-type lysozymes shared a common origin. The absence of LYI from vertebrate genomes but its presence in the genome of invertebrate deuterostomes suggests that during the vertebrate radiation the gene was lost. Arthropods contain LYZ and LYI genes and nematodes only LYI and neither possess a LYG gene (Bachali *et al.* 2002).

The presence of the three lysozyme-isotypes, G, C and I, in molluscs further substantiates the proposal that the genomes of species of this phylum are less rearranged than that of other protostomes and that they possess a more similar gene repertoire to deuterostomes than other invertebrates (Simakov *et al.* 2013). In bivalves, a greater number of lysozyme I-type and G-type members were identified. However, homologues of vertebrate Lysozyme C-type were detected in very few species of bivalves and gastropods. It is currently unclear if the three metazoan lysozyme isotypes (G, C and I) shared a common origin as their sequence conservation is very low, although it is clear that all emerged prior to the protostome and deuterostome divergence. Moreover, although the lysozymes are only one of several gene families associated with pathogen defence, their persistence in the genomes of phylogenetically distant organisms, namely early protostomes to advanced vertebrates suggests they have a critical role in innate immune defence.

2.5.3. Lysozyme activity in teleosts

The vast diversity of the fishes and the divergent evolution of *lyz* and *lyg* genes in the species analysed, means knowledge about the evolution and function of these enzymes is still poor. Lysozymes are anti-bacterial enzymes and their functional role has mostly been associated

with innate immune defence in fish (L. Tort 2003). In fish, lysozymes have a widespread and divergent tissue distribution. For example, the C-type transcript is most abundant in the liver of Asian seabass (*Lates calcarifer* Bloch, 1970), in the head kidney, gill and brain of orange-spotted grouper (*Epinephelus coioides*) and in the spleen, kidney and gill of the olive flounder (*Paralichthys olivaceus*) and catfish (*Ictalurus punctatus*) (Fu *et al.* 2013; Lee *et al.* 2014a; Wang *et al.* 2013b; Wei *et al.* 2012). The G-type lysozyme in Asian seabass and orange-spotted grouper is predominantly expressed in intestine and head kidney, respectively and in the catfish, it is abundant in the spleen, kidney and gill and is constitutively expressed in most tissues of the olive flounder. Bacterial, viral or a lipopolysaccharide challenge modulate the expression of both isoforms of lysozyme (Fu *et al.* 2013; Lee *et al.* 2014a; Wang *et al.* 2013b; Wei *et al.* 2014). Despite the divergent tissue distribution of lysozymes in fish a common feature is their abundance in tissue of the immune system suggesting this is their core function (Panase *et al.* 2017; Subbotkina *et al.* 2018; L. Tort 2003).

In vertebrates although lysozymes have an important action in the immune system they also play other roles such as cellulose hydrolysis during digestion in ruminants (Callewaert & Michiels 2010) and type-C lysozyme is the dominant form in the digestive system of terrestrial vertebrates (Irwin & Wilson 1990; Jimenez-Cantizano *et al.* 2008; Harikrishnan *et al.* 2011). Although in fish lysozyme activity is associated with immune tissues such as spleen and kidney (L. Tort 2003; Panase *et al.* 2017; Subbotkina *et al.* 2018) the high expression of *lyz* in the stomach and liver of brill (*Scophthalmus rhombus*) (Jimenez-Cantizano *et al.* 2008) has led to the suggestion that like other vertebrates it may also have functions in digestion in fish.

In our study *lyz*, *lalba* and *lyg* transcripts had variable presence and abundance in the skin transcriptomes of several species (Table 2.4). This is unsurprising when considering the importance of the skin as a barrier between fish and the external environment. However, the bacteriolytic activity of lysozyme in skin mucous from different species is highly variable and if this is related to the main isoforms expressed remains to be established (Saurabh & Sahoo 2008). The difference in lysozyme activity in Coho and Atlantic salmon has been related to the relative importance of lysozyme in the innate immune response of the two species (Fast *et al.* 2002).

2.5.4. Lysozyme activity in teleost early development

Few studies have considered the importance of the lysozyme family in fish early development. The studies that exist indicate, as observed in the present study, that enzyme

activity differs between embryos and larvae and also across species (Cecchini *et al.* 2000). Enzyme activity has been detected in oocytes, fertilized eggs, and larval stages of several fish species such as in *Plecoglossus altivelis*, *Tribolodon hakonensis* and *Salmo gairdneri*, (Kudo & Teshima 1991; Kudo 1992; Brown *et al.* 1997). In the sea bass, Caspian kutum (*Rutilus kutum*) and Siberian sturgeon (*Acipenser baerii*) the lysozyme activity decreased from fertilization to the early larval stages and increased again at later stages (Cecchini *et al.* 2000; Abdollahi *et al.* 2016; Valipour *et al.* 2018). Such observations have led to the proposal that the lysozyme RNA and protein in fish eggs and embryos are maternally derived (Wang & Zhang 2010; Yousif *et al.* 1994; Lee *et al.* 2014a; Seppola *et al.* 2009). The lysozyme along with maternal immunoglobulins, complement proteins and anti-protease (Seppola *et al.* 2009; Nayak *et al.* 2011; Mulero *et al.* 2007) are proposed to play a key role in protecting fish embryos from bacteria during early development when adaptive immunity is still not fully established and functioning (Magnadottir *et al.* 2005). During embryonic stages lysozyme activity contributes to prevent vertical transmission (from mother to offspring) of bacterial pathogens and disease (Yousif *et al.* 1994). In some fish species, maternally inherited mRNAs are selectively degraded prior to activation of embryonic transcription and in Atlantic cod *lyg* gene transcription was only activated in late stages of the gastrula period (118 hpf) and subsequently up-regulated post-hatching (Seppola *et al.* 2009). While in the olive flounder (*Paralichthys olivaceus*) *lyz* but not *lyg* maternal RNA transcripts were detected in unfertilized oocytes and *lyg* was up-regulated in late neurula and strongly up-regulated post hatch (Lee *et al.* 2014a).

Our analysis of enzyme activity and expression confirmed that lysozyme is present in embryos and that activity may vary between species, which supports previous suggestions that some fish species have more effective embryonic antimicrobial mechanisms (Mulero *et al.* 2007). The expression of the lysozyme isoforms was modified during development suggesting their importance and role differs. Curiously, the expression of teleost *lalba* during gilthead sea bream development was similar to the bacteriolytic *lyg* in the embryo and larval stages, which suggests it probably has a role in teleost fish development. The divergent expression of lysozyme between embryos and larvae of different gilthead sea bream brood stock and the link to subsequent juvenile growth performance is intriguing. It may be an innate immune parameter that should be considered along with other factors reported to contribute to embryo quality such as photoperiod (Taranger *et al.* 1998), stress (Schreck *et al.* 2001), water temperature and quality (Davies & Bromage 2002) and feeding (Bobe 2015a). Future work will be required to establish the contribution of lysozyme to embryonic quality.

2.6. Conclusion

Lysozymes are an ancient family of enzymes involved in innate immune defence. Three main gene families exist and have rapidly evolved. In the vertebrate lineage species-specific gene family expansions occurred and members of the same lysozyme gene family share poor sequence conservation and lack a conserved genome environment. The results obtained for each lysozyme-type suggest that evolution was rapid and motivated by distinct biotic and abiotic pressures and that the ancestral lysozyme genes were transposed to a different genome region in the ancestral bony fish genome. Analysis of teleost data revealed for the first time the existence of putative homologues of mammalian LYZL1/2 and LALBA in some teleost genomes, but their function remains to be elucidated. Lysozyme activity and expression were detected in fish embryos and early developmental stages. In the gilthead sea bream lysozyme transcript expression was significantly different between eggs and larvae from different brood stock. Overall, our results contribute to the understanding of the evolution of the vertebrate innate immune system and specifically to the lysozyme system in fish.

CHAPTER 3

The evolutionary analysis of complement component C5 and the gene co–expression network and putative interaction between C5a and C5a anaphylatoxin receptor (C5AR/CD88) in human and two Cyprinid fish

Lisen Li[1,&], Weining Yang[1,&], Yubang Shen[1,*], Xiaoyan Xu[1], Jiale Li[1,2,*]

[&] *Contributed equally to the study*

[1]*Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, China*

[2]*Shanghai Engineering Research Center of Aquaculture, Shanghai Ocean University, Shanghai 201306, China*

Manuscript published in Developmental & Comparative Immunology, 2021, 116: 103958
https://doi.org/10.1016/j.dci.2020.103958

Role of Lisen Li: the bioinformatics analysis for gene isolation, phylogenetic tree construction, codon usage bias analysis, gene co-expression and protein-protein interaction network analysis, homology modeling and protein docking analysis.

Acknowledgements

We thank all the laboratory members for their assistance in the experimental analysis. We also thank FreeScience Information Technology Company (Zhejiang province, China) for proofreading, and editing the grammar in the text. This research was supported by the China's Agricultural Research System (CARS-45-03), the Project of Shanghai Engineering and Technology Center for Promoting Ability (16DZ2281200).

3.1. Abstract

The complement system is a complex network of soluble and membrane-associated serum proteins that regulate immune response. Activation of complement C5 generates C5a and C5b which has a chemoattractive effect on myeloid cells and initiates the membrane attack complex (MAC) assembly. However, the study of evolutionary process and systematic function of C5 are still limited. In this study, we performed an evolutionary analysis of C5. Phylogeny analysis indicated that C5 sequences underwent complete divergence in fish and non-fish vertebrate. It was found that codon usage bias improved and provided evolutionary evidence of C5 in species. Notably, the codon usage bias of grass carp was closer to the zebrafish genome compared with humans and stickleback. This suggested that the zebrafish cell line may provide an alternative environment for heterologous protein expression from grass carp. Sequence comparison showed a higher similarity between human and mouse, grass carp, and zebrafish. Moreover, selective pressure analysis revealed that the *C5* genes in fish and non-fish vertebrates exhibited different evolutionary patterns. To study the function of C5, gene co–expression networks of human and zebrafish were built which revealed the complexity of C5 functional networks in different species. The protein structure simulation indicated that C5 from grass carp and zebrafish are more similar than to human, however, differences between species in C5a proteins are extremely smaller. Spatial conformations of the C5a–C5AR (CD88) protein complex were constructed, which showed that interactions may exist between C5a and CD88 proteins. Furthermore, the protein docking sites/residues were measured and calculated according to the minimum distance for all atoms from C5a and CD88 proteins. In summary, this study provides insight into the evolutionary history, function and potential regulatory mechanism of C5 in the fish immune responses.

Keywords: Innate immune system; Fish; Protein interaction; Spatial conformation

3.2. Introduction

The vertebrate immune system is composed of two parts; innate immunity and adaptive immunity (Medzhitov & Janeway 1997). The complement system contains about 35 individual plasma proteins and cell receptors, it also regulates host defenses by interacting with components of the innate and adaptive immunity (Holland & Lambris 2002; Fishelson 1991). The complement system and its components have been investigated in several fish species such as jian carp (*Cyprinus carpio* var. Jian) (Yuan *et al.* 2017), half–smooth tongue sole (*Cynoglossus semilaevis*) (Li & Sun 2017), rainbow trout (*Oncorhynchus mykiss*) (Wu *et al.* 2014), gilthead sea bream (*Sparus aurata* L.) (Guardiola *et al.* 2018), grass carp (*Ctenopharyngodon idella*) (Dang *et al.* 2016). In teleost grass carp, 34 members have been identified in the complement system (Liao *et al.* 2019).

Complement component C5, C3, and C4 belongs to a single gene family and which first diverged from a common ancestor of C3 and C4 (Gehring *et al.* 1987; Hughes 1994). C5 is the fifth complement component formed by alpha and beta polypeptide chains that are linked by a disulfide bridge (Merle *et al.* 2015). Several evolutionary studies on C5 in fish have focused on the phylogeny, gene structure, and synteny (Johansen *et al.* 2019; Lv *et al.* 2020; Liao *et al.* 2019). Findings from such studies have been contradictory with regard to the evolutionary origin of C5. A prior study found that the C5 of cartilaginous fish nurse shark (*Ginglymostoma cirratum*) was more similar to the mammalian C5 than to C5 of teleost fish at the amino acid level. This suggested an evolutionary origin of the C5 complement factor that predates the emergence of sharks (Graham *et al.* 2009). Currently, no comprehensive evolutionary study has been performed on C5.

The complement polypeptide chains have distinct functions in different biological processes. C5a is an anaphylatoxin derived from the alpha polypeptide of C5 that potentially exerts spasmogenic and chemotactic activity (Wu *et al.* 2014). Han et al. (2010) proved that C5a functions by effectively amplifying IL–8 expression in human during sepsis (Wang *et al.* 2010). Besides, Huber-Liang et al. (2000) demonstrated that C5a improved and increased the survival rate by disrupting C5a and C5a receptor in rat and mouse sepsis models (C5AR) (Guo *et al.* 2000). Multiple functions of C5a have been reported including superoxide release, release of granule enzymes and cytokines. In addition, C5a is located in neutrophils, monocytes, and macrophages (Manthey *et al.* 2009). C5b is derived from beta-polypeptide chains that participate in the assembly of membrane attack complex (MAC) by C5b–C9 which possess

cytolytic activity (Ricklin *et al.* 2010; Graham *et al.* 2009). In general, C5b exerts functions on MAC intra-regulation and C5a has been proposed to participate in additional pathways. C5AR and C5L2 have been reported as two functional receptors of C5a, and the consequence of C5a activation is dependent on the location of C5a receptors (Manthey *et al.* 2009). C5AR is also known as CD88. A previous study reported that several functional domains/residues are involved in the interaction between C5a and CD88 (Monk *et al.* 2007). In human and rats, residue Asp282, at the extracellular face of helix 7, has been reported to interact with the Arg74 side chain of C5a (Cain *et al.* 2001; Cain *et al.* 2003). Of note, loop 1 (residues C5a 12–20), loop 3 (C5a 39–46), and the C-terminal 6–8 residues (especially Arg74) are important for binding to C5AR in mammals. Based on the above reports, it is important to construct spatial conformation of C5a–CD88 protein complex in experimental animals to explore the mechanisms of interaction between the complement system and receptors.

Protein−protein recognition and interactions play important roles in many biological processes (Huttlin *et al.* 2017; Li *et al.* 2018; Jiang *et al.* 2017). Studying protein structure could offer a valuable perspective for understanding the regulation of biological process via protein interaction. As proteins evolve from a common ancestor, their sequences and structures diverge from each other (Povolotskaya & Kondrashov 2010). The observed distribution of protein structures can also give us important clues about the underlying evolutionary process (Goldstein 2008). In terms of existing studies, the comparison of protein structures between species from different evolutionary positions is not well employed at present. With more detected the protein's structures at high resolution and the applications of more analyzing models and websites, this work will provide valuable information for gaining the knowledge of the association between protein structure and evolution. However, the current experimental methods are not effective in detecting the structure of protein complexes. Protein docking technology is a promising approach for simulating and predicting the structure of a protein complex (Smith & Sternberg 2002; Huang 2015; Halperin *et al.* 2002). The technology selects valid ligand-protein complexes showing the required surface complementarity. The retained poses are transformed to the real Cartesian space to implement site constraints and atomic scoring (Kong *et al.* 2019). This allows the study of the spatial conformation of a protein complex. The protein data bank (PDB) file from protein data bank website is available and it is crucial for studying and understanding the structures of protein and protein complex. The available PDB file is also fundamental for building the protein structure. Overall, this methodology could also provide theoretical support to structures of unknown proteins or

undiscovered interactions between two proteins.

Grass carp is the primary aquaculture species in China (Wang *et al.* 2017). According to the data from The Food and Agriculture Organization (FAO), the global production of grass carp was approximately 6.07M tons in 2016. A previous study deciphered the grass carp genome and reported significantly expanded gene families, which included many immune–associated functional domains consistent with the adaption of the grass carp to variable environments (Wang *et al.* 2015). Notably, such characteristics imply that grass carp is considered as a fish representative; thereby, is a potential optimal model for immune evolutionary study.

In this study, the evolutionary process of C5 was analyzed based on phylogeny, codon usage, multiple sequences alignment, and selective pressure. The phylogeny of the species with C5 sequences was improved and verified through codon usage bias and multiple sequences alignment for the first time. The selective pressure analysis was conducted according to the different species clusters, which revealed fish and non–fish vertebrates that exhibited different evolution patterns. In addition, we simulated and built the species C5 protein structures and spatial conformations of the C5a–CD88 protein complex. The protein docking sites/residues were thereafter calculated. Notably, this study provides an insight into the evolutionary history and potential regulatory mechanism of C5 in the fish immune response.

3.3 Materials and methods

3.3.1. Phylogeny

Phylogeny of fish C5 amino acid sequences was generated from coelacanth (*Latimeria chalumnae*), spotted gar (*Lepisosteus oculatus*), medaka (*Oryzias latipes*), tetraodon (*Tetraodon nigroviridis*), stickleback (*Gasterosteus aculeatus*), rainbow trout (*Oncorhynchus mykiss*), tilapia (*Oreochromis niloticus*), zebrafish (*Danio rerio*), common carp (*Cyprinus carpio*), grass carp (*Ctenopharyngodon idella*), half–smooth tongue sole (*Cynoglossus semilaevis*), Japanese flounder (*Paralichthys olivaceus*). Another two cartilaginous fishes the elephant shark (*Callorhinchus milii*), whale shark (*Rhincodon typus*) were used for the outgroup. The human (*Homo sapiens*) and mouse (*Mus musculus*) were selected for mammals. Chicken (*Gallus gallus*) and turtle (*Pelodiscus sinensis*) were represented for the reptiles/birds. Amphibian group was represented by xenopus (*Xenopus tropicalis*). All sequences were retrieved from the online Ensemble (http://asia.ensembl.org/index.html) and National Center for Biotechnology Information Search database (NCBI; https://www.ncbi.nlm.nih.gov/) websites. Phylogenetic

tree was built using Maximum-Likelihood (ML) in PhyML 3.0 on ATGC platform (Guindon *et al.* 2010). The ML tree was constructed using a WAG matrix and 100 bootstrap replicates. ML tree was displayed in FigTree software (http://tree.bio.ed.ac.uk/software/figtree/) and edited with Inkscape.

3.3.2. Codon usage bias

The codon usage bias analysis was performed according to the published methodology (Fuhrmann *et al.* 2004). The relative adaptiveness plot of the codon was regarded as a crucial reference index to assess the codon usage bias by the value ranging from 0 to 100 in this study (Sharp & Li 1987). Mean differences were calculated between the genomes of each two species. Human, stickleback, zebrafish and grass carp genome were used as genome backgrounds based on the phylogeny for codon usage bias analysis. The codon usage bias of grass carp C5 amino acid sequence was also calculated.

3.3.3. Multiple sequences alignment

Deduced amino acid sequences of selected species were aligned using AliView software (Edgar 2004). The amino acid sequence similarity was calculated using the GeneDoc program (http://www.nrbsc.org/gfx/genedoc/). The physio–chemical property of the amino acid sequence was highlighted in different colors. Sequence similarity was yielded from the GeneDoc and visualized by heatmap. The sequence conservation was analyzed on CLUSTALW online website (https://www.genome.jp/tools-bin/clustalw). The signal peptide was predicted using the online software SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP/index.php) (Krogdahl *et al.* 2005).

3.3.4. Selective pressure analysis

The selective pressure analysis was performed according to the previous methodology (Weaver *et al.* 2018). The non–synonymous to synonymous nucleotide substitution rate (ω) was calculated for all selected nucleotide acid sequences. The site–wise synonymous was estimated using the FEL (Fixed Effects Likelihood) and SLAC (Single Likelihood Ancestor Counting) models. The non–synonymous / synonymous (d_N/d_S) rate ratio > 1 or < 1 represents for positive selection or negative selection from above two models, respectively. Positive and negative selected sites were detected and selected from both two analyzed models according to p–value < 0.05.

3.3.5. Gene co–expression network analysis

The human and zebrafish C5 were used to build gene co–expression network on GeneMANIA online services (Warde-Farley *et al.* 2010). The networks yielded from GeneMANIA website were further validated by the rebuilt C5 protein co–expression networks from STRING online service (Szklarczyk *et al.* 2019). The generated C5 co–expressed proteins were retrieved for GO (Gene Ontology) enrichment analysis on STRING online analysis. GO biological process was enriched and selected according to the FDR (False Discovery Rates) < 0.001. The GO enrichment dot bubble diagram was plotted by ggplot2 R program for data visualization. The program generated svg file was edited with Inkscape.

3.3.6. Protein structure analysis

The C5 amino acid sequences of human, zebrafish and grass carp were selected to build the protein structure according to the published methodology (Waterhouse *et al.* 2018). Human C5a chain was taken from full–length C5 sequence according to previous human C5 study (Manthey *et al.* 2009). The sequences of putative zebrafish and grass carp C5a chains were also taken out from full–length C5 sequences based on human C5a. The protein structures of three species were simulated separately.

3.3.7. Protein complex simulation

The putative CD88 amino acid sequences were collected by selecting the species from constructed phylogenetic tree of C5a. All putative CD88 amino acid sequences were used to build the phylogenetic tree. The simulated protein interaction between C5a and CD88 of human and grass carp were carried out according to the previously established methodology (Kong *et al.* 2019). The PDB files of C5a and CD88 from human and grass carp were download from protein data bank and they were used for structural simulation. Human and grass carp CD88 sequences were regarded as receptor proteins, C5a is smaller as ligand protein inputted. Before data processing, it is essential to anticipate and fix any potential PDB error in the Molprobity website (http://molprobity.biochem.duke.edu/). The top ten conformations of protein complexes of each species will be simulated and acquired.

3.4. Results

3.4.1. C5 Phylogeny

The amino acid sequences of the complement C5 species were used to construct the phylogenetic tree using the ML method. All sequence information is shown in the Supplementary Table 3.1 annex I. Two main groups of the phylogenetic tree successfully diverged based on the 100 bootstrap values between fish (pink region) and non-fish (blue region) vertebrate clusters (Figure 3.1). Coelacanth C5 diverged before teleost and non-fish (blue region) vertebrate clusters after the cartilaginous fish. Xenopus clustered with a fish branch at a relatively low bootstrap value (62). Common carp showed two similar C5 isoforms (C5a and C5b) and clustered with grass carp. Furthermore, the codon usage bias was calculated based on the mean differences between the xenopus genome and other species (Figure 3.1). Of note, the mean difference values between the genomes of xenopus and other species were yielded except for no useful available information on the spotted gar genome. The mean differences between xenopus and species in the Teleost–1 cluster were relatively lower than Teleost–2 and non-fish vertebrate clusters, indicating slight affinity.

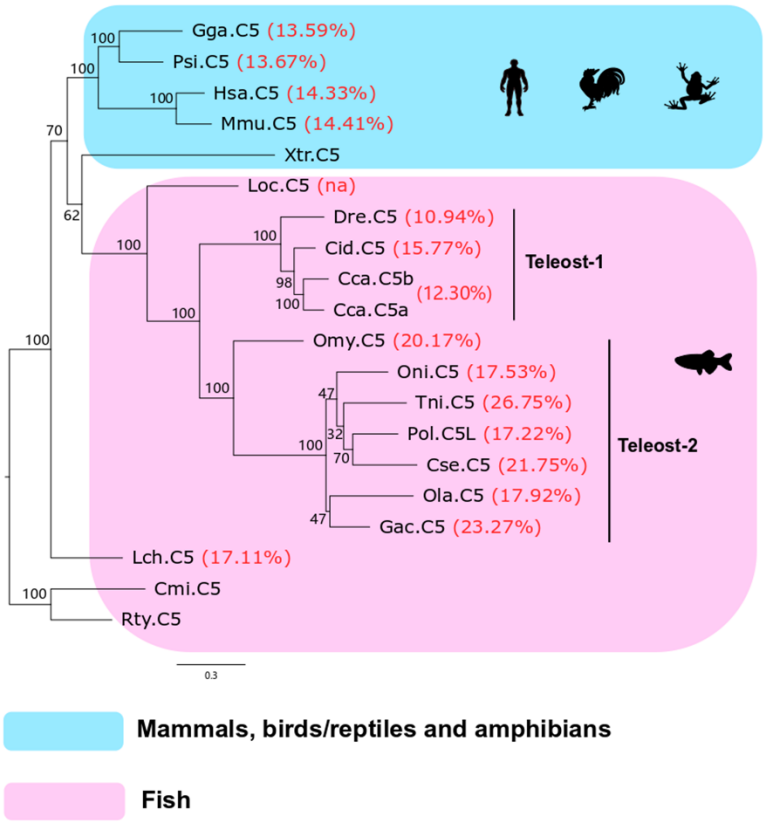


Figure 3.1. Phylogeny of vertebrate C5 amino acid sequences. Maximum likelihood (ML) method was used to run the phylogenetic tree. Branch points were validated by 100 bootstrap replications. The cartilaginous

fishes were used to root the tree. The codon usage bias based on mean differences between the genome of xenopus and other species were displayed in percentages and marked by red color in the bracket. “na” represent the useful genome information is not available.

3.4.2. C5 codon usage

The codon usage bias analysis was conducted on four species, including human, stickleback, zebrafish, and grass carp, as shown in Figure 3.2 and Supplementary Figure 3.1 annex I. Distinct codon usage bias in the genomes of different species was reported. Five conserved codons, including glutamine (Gln), leucine (Leu), methionine (Met), tryptophan (Trp), and valine (Val) were found in these species based on the highest value of amino acid relative adaptiveness plot (Figure 3.2; Supplementary Figure 3.1 annex I). More codons were identified in zebrafish (8) and grass carp (7) compared with the stickleback (6) and human (5), (Figure 3.2). The codon usage bias in the genomes of four species was similar, but exhibited a slight difference. The mean differences between the genomes of the three species and grass carp C5 gene were 27.34 (human), 34.31 (stickleback), 23.20 (zebrafish), and 26.22 (grass carp), (Supplementary Table 3.2 annex I).

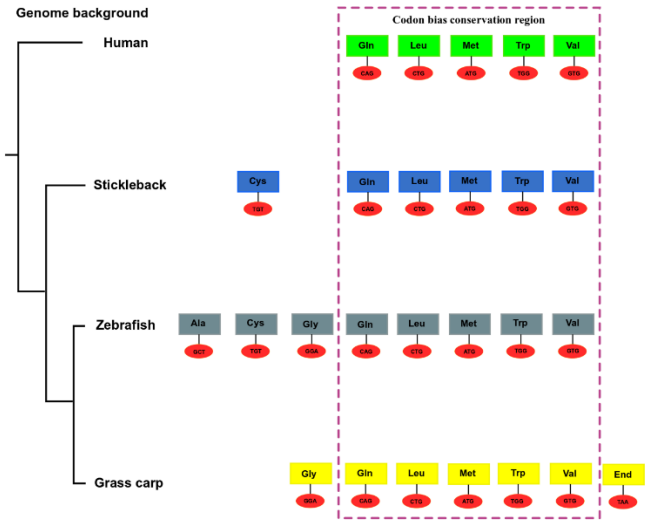
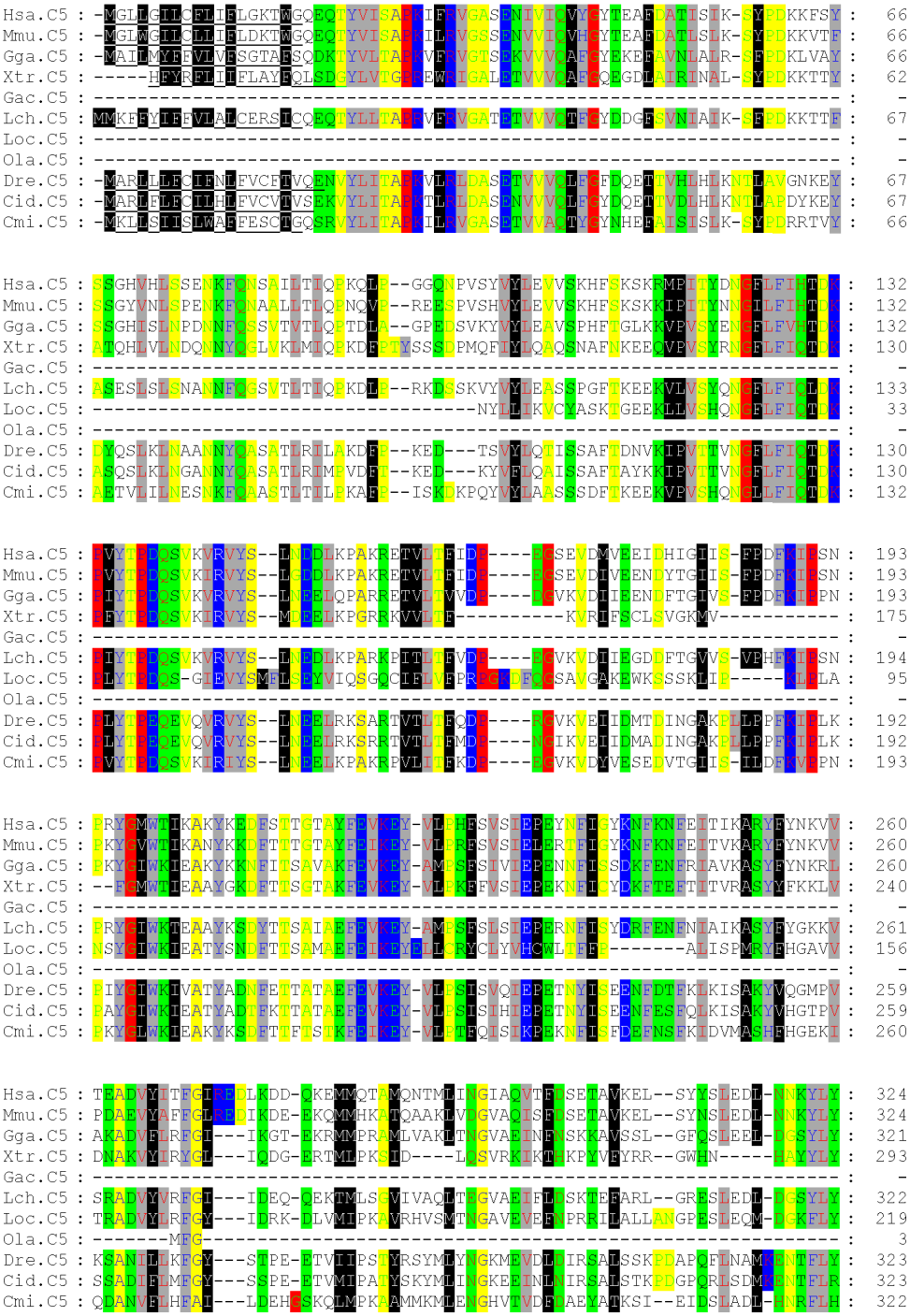


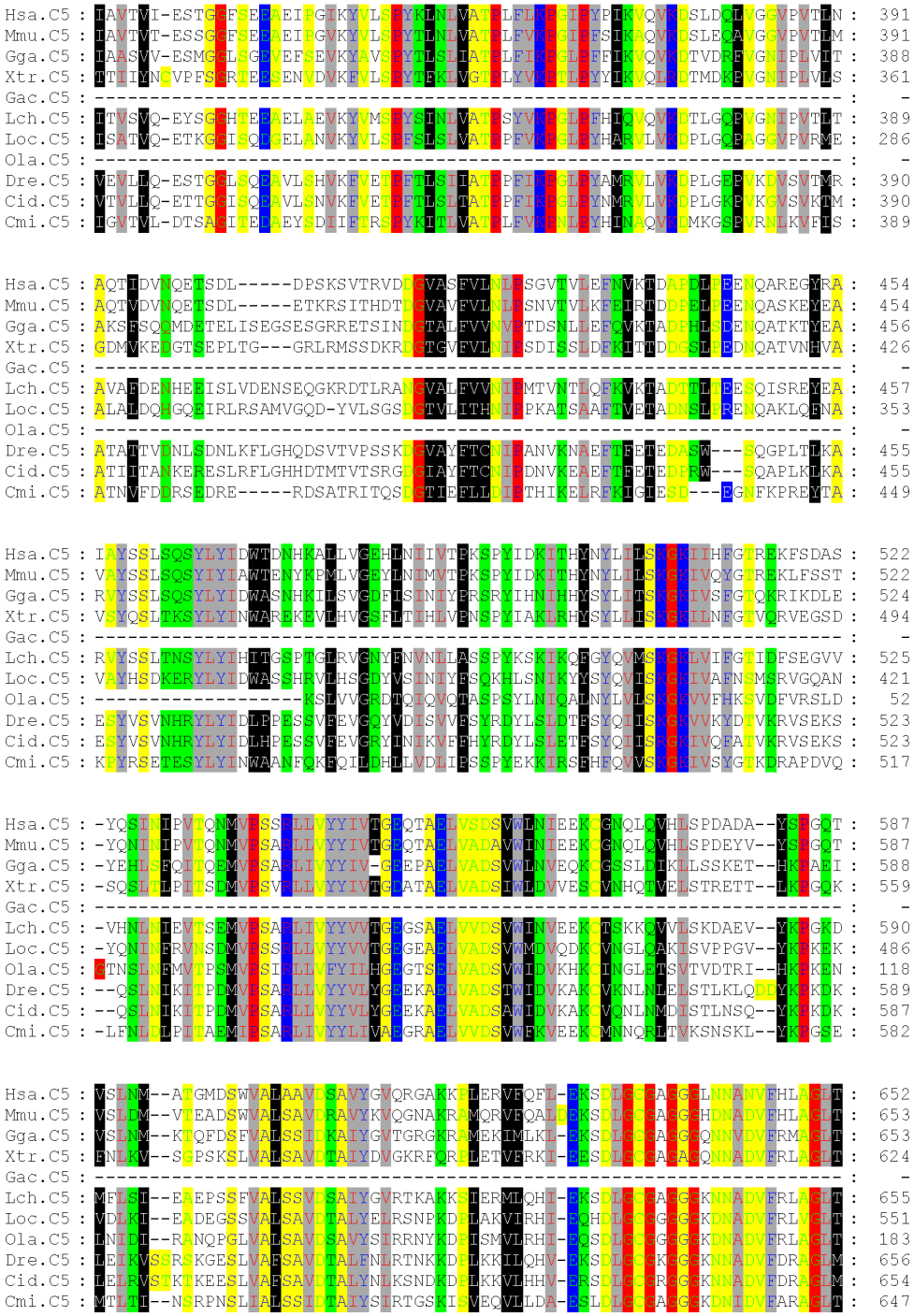
Figure 3.2. Codon usage bias analysis. The relative adaptiveness plot of codons from grass carp C5 sequence and other vertebrates’ genomes are both equal to 100 were selected. Each colorful square represents the amino acid of corresponding species. The red ellipsoid stands for the codon of each amino acid with the highest relative adaptiveness plot value (100) both in grass carp C5 and species genome.

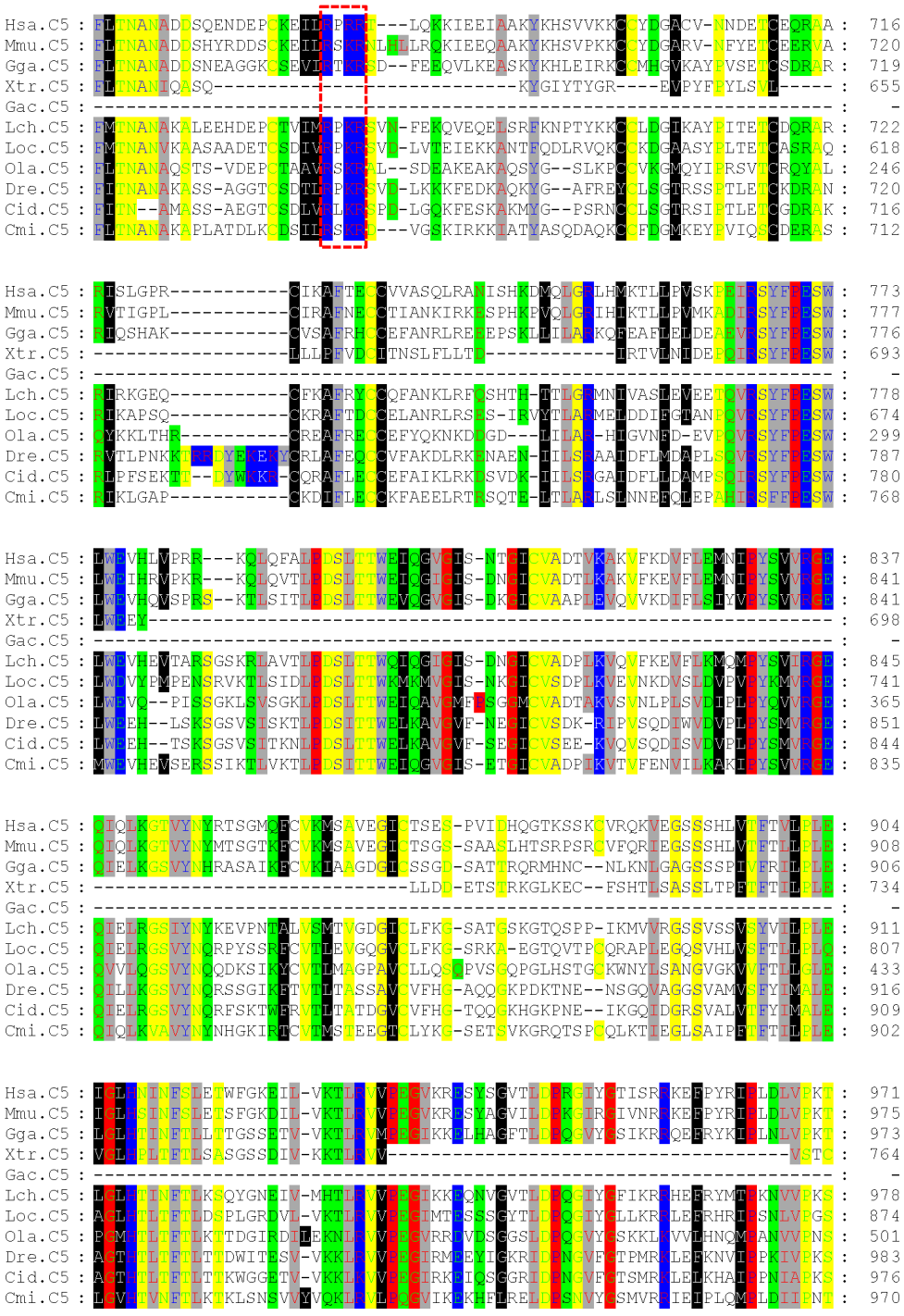
3.4.3. Multiple sequence alignment of C5

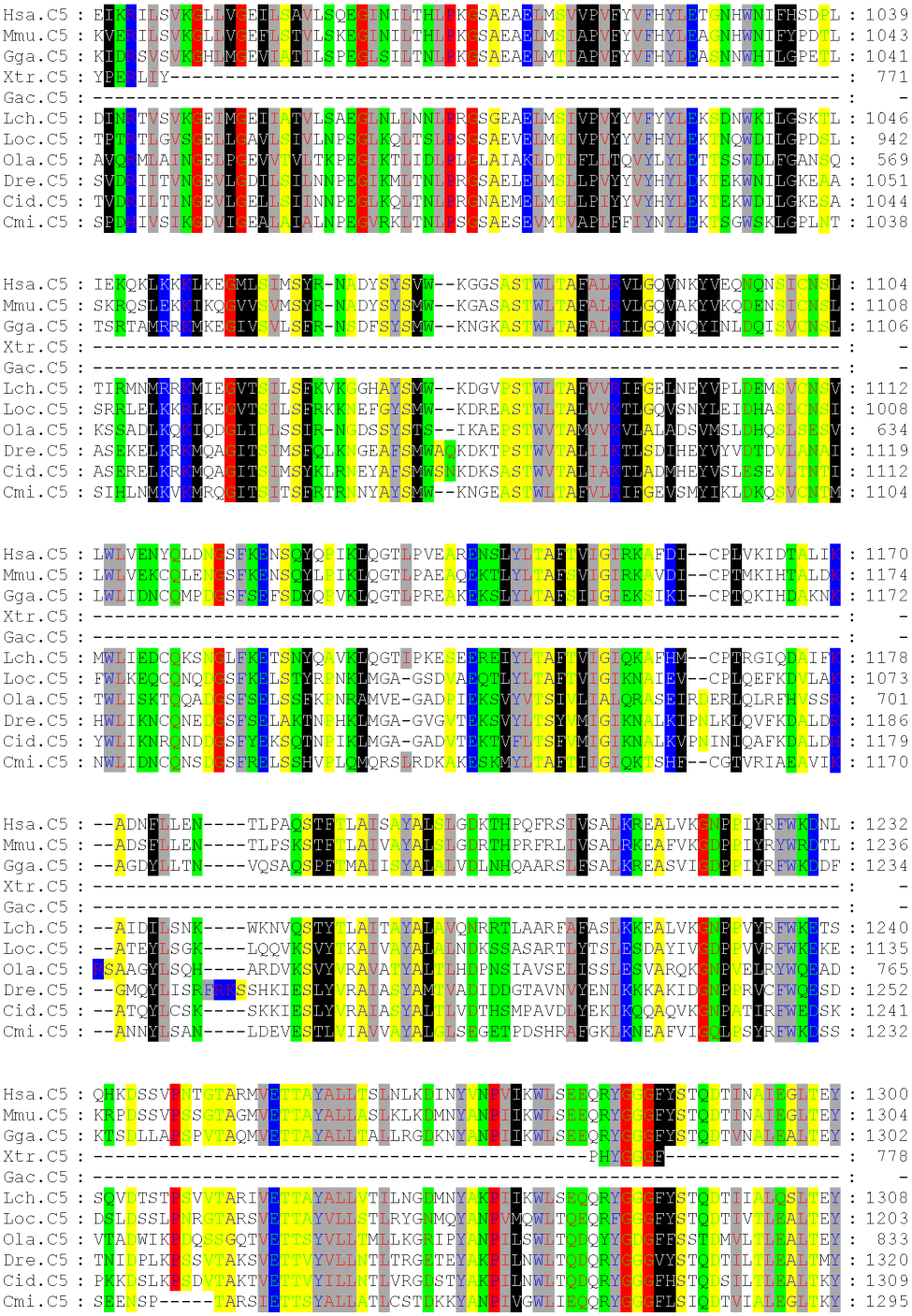
Moreover, multiple sequence alignment (MSA) of the C5 amino acid sequences was performed (Figure 3.3). Species amino acid sequences indicated the physicochemical properties and amino acid conservations. Notably, b–a cut sites were identified for all the selected species,

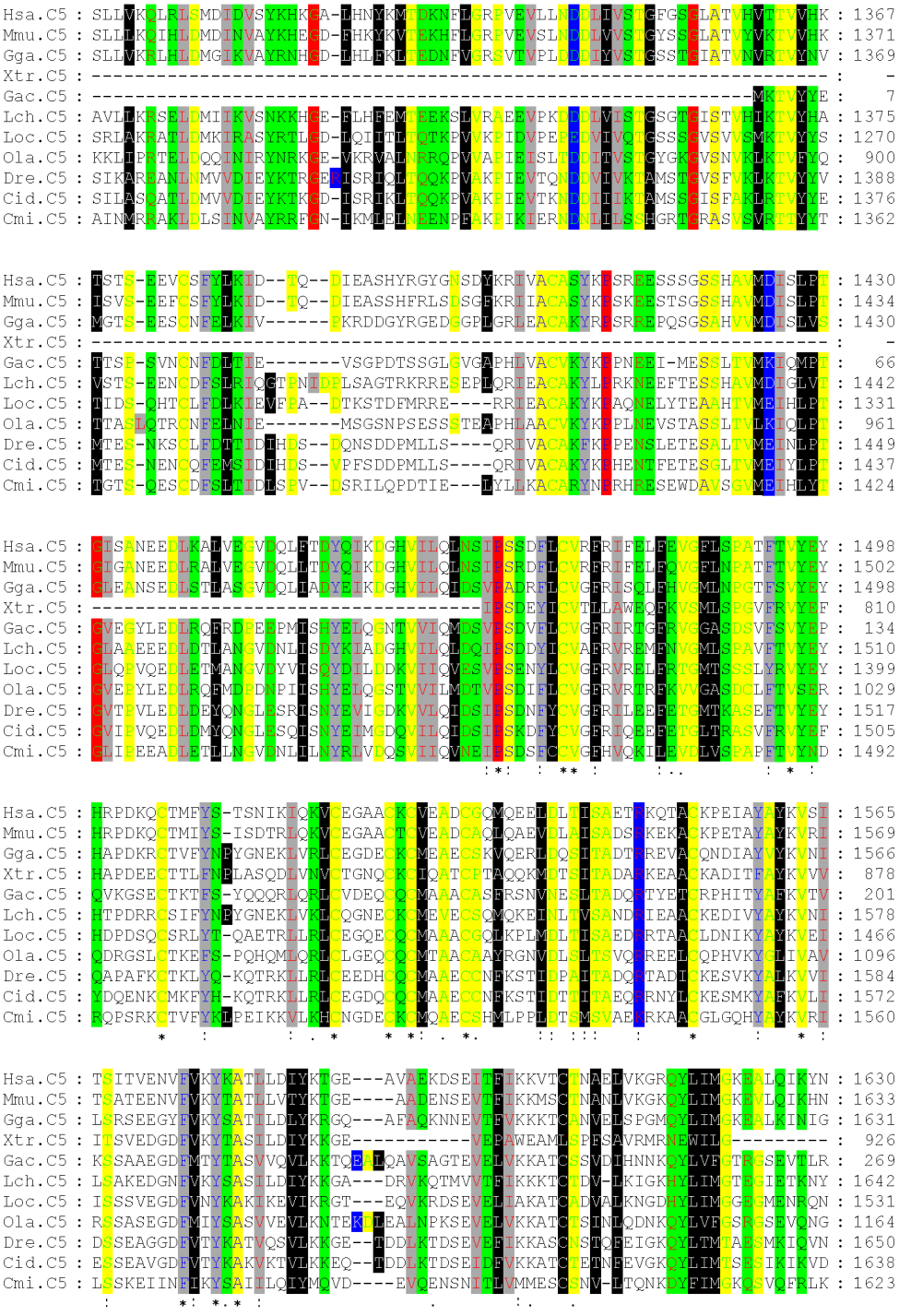
but were absent in xenopus and stickleback (Figure 3.3). The stickleback C5 sequence was short, incomplete, and only aligned to the homolog region of the human a chain. Besides, the C5 amino acid sequences were predicted to have signal peptides except for spotted gar, medaka, and stickleback (Figure 3.3). Species sequence similarities are displayed in Figure 3.4. Humans shared higher sequence similarity with the mouse (89 %) based on the heatmap results. The sequence similarity between humans and fish was between 10-63 %. Additionally, the sequence similarities were between 9-85 % in fish species. The highest similarity was observed between zebrafish and grass carp (85 %), whereas, the lowest similarity was observed between stickleback and elephant shark (9 %). The xenopus and stickleback showed relatively low similarity compared with other species, as indicated by the light blue region of the heatmap (Figure 3.4).











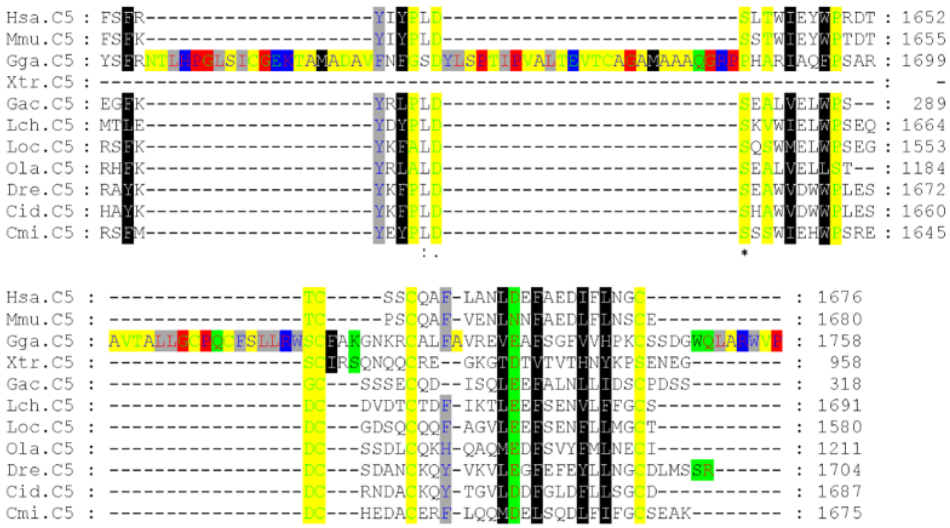


Figure 3.3. Multiple sequence alignment of complement C5. The predicted signal peptide is underlined at the beginning of the amino acid sequence. Gaps inside the sequence are marked by dashes. Complete residue conservation is annotated with a "*" and highlighted in bold. Partially conserved residues are denoted by "." and the position of the consensus amino acids present in the greatest number of sequences are indicated with ":". Alignment columns were highlighted using GeneDoc according to physiochemical properties (Text color/Shade color): Proline Blue/Red; Glycine Green/Red; Tiny Blue/Yellow; Small Green/Yellow; Positive Red/Blue; Negative Green/Blue; Charged White/Blue; Amphoteric Red/Green; Polar Black/Green; Aliphatic Red/Gray; Aromatic Blue/Gray; Hydrophobic White/Black). The b–a cut sites of C5 in the selected species are marked by red dash frame.

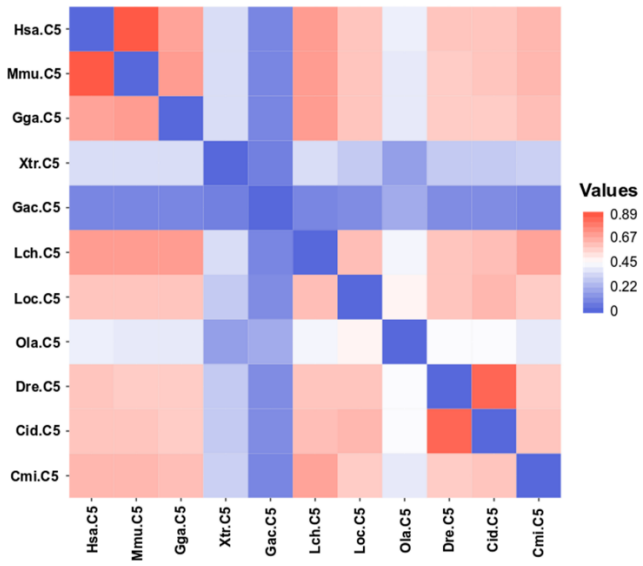


Figure 3.4. Sequences similarity heatmap. Each colored cell represents a sequence similarity value; Colors ranging from red to blue represent sequence similarity from the highest level to the lowest level, respectively; Values represent percentage by reserving two decimal places.

3.4.4. C5 selective pressure

The C5 selective pressure analysis was performed by separating the species groups. The nucleotide acid sequences of fish, non–fish vertebrate groups, and all species were analyzed using FEL and SLAC models (Table 3.1; Supplementary Table 3.3 annex I). The sites (amino acid residues) were calculated and retained (p–value < 0.05). The d_N/d_S for FEL and SLAC models in the fish group were close to 1. Notably, one amino acid residue 83 in the fish group was identified $d_N/d_S < 1$ from both two models with p–value < 0.05, no residue showed the d_N/d_S higher than 1. However, the d_N/d_S for non–fish and all species groups were lower than 1. Five sites in each group were identified from the two models with a p–value < 0.05. Besides, no site was identified in the non-fish group with $d_N/d_S > 1$, two sites were acquired from all species group with $d_N/d_S > 1$, p–value < 0.05.

Table 3.1. Selective pressure analysis of *C5* genes. ***a***: Non-synonymous/synonymous rate ratio is shown as d_N/d_S; ***b***: Numbers of sites are both successfully calculated and significant from FEL (Fixed Effects Likelihood) and SLAC (Single-Likelihood Ancestor Counting) models; Statistic p–value lower than 5 on same sites from two models was considered under selective pressure. Na: not available.

Genes	d_N/d_S (FEL model)	d_N/d_S (SLAC model)	Sites ($d_N/d_S < 1$)	Sites ($d_N/d_S > 1$)
***C5*–Fish**	1.00	1.03	83	Na
***C5*–Other vertebrates**	0.67	0.67	22, 89, 110, 206, 236	Na
***C5*–All species**	0.80	0.81	32, 56, 191, 527, 1404	453, 1234

3.4.5. C5 gene co–expression network analysis

Comparisons between human and Zebrafish *C5* gene co-expression and protein–protein interaction (PPI) network have been carried out and are displayed in Figure 3.5A - C. In human, a number of genes which were co–expressed with *C5* were found, in addition, C5 protein was found to interact with many proteins (Figure 3.5A and 3.5C). The gene co–expression network was built using the GeneMANIA website, and the result showed that the *C5* gene has many types of interactions with other genes. For instance, the direct interaction (e.g. Physical interactions) and indirect interaction (e.g. Co–expression) (Figure 3.5A). Human gene co–expression network was further confirmed using STRING PPI network (Figure 3.5C). Notably, 16 of 21 co–expressed genes were found from the GeneMANIA network and 16 of 41 interacted proteins were found from the STRING PPI network. The Zebrafish GeneMANIA network revealed that a few of genes had a physical interaction with C5. Moreover, most genes from the zebrafish GeneMANIA network shared same domains (Figure 3.5B). Only 2 common genes

from the GeneMANIA network were also found in STRING PPI networks, which are complement C6 (C6) and carboxypeptidase B2 (CPB2) (Figure 3.5B and 3.5D).

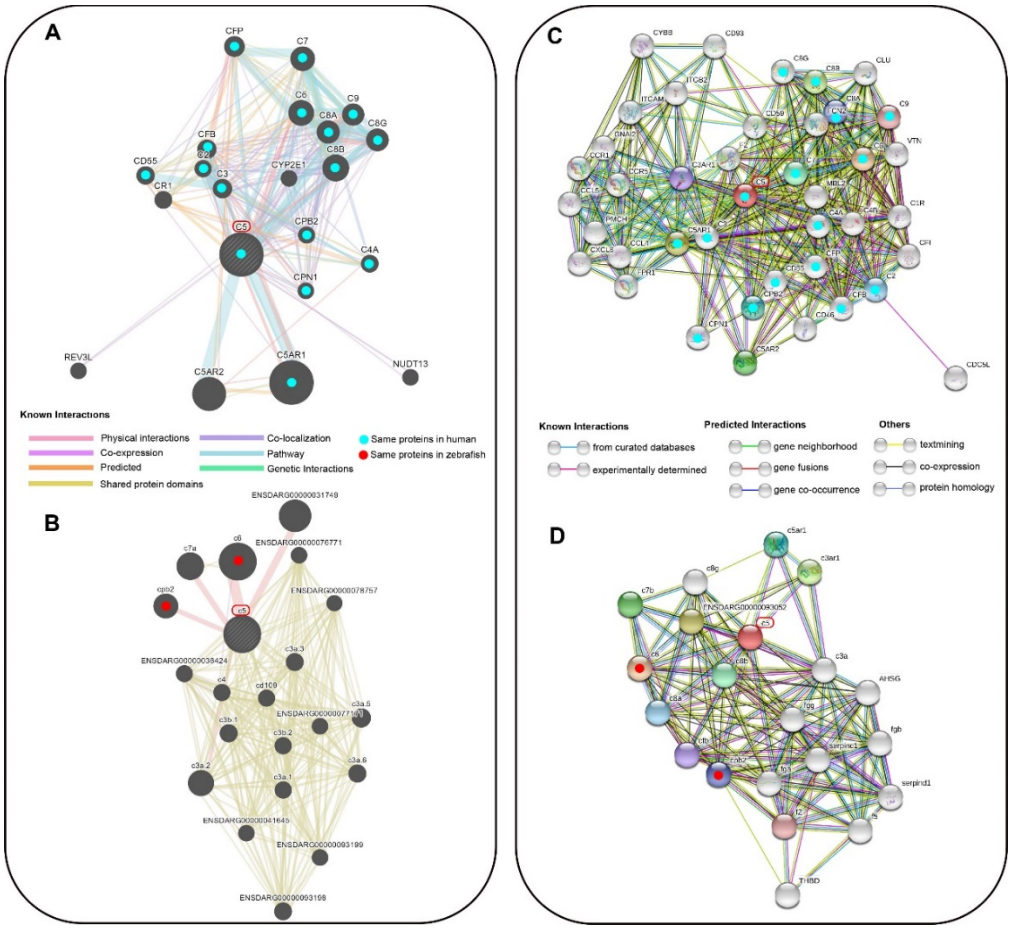


Figure 3.5. *C5* gene co–expression and PPI networks. A. Human *C5* gene co–expression network–GeneMANIA; B. Zebrafish *C5* gene co–expression network–GeneMANIA; C. Human C5 PPI network–STRING; D. Zebrafish C5 PPI network–STRING. Human and zebrafish C5 were denoted with red solid box. GeneMANIA network shows the relationships between genes in the list. Nodes represent genes and links represent networks, genes can be linked by more than one type of network. STRING network: Network nodes represent proteins, edges represent protein-protein associations. Node Color: Colored node represents query proteins and first shell of interactors; White node represents second shell of interactors; Node Content: empty node represents proteins of unknown 3D structure; Filled node represents some 3D structure is known or predicted.

GO biological process enrichment showed that the human C5 interacted proteins were involved in many regulatory processes; the complement, inflammation, immune, and response to external stimulus related processes were enriched (Figure 3.6A). The gene number distributed in each biological process was generally high. However, zebrafish C5 interacted proteins were also enriched and mapped to several biological processes (Figure 3.6B). The

complement receptor mediated signaling pathway (GO: 0002430) and inflammatory response (GO: 0006954) from zebrafish were also enriched in human. Other biological processes were associated with cellular signal transduction (GO: 0007200), regulation of cytosolic ion (GO: 0007204), and chemotaxis (GO: 0060326) (Figure 3.6B).

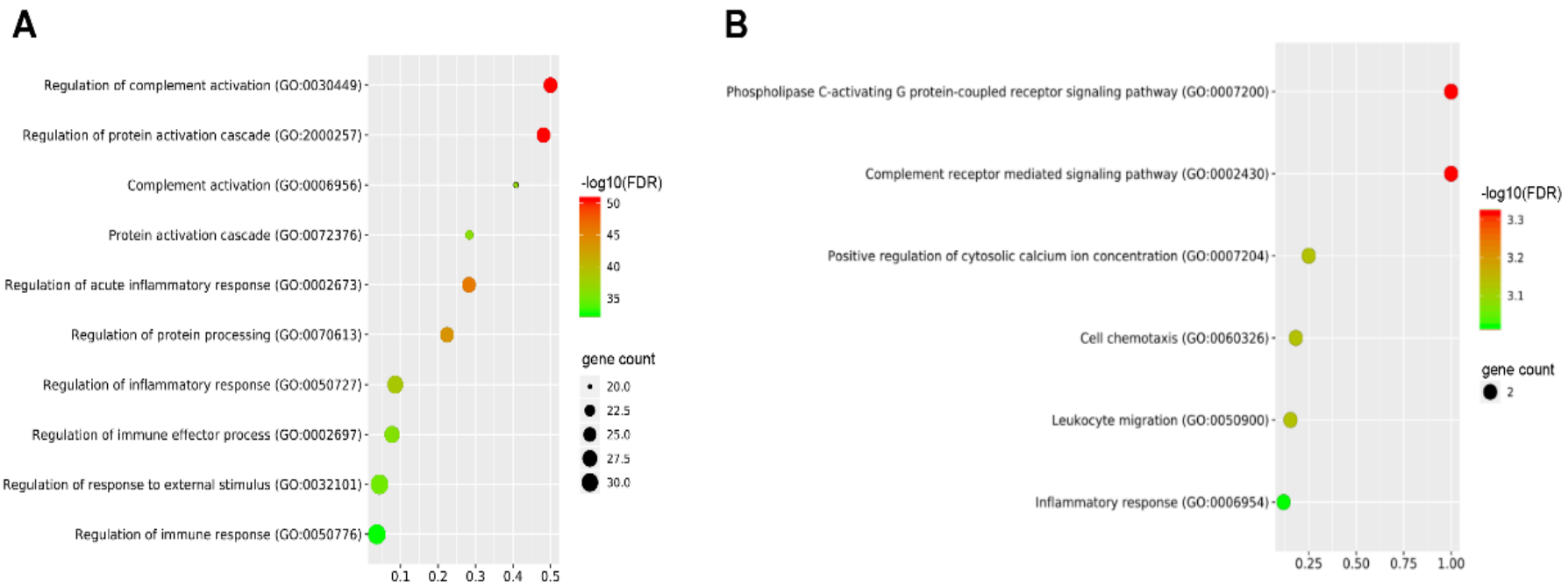


Figure 3.6. Dot bubble diagram of C5 GO biological process enrichment. A. Human C5 co–expressed GO biological processes; B. Zebrafish C5 co–expressed GO biological processes; GO pathway FDR < 0.05 was enriched and displayed. Gene number represents the number of co–expressed genes in this study. Annotations of Enriched GO biological processes were shown in y–axis. X–axis represents gene ratio (co–expressed gene count/background gene count).

3.4.6. Protein structure of C5 and C5a–CD88 complex

The protein structures of humans, zebrafish, and grass carp complement C5 were simulated (Figure 3.7). Similar protein structures were reported between zebrafish and grass carp; however, different conformations were reported between fish and human. The b–a cut sites were identified and were found to be located in the human C5 protein between the a and b chain. However, the b–a cut sites were not successfully simulated in the C5 protein structures in zebrafish and grass carp, suggesting that it may be disconnected (green dashed lines) between the a and b chain. The spatial conformations of C5a protein were similar among humans, zebrafish, and grass carp (pink frame). Before structural simulation, we firstly built the phylogenetic tree for identifying the authenticity of CD88 sequences (Supplementary Figure 3.2 annex I; Supplementary Table 3.4 annex I). According to the tree, two main branches were found after the appearance of coelacanth. The mammals, birds and reptiles were clustered and fish were another individual group. Grass carp putative CD88 sequences were clustered with common carp and zebrafish in fish group. The protein complexes of the human and grass carp C5a–CD88 was simulated (Figure 3.8). The C5a amino acid sequences were acquired from the human and grass carp full–length C5 sequences (Supplementary Figure 3.3 annex I). The top 10 spatial conformations of C5a–CD88 protein complex were successfully simulated in human and grass carp, respectively (Figure 3.8). Human CD88 was input as protein dimer, different from the monosome of grass carp CD88. Multiple spatial conformations of the human and grass carp C5a–CD88 protein complex were observed. Notably, when the CD88 position was fixed, the potential docking regions and C5a sites in CD88 were several and distinct.

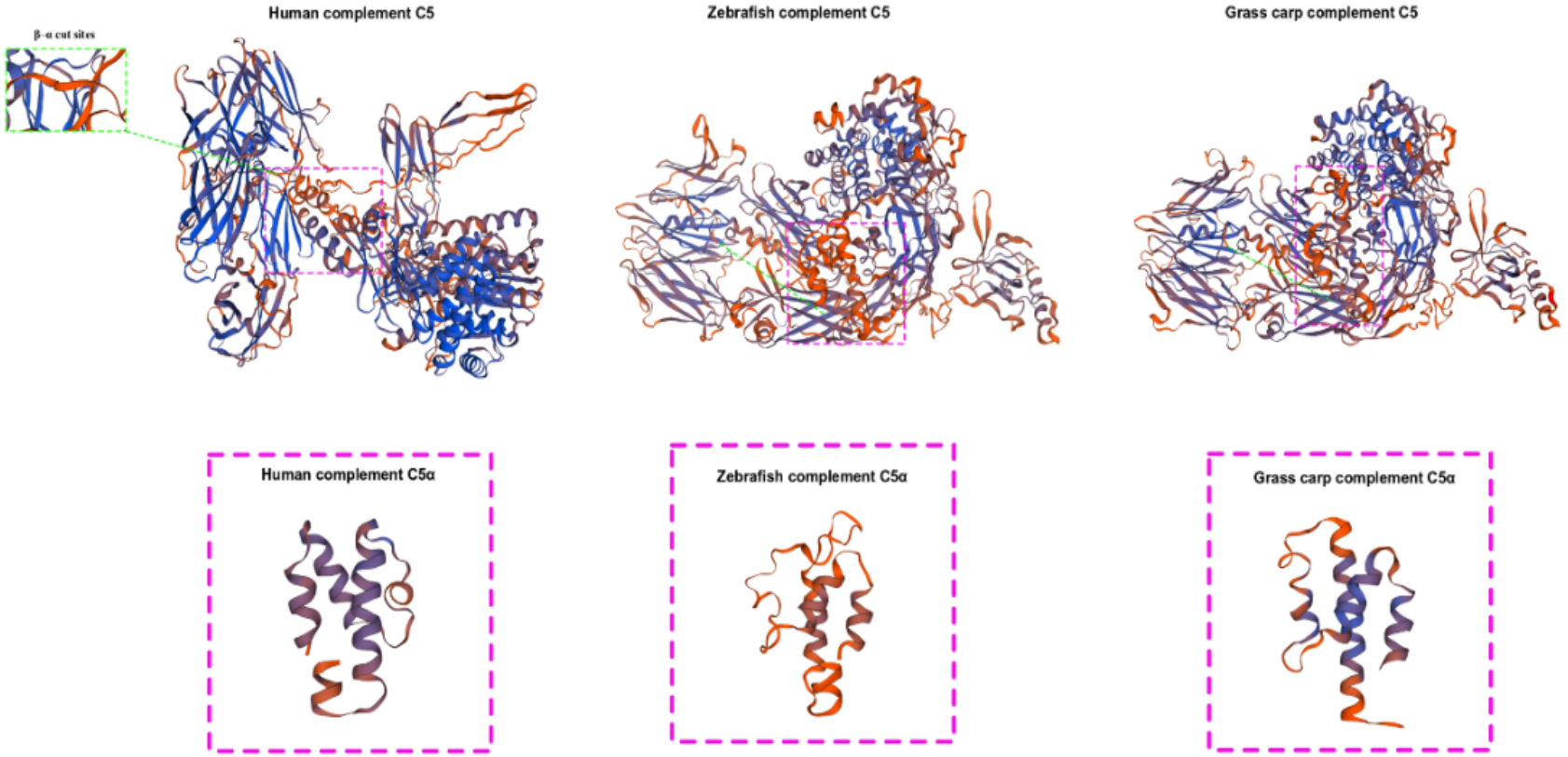


Figure 3.7. Protein structures of humans, zebrafish and grass carp C5. Protein structure was visualized by a colorful ribbon diagram. The b–a cut sites in a human was marked by green dash frame, in zebrafish and grass carp were absent and connected by green dash lines. Protein structures of C5a were marked by the pink frame.

Chapter 3

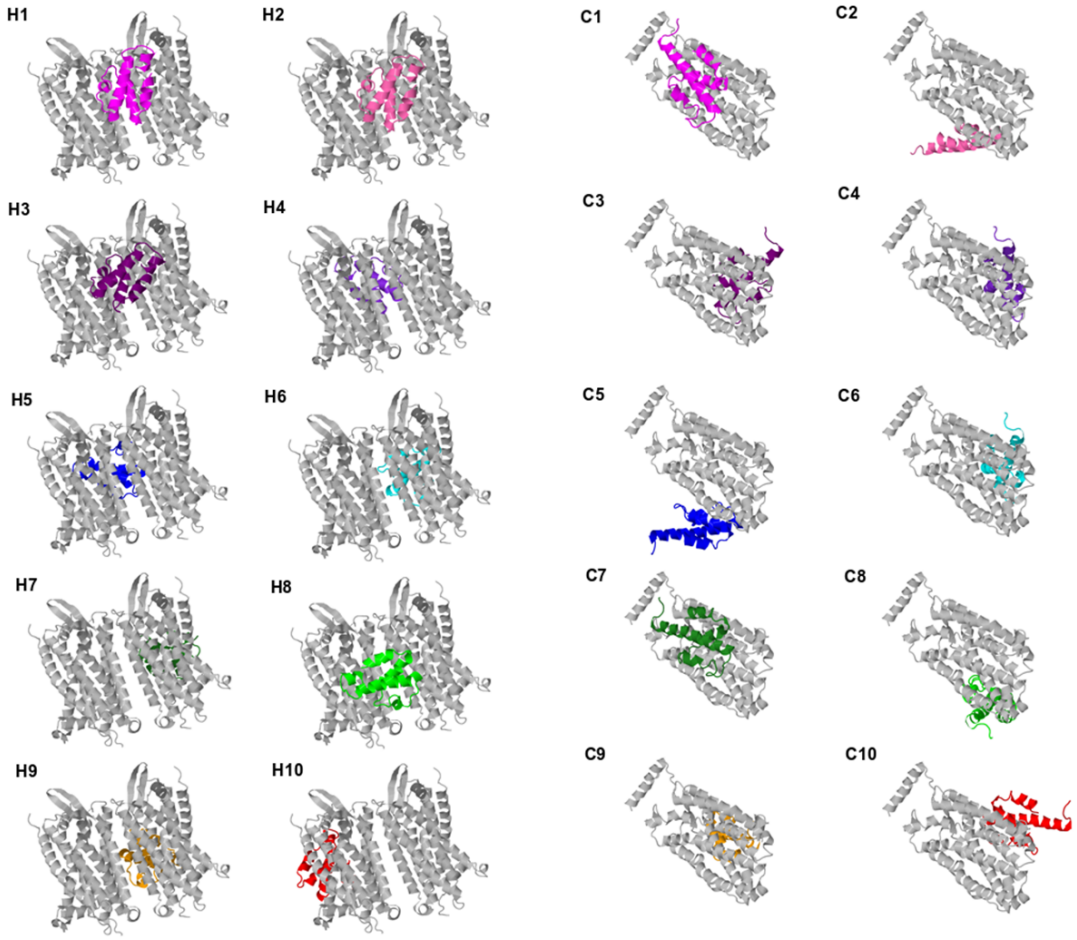


Figure 3.8. Simulated spatial conformation of human and grass carp C5a–CD88 protein complex. CD88 protein was denoted by uniform silver in human and grass carp, different docking patterns of C5a were marked by different colors; Human CD88 was inputted as protein dimer; grass carp CD88 was inputted as protein monosome. H1–H10 represents human top 10 spatial conformations of simulated C5a–CD88 protein complex; C1–C10 represents grass carp top 10 spatial conformations of simulated C5a–CD88 protein complex.

Additionally, the spatial conformations of grass carp C5a–CD88 protein complex and protein docking sites/residues were analyzed and calculated (Figure 3.9). Two protein conformations C1 and C7 were selected among the top 10 protein complex models according to the 290Ala interaction site from CD88 protein. Three structures of the human CD88 homolog region, including Loop 1, loop 3, and the C-terminal 6–8 residues are important for binding to the C5a receptor (C5AR); therefore, homolog regions were mapped and located in the grass carp C5a protein. The 290Ala residue from the simulated CD88 protein in C1 and C7 models were calculated to dock to the 28Thr, 72Val from C5a protein at the minimum distance of 1.101 and 1.291 nm, respectively.

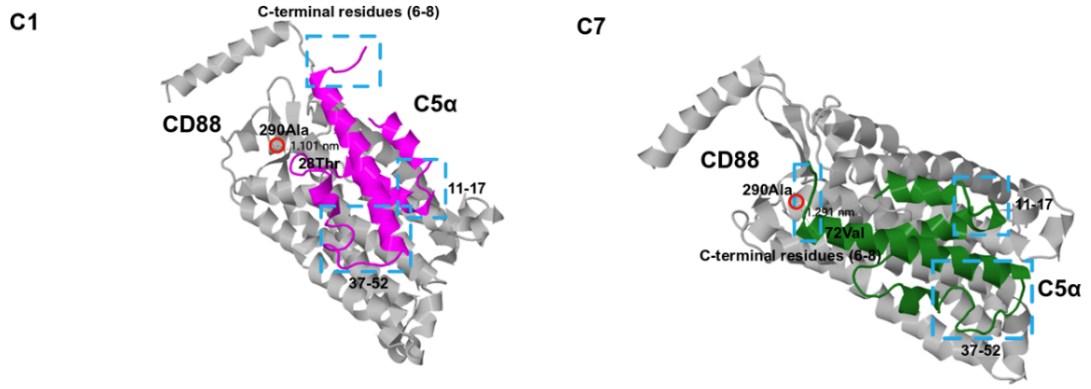


Figure 3.9. Two simulated spatial conformations of grass carp C5a–CD88 protein complex. C1 and C7 represent first and seventh simulated conformations of C5a–CD88 protein complex from ten simulated conformations in total according to the 290Ala interaction site from CD88 protein. CD88 protein was denoted by uniform silver in grass carp, different docking patterns of C5a were marked by different colors; Blue frame represents the important region of C5a for binding to C5AR; Red circle stands for the human homologue Asp282; Minimum atom distance is labeled and marked by dashes between two closer atoms.

3.5. Discussion

The complement system is an integral part of the innate immune system that consists of several soluble and membrane-bound proteins (Muller 1988; Volanakis 1998; Lambris *et al.* 1999). Complement C5 plays important roles in two major processes, i.e., the production of anaphylatoxins (C5a) and the formation of a membrane-attack complex (MAC), which lyse microorganisms (C5b) (Hughes 1994). Several studies have investigated the vertebrate complement C5. However, limited reports exist on the systemic evolutionary and structural study of C5 in vertebrates, particularly in fish. In this study, phylogenetic analysis of the C5 in vertebrates suggested that it underwent complete divergence. Codon usage bias analysis showed that the grass carp C5 gene has closer codon usage to the genome of zebrafish and grass carp compared with the stickleback and humans. Besides, the multiple sequence alignment confirmed that deduced amino acid residues of the human b–a cut sites exist in all the selected species except for xenopus and stickleback. High sequence similarity was observed between grass carp and zebrafish (85 %). Moreover, selective pressure analysis indicated that fish underwent neutral selection, whereas, non-fish vertebrates underwent negative selection pressures. Spatial conformations of C5 in zebrafish and grass carp were similar but different from the human. Moreover, this study simulated the top 10 different spatial conformations of C5a–CD88 in human and grass carp, respectively, suggesting that this possible protein interaction may also exists in grass carp. The potential protein docking sites/residues were further assessed and calculated.

Several studies on fish have researched on the phylogeny of C5 (Graham *et al.* 2009; Johansen *et al.* 2019). However, a majority of such studies have limited species to display the C3/C4/C5 phylogenetic tree. Therefore, the present study included more species and specifically focused on the phylogeny of C5 in fish and non-fish vertebrate species. The results indicated that the *C5* gene underwent full divergence evolution, which is congruent with the findings of previous studies (Graham *et al.* 2009; Franchini *et al.* 2001). Notably, xenopus clustered with the teleost group in the phylogenetic tree. The genome codon usage bias between xenopus and other species for the first time was analyzed to verify the phylogenetic results. It was established that the codon usage of xenopus is closer to the teleost–1 cluster compared with others, which supported the phylogeny results. This suggests that the genomic codon usage bias could have potentially affected the xenopus C5 evolution. Similar studies have previously been performed in catfish (*Plotosus lineatus*) (Barathkumar *et al.* 2014). Reports indicated that the complexity in the course of the toxin evolution of catfishes in the par view of the neutral theory of natural selection could be efficiently interpreted through codon usage analysis. In addition, selection favors specific codons that promote the efficient and accurate translation of highly expressed genes. Therefore, the presence or absence of optimal codons can indicate whether selection for translational efficiency and accuracy plays a major role in the genomic evolution of an organism (Yi *et al.* 2018). However, the evolutionarily related codon usage bias analyses for fish genes are limited. The codon usage bias between grass carp *C5* gene and genomes of other species were calculated and analyzed to illustrate the evolutionary relationships between grass carp and other species. The results confirmed that the codon usage bias of the grass carp *C5* gene is closer to the genome of grass carp and zebrafish than in humans and stickleback. This confirmed their evolutionary relationship and suggested that the optimal heterogeneous expression environments of the *C5* gene are zebrafish and grass carp. However, the human cell line may not be an ideal vehicle for the heterogeneous expression of the grass carp *C5* gene. The mean difference of codon usage bias between the grass carp *C5* gene and the genome of zebrafish was lower than that between the grass carp *C5* gene and the genome of grass carp. We speculated that this result may be caused by different genome sequencing integrities. Additionally, this study reported distinct genomic codon usage biases in different species. Notably, the codon usage biases of different species are species-specific, possibly affected by transfer RNA abundance, gene copy numbers, and gene expression levels (Lavner & Kotlar 2005; Urrutia & Hurst 2003; Ikemura 1985). A previous study also reported that the appearance of translational selection can influence synonymous codon usage in fishes (Romero *et al.* 2003).

Multiple sequence alignment is a useful tool for studying closely related genes or proteins to determine the evolutionary relationships between the genes and the shared patterns among functionally or structurally related genes (Yu *et al.* 2006). The multiple sequences alignment in this study reported low sequence conservation in selected species. However, the sequence similarity between (human and mouse) and (zebrafish and grass carp) were significantly higher than other species. This confirmed the phylogeny results and codon usage bias. Besides, the signal peptide was predicted in most species following the previous fish study (Graham *et al.* 2009). Putative b–a cut sites were found in the species except for xenopus and stickleback, suggesting that the structures of the C5 amino acid sequences in these species, which were selected for sequence alignment, may be slightly distinct or are incomplete sequences yielded by the sequencing technology.

Since previous studies reported several differences between fish and non-fish vertebrate species, we explored whether these changes arise from gene selective pressure. A previous evolutionary study on fish complement genes analyzed and reported different selective pressure in the lineages of fishes and mammals (Wang *et al.* 2013c). Therefore, this study performed selective pressure analysis according to species taxonomy. The results showed that the selective pressure in the fish group is different from the non-fish vertebrate group. This indicates that fishes and non-fish vertebrate *C5* genes exhibit different evolution patterns. In addition, the aquatic environment contains countless groups of bacteria and viruses compared with the terrestrial environment. The neutral evolution of the fish *C5* gene possibly arises because of not being affected by the selective pressure during the evolutionary process through a specific defense strategy against the pathogen. An initial study reported that three periods of positive selection events had happened on *C3* genes during the evolutionary history and *C3* genes experienced different evolutionary patterns on fishes and mammals based on their living environments (Meng *et al.* 2012). Elsewhere, a similar study analyzed the evolutionary process in the *Myostatin* (*MSTN*) gene of teleost species, including Atlantic salmon (*Salmo salar*) and other vertebrates. They reported at least two separate gene duplication events in the fish lineage using a combination of methodologies between phylogenetic and selective pressure analyses (Ostbye *et al.* 2007).

Gene co–expression is widely used for gene functional annotation, and pathway analysis. The gene regulatory network was reconstructed by the theory of gene co–expression based on the expression data (Wang *et al.* 2014b). Gene co–expression network is a potent and promising approach to gather biologically relevant information, e.g., for the identification of genes not yet

associated with explicit biological questions, and accelerating the interpretation of molecular mechanisms at the root of significant biological processes (Tieri *et al.* 2018). In the present study, to explore the biological function of C5, we built the *C5* gene co–expression and PPI networks for human and zebrafish. We found that the number of co–expressed genes and interacted proteins in the human networks were more than in zebrafish. In addition, more GO biological processes were enriched in human than zebrafish. These results implied the C5 plays a vital role in the regulation of physiological and immune processes. Our results also revealed that the functional complexity of C5 may be distinct in species from different evolutionary positions. A similar study used the approach of gene co–expression to investigate human brain evolution by examining the large–scale organization of gene co–expression networks in human and chimpanzee brains (Oldham *et al.* 2006). They found gene co–expression relationships were poorly conserved in the cerebral cortex relative to the subcortical brain regions, revealing a striking gradient that parallels known evolutionary hierarchies. Combining the above results, building gene co–expression network could help identify key drivers of evolutionary changes and co–expression network conservation in species is indicative of the central functional importance of the hub gene (Langfelder *et al.* 2012; Langfelder *et al.* 2011).

Furthermore, protein structure prediction is a central setback in molecular biology. The structure of a protein provides several features that can be used to determine the function of the protein (Langlois & Lu 2008). The structure of a thermostabilized C5a receptor complexed with extra-helical antagonist NDT9513727 was identified in 2018 (Robertson *et al.* 2018). This work also reported one key interaction between the small molecule and residue Trp213$^{5.49}$ that potentially determined the species selectivity of this compound. The study provided new insights into the interaction between micro–molecule allosteric antagonists and the receptor. Besides, it provided a potential breakthrough in disease treatment based on complement related targeted medicine. In the present study, to improve the comparison of C5 genes and proteins in different species, we also included the zebrafish C5 protein structure. The significant difference between humans and two fish species constitutes the deduced protein structure. However, the C5a structures were similar, suggesting close similarity on the putative functions of C5a through evolution. In order to identify the authenticity of the CD88 sequences, we built the phylogenetic tree. We found the grass carp putative CD88 sequences were clustered in fish group with common carp and zebrafish. These sequences were clustered into two main branches according to the taxa, indicating the the authenticity of the collected CD88 sequences. Further, simulation of each top 10 C5a–CD88 protein complex conformations for human and grass carp suggested

that a theoretical interaction might exist between C5a and CD88 proteins. Following findings of the previous studies, in the human C5a protein structure, Loop 1 (residues C5a 12–20), loop 3 (C5a 39–46), and the C-terminal 6–8 residues (especially Arg74) are important for C5a receptor (C5AR) binding (Monk *et al.* 2007). Other studies on human C5AR have reported that Asp282, at the extracellular face of helix 7, interacts with the Arg74 side chain of C5a and the C-terminal Arg in peptide analogs (Cain *et al.* 2001; Cain *et al.* 2003). The potential docking/interacting sites were calculated and acquired based on the minimum distance from all atoms between C5a and CD88 proteins. Based on the previous study, atom pairs with more than 1.5 nm separation were not considered in the calculation of potentials (Shirota *et al.* 2009). The minimum distance from the atoms were considered as a potential protein interaction site for molecular function study. The 290Ala residue from the simulated CD88 protein in C1 and C7 models were calculated to dock to the 28Thr, 72Val from C5a protein at the minimum distance of 1.101 nm and 1.291 nm, respectively. The simulation and calculation of potential docking/interacting residues will improve and provide the mechanism of study on complement regulation.

In conclusion, this study explored the evolutionary process of C5 genes in different organisms. A gene co–expression network was constructed which revealed the systematic function of C5 during species evolution. The structural simulation of the C5 and C5a–CD88 protein complex provides insights into the putative functions of complement proteins in the immune system.

CHAPTER 4

Dynamic changes of the proteome in three Mediterranean fishes during early development stages

Lisen Li[1#], Liliana Anjos[1#*], Soraia Santos[1], Sara Castanho[2], Florbela Soares[2], Cátia Santa[3,4], Bruno Manadas[3,4], Deborah M Power[1, 5*]

 # These authors contributed equally to the study.

[1]*Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal*

[2]*IPMA—Portuguese Institute for the Sea and Atmosphere/EPPO—Aquaculture Research Station, Av. 5 de Outubro, 8700-305 Olhão, Portugal*

[3]*CNC - Center for Neuroscience and Cell Biology, Universidade de Coimbra, 3004-517 Coimbra, Portugal*

[4]*Institute for Interdisciplinary Research (IIIUC), Universidade de Coimbra, 3004-517 Coimbra, Portugal*

[5]*Shanghai Ocean University International Center for Marine Studies, Shanghai, China*

Manuscript submitted for publication

CRediT statement; LL- Investigation and experimental methodology, data analysis, writing - original draft, DMP – conceptualization, supervision, data analysis, writing – drafting, review & editing, LA - supervision, proteomics data analysis, writing – drafting, review & editing, SS – methodology & data analysis, writing - review, SC – animal maintenance and sampling, writing - review, FS - animal maintenance and sampling, writing – review, CS – proteomics, data analysis, writing - review & editing, BM - proteomics, data analysis, writing - review & editing.

Acknowledgements

This work was supported by funding from the European Union Horizon2020 Programme (PERFORMfish, grant nº 727610), from the Portuguese national funds from FCT - Foundation for Science and Technology through project UID/Multi/04326/2020, UIDP/04326/2020 and LA/P/0101/2020 and also from the operational programmes CRESC Algarve 2020 and COMPETE 2020 through projects EMBRC.PT ALG-01-0145-FEDER-022121 and BIODATA.PT ALG-01-0145-FEDER-022231. Liliana Anjos was funded by FCT, under the "Norma Transitória"-DL57/2016/CP1361/ project CT0011. The authors are especially grateful to Dr. Pedro Pousão from IPMA (Instituto Português do Mar e da Atmosfera, Olhão Portugal)) for kindly providing the egg and embryo samples, to Dr. Rute Felix from CCMAR for her support in QPCR analysis and to Dr Cymon J. Cox from CCMAR for his contribution in bioinformatic searches to convert homolog sequence.

4.1. Abstract

Comparative studies of early development can reflect the changes in the spatio-temporal use of regulatory molecules, as well as changes in the specificity of regulatory proteins and are correlated with important differences in morphology between phylogenetically distant species. The three Mediterranean fishes: gilthead sea bream (*Sparus aurata*; SA), white sea bream (*Diplodus sargus*; DS), and meagre (*Argyrosomus regius*; AR) are phylogenetically proximate. However, the change of functional molecules during embryogenesis and post embryonic development remains poorly known. In the current study, the proteome profiles in different embryonic developmental stages (before hatch-BH and hatch-H) of the three fish species (SA, DS, AR) was analysed by quantitative SWATH-MS (Sequential Window data independent Acquisition of the Total High-resolution-Mass Spectra) proteomics technology (data available via ProteomeXchange with identifier PXD033417). A relatively high conservation of the total identified and quantified proteins in different developmental stages was found for the three species. Differentially expressed proteins (DEP) specific to each species were acquired to provide a potential reference marker for species identification during embryo development. Gene Ontology and pathway analysis showed very few immune-related processes but the response to estrogen and cellular response to estrogen stimulus were found to be significantly enriched processes, suggesting the development and regulation of the endocrine system occurs in fish embryos. Of note, was the presence of hatching enzyme (He) a crucial protein for driving hatching, and significantly decreased protein levels were found from BH to H ($p < 0.05$). However, the relative expression of the three *he* genes (p value < 0.001) at the two stages indicated they have a distinct expression pattern, suggesting functional divergence of *he* isoforms during SA embryonic development. This enzyme may be useful as a potential marker of embryo developmental stage due to its distinct expression pattern between BH and H. The present study should unlock fundamental understanding of Mediterranean fish ontogeny contributing to improve aquaculture management practices linked to species specific characteristics.

Keywords: Biomarker; Embryogenesis; Fish; Quantitative proteomics

4.2. Introduction

Early development of oviparous organisms, such as fish, reptiles, birds and monotremes, that lay eggs that mature and hatch, is dependent on maternally derived structural, nutritional, and regulatory molecules. The maternal factors that accumulate in eggs influence their quality and this determines the future developmental potential of the embryo and this is a characteristic that is of high interest for fish aquaculture (Bizuayehu *et al.* 2019). Despite the evolutionary distance between vertebrates, including fish, embryo development shares the same general characteristics (Arezo *et al.* 2005), and gene expression profiles have conserved patterning in embryos of *M. musculus*, *G. gallus*, *X. laevis*, and *D. rerio* (Irie & Kuratani 2011).

Embryogenesis in teleosts has been extensively studied and presumably contributed to the success of the teleosts of which over 28,000 extant species exist (Cucchi *et al.* 2012; Honji *et al.* 2012; Nelson *et al.* 2019). Since most fish species are oviparous, this means egg and embryo development is directly influenced by the environment and despite the almost impermeable outer chorion, temperature, oxygen availability, salinity, pH etc. all affect embryogenesis, although not always in the same way (Jonsson & Jonsson 2019; Speers-Roesch *et al.* 2012; Di Santo *et al.* 2016; Vetemaa & Saat 1996). For example, in the maraena whitefish (*Coregonus maraena*) temperature alters *igf* gene (*igf1* and *igf2*) expression and affects embryonic development (Nipkow *et al.* 2018). In embryos of the Chinese black sleeper (*Bostrichthys sinensis*) and mudskipper (*Boleophthalmus pectinirostris*) the perivitelline fluid plays an important role in tolerance to hypoxia and desiccation (Shixi *et al.* 2006). Salinity influences embryo viability and hatching rate of Nile tilapia (*Oreochromis niloticus*) and this is proposed to be due to the increased permeability of the chorion during gastrulation, which may result in osmoregulatory strain since the chloride cells are still not fully functional (Fridman *et al.* 2012). The effect of pH on fish species is variable and in the tetra (*Astyanax lacustris*) it has a negligible effect, but in the small scaled pacu (*Piaractus mesopotamicus*), it affected the hatching rate and survival (Santos *et al.* 2020).

Fish larvae are exposed to profound environmental changes when they break out of the virtually impermeable chorion at hatching (Korwin-Kossakowski 2012). Nearly all environmental factors affect hatching, including water temperature (Nipkow *et al.* 2018; Kaminski *et al.* 2006), oxygen content (Ciuhandu *et al.* 2005), light (Bruning *et al.* 2011) and chemicals such as ammonia (Wang *et al.* 2012). All these factors may directly influence embryogenesis and therefore embryo quality and in this way are also tightly associated with the quality of post embryonic (larval) stages. Oocyte-derived compounds and proteins contribute to the success of fertilization and embryogenesis (Lubzens *et al.* 2017). Vitellogenin (Vtg), an

abundant lipoprotein in fish eggs, is crucial for oocyte maturation and contributes to embryogenesis and larval development (Arukwe & Goksøyr 2003). In addition to abundant yolk proteins such as Vtg, fish eggs also contain other factors that influence their quality such as maternal mRNAs, non-yolk proteins, vitamins and hormones (Bobe & Labbe 2010).

Maternally inherited mRNAs and proteins accumulate and support early embryonic development until activation of zygotic transcription and thus play a pivotal role during oogenesis and early embryogenesis (Tata 1986; Howley & Ho 2000; Pelegri 2003; Bobe & Labbe 2010; Moravec *et al.* 2021). The importance of maternally inherited molecules is demonstrated when they are depleted as they result in embryonic lethality (Moravec *et al.* 2021). In rainbow trout (*Oncorhynchus mykiss*), maternal mRNA translation delays embryonic cleavage (Nagler 2000) and in zebrafish (*Danio rerio*), their spatial distribution within the oocyte specifies the dorso/ventral axis of the embryo (Bally-Cuif *et al.* 1998; Howley & Ho 2000). Maternal sex steroids, cortisol, and other lipophilic hormones like thyroxin contribute to teleost egg and embryo development although their exact function is still not well established (Feist *et al.* 1990; Mommsen *et al.* 1999; Hwang *et al.* 1992; Arukwe & Goksøyr 2003).

The composition of eggs and embryos from a range of fish species has been reported. For example *Salmonidae* eggs contain 63.45% - 63.95% moisture, a total protein content of 23.83% - 24.93%, a crude fat content of 4.26% - 4.58%, and 1.57% -1.87% crude ash and *Acipenseridae* eggs contain 64.03% - 64.10% moisture, 19.82% – 19.94% total protein, 10.34% – 10.46% crude fat and 1.94% – 2.18% crude ash (Kowalska-Goralska *et al.* 2020). In the tilapia (*Oreochromis niloticus*) and catfish (*Siluriformes* sp.) the eggs contain 60% - 70% moisture and approximately 20% - 30% and 30.56% crude proteins, respectively (Oko 2019). A higher crude fat (0.53% – 10%) content occurs in tilapia eggs compared to catfish eggs (0.53%). Again, a higher ash content (1.89% - 10%) occurs in tilapia eggs compared to catfish eggs (1.89%), suggesting a general similarity in egg nutrient composition of different fish species. A similar global composition even occurs in non “white fish” species, and Intarasirisawat *et al*. (2011) reported that the roe of three tuna species contained 72.2%. – 73.0% moisture, 18.2% – 20.2% protein, 3.4% – 5.7% lipid and 1.8% – 2.1% ash (Intarasirisawat *et al.* 2011), similar results were also found in skipjack tuna (*Katsuwonus pelamis*) (Yoon *et al.* 2018). Intriguingly, there appears to be a much lower crude fat content in catfish eggs compared to other fish eggs but a relatively higher crude protein content, suggesting the importance of crude protein for catfish early development. Further studies will be required to confirm and explain the basis of the divergent nutrient contents of catfish eggs and embryos. Apart from moisture, proteins are an important molecular cargo of fish eggs due to their functional role in the embryo and are

important as a source of cellular energy and for the formation of fish embryos and larvae (Lubzens *et al.* 2017). Several protein families have been identified in addition to Vtgs and Zona pellucida proteins (ZPs) in the fish egg. These protein families include proteins associated with metabolic functions, chaperones, peroxiredoxins and translation regulatory proteins (Lubzens *et al.* 2010). A number of studies have assessed the role of specifically derived maternal proteins contributing to immune protection of the egg and embryos, such as complement 3 (C3) protein in Atlantic salmon (*Salmo salar*) (Lovoll *et al.* 2007) and phosvitin (Pv) in zebrafish embryos (Wang *et al.* 2011). The functional importance of proteins allied with new high throughput methods has led to an increased number of studies of adult tissues in model teleost species, such as zebrafish (De Souza *et al.* 2009; Groh *et al.* 2011; Singh *et al.* 2010; Zhang *et al.* 2010) and a few studies in aquaculture species such as cod (*Gadus morhua*) (Sveinsdottir *et al.* 2008), European sea bass (*Dicentrarchus labrax*) (Crespel *et al.* 2008), and rainbow trout (Niksirat *et al.* 2020).

In the present study, three Eupercaria fish, the gilthead sea bream (*Sparus aurata*; SA), white sea bream (*Diplodus sargus*; DS) and meagre (*Argyrosomus regius*; AR) were studied. The three species studied include an important Mediterranean aquaculture species, SA, and two emerging aquaculture species of this region. Despite the significant rise in total production of all the species from 87.30 K tons in 2000 to 185.98 K tons in 2016 due to advances in husbandry and management practices, defining and controlling embryonic quality remains a challenge. Previous research that has reconstructed the phylogeny of Mediterranean teleost species reveals a close phylogenetic relationship exists between the three species (Meynard *et al.* 2012). As a first step to better understanding the role of egg proteins in embryogenesis and the early development of the target species the core proteome of their embryos was obtained and a comparative analysis between species was carried out as well as within species at two different embryonic stages. Proteomic technology was applied to develop accessible and usable resources of scientific utility for these three Mediterranean teleost fish species.

4.3. Materials and methods

4.3.1. Experimental fish and embryonic sample collection

All the sampling procedures were carried out as part of the routine hatchery management and were in compliance with 2009/58/EC legislation (protection of animals kept for farming) and under GLOBAL G.A.P. (Good Agricultural Practice, including good animal welfare practices) certification. Fertilized eggs (embryos) from gilthead sea bream (*Sparus aurata*; SA), white sea bream (*Diplodus sargus*; DS) and meagre (*Argyrosomus regius*; AR) (henceforward

designated as: SA, DS and AR respectively) were obtained from broodstock maintained in the Aquaculture Research Station of the Portuguese Institute for the Ocean and Atmosphere (EPPO-IPMA, Olhão, Portugal). AR fertilized eggs were collected from IPMA's F2 generation broodstock, which had an average weight of 2.5 ± 0.59 kg and were kept at a density of 5.9 kg/m^3 (17 breeders). DS fertilized eggs were collected from a group of breeders with an average weight of 1.0 ± 0.41 kg kept at a density of 2.5 kg/m^3, and SA fertilized eggs were collected from a group of breeders with an average weight of 2.0 ± 0.37 kg kept at a density of 3.5 kg/m^3. Both DS and SA breeders were from wild broodstock collected from the Portuguese Atlantic coast (26 and 18 breeders respectively).

Spawned eggs and embryos were harvested into collectors placed under the outlet pipe positioned to remove water from the surface of the broodstock tanks. Floating viable embryos in the collectors were used to stock one independent 200 L incubator for each fish species (SA, DS and AR), and embryos were maintained in incubators until larval hatching (approximately two days after collection). Incubator tanks containing the embryos were operated as a flow-through system and received sea water at 19.5 ± 0.87 °C with a salinity of 36 ± 1 ppt and oxygen at 6.4 ± 0.55 mg/L. The conditions of the sea water in the incubator tanks were recorded daily with handheld monitors. Triplicate samples composed of independent pools of embryos (n = 3, approximately 80-120 mg of embryos) were collected from the egg incubator for each fish species at two developmental time points: 1 day before hatch (BH) and at hatch (H) as indicated in Figure 4.1 and Supplementary Figure 4.1 in annex I. In total, 18 pooled embryonic samples were collected for the three species at two developmental stages and were snap frozen in liquid nitrogen in 1.5 ml microcentrifuge tubes and stored at – 80 ºC until analysis.

An additional experiment was performed using fertilized eggs from commercial sea bream and sea bass hatcheries with the objective of establishing comparisons between the expression profile of transcripts of some of the identified proteins from the proteome study. Embryonic samples of SA were obtained from several different commercial hatcheries 1 day before hatch (BH, n = 6) and at hatch (n = 18). For European sea bass (*Dicentrarchus labrax*; DL) embryo samples were collected from one commercial hatchery from different production batches at hatch (n = 16). Sampled embryos were preserved in methanol and stored at -20 ºC until total RNA extraction and cDNA synthesis for subsequent Quantitative PCR analysis.

4.3.2. Protein extraction from emrbyoes and protein profile analysis

Total protein was extracted from the frozen embryos (n = 3 pooled embryonic samples/developmental stage/fish species, making a total of 18 protein extracts), as shown in

Figure 4.1. Fertilized eggs (80 - 120 mg/sample) were homogenized in ice-cold protein extraction buffer (PEB - 1.7% SDS, 50 mM Tris HCl pH 6.8, 100 mM DTT, 5% (v/v) protease inhibitor cocktail (Sigma, US)) in a proportion of 500 µl of PEB/100 mg of embryos) on ice using a 1.0 ml manual glass homogenizer (Jencons, UK). Homogenates were kept on ice for 30 min and vigorously mixed at 10-minute intervals. To denature the protein, the homogenates were boiled at 95 ºC for 10 min in a dry bath, vigorously mixed at 5 min intervals and then allowed to cool at room temperature. Soluble protein fractions were separated by centrifugation at 16900 × g for 20 min at 4 ºC, alkylated with 40% acrylamide solution (1:15 of acrylamide: protein solution v/v) and stored at - 80 ºC.

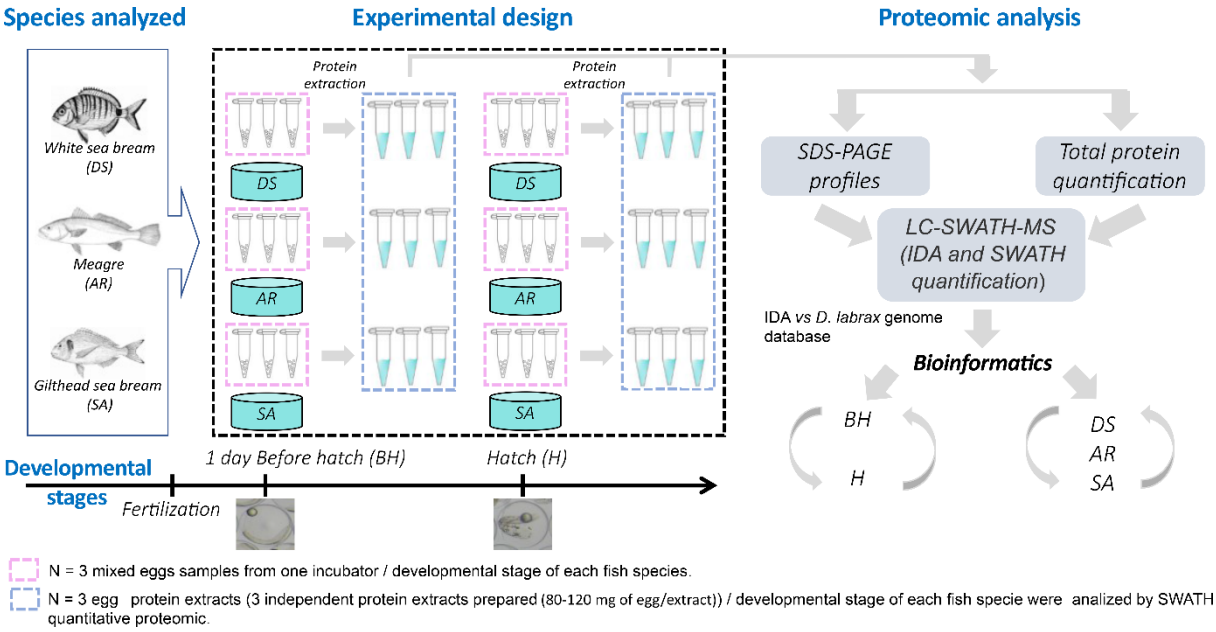


Figure 4.1. Overall workflow of the experimental design and the SWATH quantitative proteomic analysis in the current study. Pools of embryos at two developmental stages (BH and H) from three Mediterranean fish species [gilthead sea bream (*Sparus aurata*; SA), white sea bream (*Diplodus sargus*; DS), meagre (*Argyrosomus regius*; AR)] were used for proteomic analysis.

The protein content of the embryo protein extracts was determined in duplicate using a Bradford assay with bovine serum albumin as the standard (Bio-Rad, US) in a 96 well microtiter plate following the procedure of the manufacturer. Absorbance was measured in a spectrophotometer at 595 nm (Techan Infinite® M200 Microplate Reader, Switzerland). The profile and integrity of extracted soluble proteins was assessed by one dimension 12% polyacrylamide gel electrophoresis (SDS - PAGE) using the Laemmli method (Laemmli 1970) followed by Coomassie blue staining. Samples were prepared by mixing 7.2 µg of total soluble embryo proteins with 1x SDS gel loading buffer (100 Mm Tris buffer pH 6.8, 200 mM DTT,

4% SDS solution, 0.2% bromophenol blue, 20% Glycerol) and after thermal denaturing (95 ºC and 5 min) the mixture was fractionated by SDS-PAGE under a constant current (50 mA). The gels were stained with Coomassie blue R solution (50% v/v methanol, 10% v/v glacial acetic acid and 0.1% w/v Coomassie Brilliant Blue R-250, Sigma, US) at room temperature for 1 hour and destained for 12 hours in a solution containing 40% v/v methanol and 10% v/v glacial acetic acid.

4.3.3. Quantitative proteomics of embryos

The embryonic proteome of the different developmental stages (H, BH) and different fish species under analysis (SA, DS, AR) were obtained using label-free quantitative proteomics technology SWATH-MS (sequential windowed acquisition of all theoretical fragments ion mass spectra), following the procedure reported in Anjos et al. (2022) (Anjos *et al.* 2022). In brief, triplicate samples (n = 3) constituted of 3 independent pools of embryos from which protein was extracted/developmental stage/fish species were used for proteomics analysis. For short GeLC-SWATH-MS analysis 50 µg of each of the triplicate samples or a pool (45 µg) of the same three samples were fractionated as previously reported (Anjo *et al.* 2015). Gel regions containing the extracted protein were excised (4 fractions) and processed (destaining, in-gel protein digestion with trypsin, peptide extraction and solid-phase extraction with C18 sorbent (OMIX tip, Agilent Technologies)) as reported in (Santa *et al.* 2016).

LC-SWATH-MS was carried out using an Eksigent® NanoLC-Ultra 2D coupled to a TripleTOF^{TM} 6600 system (Sciex®, USA) using two operative phases (Figure 4.1): 1) Information-dependent acquisition (IDA) of the pooled samples. This data was used to generate a reference spectral library, named, "*Library_Pools H and BH of AR_DS_SA*" by combining all files from the IDA experiment; and 2) SWATH-MS acquisition for the relative quantification and identification of the proteins in the individual samples (3 biological replicates/developmental stage/fish species). The SWATH setup was as reported in Gillet et al (Gillet *et al.* 2012), and the same chromatographic conditions were used for SWATH and IDA acquisitions. Protein identification and library construction was performed using ProteinPilot™ software (v5.0.1, Sciex) with the following search parameters: 1) predicted proteins from the sea bass genome database (June 2012 draft assembly dicLab v1.0c with annotation from July 2013; file diclab1_pep.faa.gz downloaded from http://seabass.mpipz.mpg.de, (Tine *et al.* 2014)); 2) acrylamide alkylated cysteines as fixed modification; 3) the gel based special focus option.

Relative quantification was performed using the SWATH^{TM} processing plug-in for

PeakView™ (v2.2, Sciex). Identification of proteins was accepted when the confidence level was 95% local-FDR (Sennels *et al.* 2009) and protein quantification was only accepted when at least one peptide reached an FDR confidence level of 99% and complied with the criteria described in Lambert et al (Lambert *et al.* 2013). That is, (i) the peptides should be common to all three DDA (data-dependent acquisition) runs; (ii) the peptides should have been identified with ID confidence ≥ 99% in at least one of all runs; (iii) all peptides, including modified peptides, are considered; (iv) each charge state is considered as a separate instance for quantification; (v) instances of the same peptide (with same charge) are considered the same peptides if they are within ±5 min of the retention time window. Protein levels were estimated by summing all the transitions from all the peptides for a given protein (Collins *et al.* 2013) and normalized to the total intensity at the protein level. The SWATH processing quantification procedure was performed in two ways: a) separately for each fish species in order to get the maximum number of proteins confidently quantified in each species, allowing comparisons between embryonic stages within each species; and b) for all the samples from the 3 species at once to allow the comparisons to be performed directly between species within the same stage. The library generated by IDA was used for all the SWATH processing quantification procedures, by importing the proteins and peptides of confidently quantified proteins. Detailed methods and rigorous quality criteria used for protein identification and quantification are described in Supplementary File 4.1 annex I. The resulting mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE (www.ebi.ac.uk/pride, (Perez-Riverol *et al.* 2022)) partner repository with the dataset identifier PXD033417. Fold-change in protein abundance was calculated between the two developmental stages (BH and H) in each fish species (SA, DS, AR) or between the different fish species at the same developmental stage, using the medians between relative protein abundance in the experimental groups.

4.3.4. Bioinformatic analysis

Zebrafish (*Danio rerio*) orthologues for all the DE proteins were obtained using stand-alone Blastp (with E value < $10e^{-4}$) against the Ensembl zebrafish protein predictions database (GRC Zebrafish Build 10, INSDC Assembly GCA_000002035.3 assessed at https://www.ensembl.org in July 2019, (Curwen *et al.* 2004)). Protein lists that changed with the experimental conditions were analysed using the eulerr website (http://eulerr.co/, (Larsson 2020)). Venn diagrams were plotted using jvenn an open access platform for data analysis and visualization http://www.bioinformatics.com.cn (Bardou *et al.* 2014). Hierarchical clustering of

proteome profiles was carried out using the open access platform for data analysis and visualization (http://www.bioinformatics.com.cn). Principal component analysis (PCA) was also performed using ClustVis software (https://biit.cs.ut.ee/clustvis/ (Metsalu & Vilo 2015)) to analyze sample dispersion. In the data processing a 95% confidence interval (CI) was chosen to group different samples and illustrated using different colours in the elliptical area. Functional analysis of gene ontology (GO)-biological process (BP) and KEGG (Kyoto Encyclopedia of Genes and Genome) pathway enrichment were run for proteins significantly modified using the Metascape plug-in (http://metascape.org/gp/index.html#/main/step1 (Zhou *et al.* 2019b)) and using as the input zebrafish orthologues, after removing duplicate zebrafish accession numbers. Enrichment analyses (well-adopted hypergeometric test (Zar 1999) and a Benjamini-Hochberg p-value correction algorithm (Hochberg & Benjamini 1990)) were run selecting the GO-BP terms for zebrafish. Terms were considered significantly enriched at a p-value ≤ 0.01 and when a minimum of 3 proteins of the GO-BP proteins were represented on the list. The same parameters and strategy as described for GO-BP analyses were used for the KEGG pathways enrichment analyses.

4.3.5. Quantitative (Q) PCR

QPCR was used to complement the proteomics analysis and gene transcripts were targeted that code for proteins involved in development. Hatching enzymes (He), which are crucial proteins secreted from the hatching gland cells, which disrupt the egg envelope and promote hatching were targeted for analysis (Inohaya *et al.* 1995; Inohaya *et al.* 1999). The abundance of He proteins during embryonic development was compared with change in gene transcript abundance using QPCR. The experimental samples utilized were pools of embryos from SA (BH and H) and DL (H) obtained from commercial hatcheries across the south of Europe (see 4.3.1). Total RNA (tRNA) was extracted from embryos as previously described (Li *et al.* 2021). Briefly, 15 individual embryos were, weighed and homogenized in 700 µl of lysis buffer from an E.Z.N.A total RNA extraction kit (VWR, USA) containing 2-Mercaptoethanol (20% v/v). Homogenization of the embryos was performed by mechanical disruption with two iron beads (5 mm) using a Tissue lyser II Qiagen (Scansci, Moinho de Bolas MM 400) with 1 cycle of 30 s at room temperature and tRNA was recovered following the manufacturer's instructions.

For cDNA synthesis, 500 ng of DNAse treated tRNA was used and reactions were performed with RevertAid-RT (Thermo Fisher, USA) for a 20 µL final volume with 100 pmol random hexamers (Jena Bioscience, Germany). The quality of cDNA was assessed by amplification of ribosomal subunit 18s (Supplementary Table 4.1 annex I) using the following

PCR conditions: 95 °C, 3 min; 27 cycles x (95 °C, 10 s; 60 °C, 10 s; 72 °C, 10 s); 72 °C, 5 min.

Primer pairs were designed for gilthead sea bream and DL *he* genes (*SA-he1*, *SA-he2*, *SA-he3* and *DL-he1* and *DL-he2*, Supplementary Table 4.1 annex I). Quantitative PCR of samples was performed in 96-well low volume microplates in duplicate reactions using SsoFast EvaGreen supermix (Bio-Rad, Portugal). The QPCR reaction contained Evagreen (5 µl /reaction), the forward (0.3 µl) and reverse (0.3 µl) primers, template cDNA (2 µl; diluted 1:10) and sterile water to give a final volume of 10 µl. Thermocycling was performed with a BioRad CFX Connect Real Time System using optimized cycling conditions consisting of 95 °C for 30 s followed by 39 cycles of 95 °C for 5 s. Melting curves were performed for all reactions to detect non-specific products and primer dimers at 56-60 °C for 10 s. A final melting curve was generated by increasing the temperature from 60-95 °C to confirm a single reaction product was amplified. *18 ribosomal subunit* (*18s*), which did not vary across samples was selected as a reference gene for sea bass (Pinto *et al.* 2022). For sea bream the geometric mean of farnesyltransferase, *CAAX box,* and *alpha* (*fnta*), which did not vary between samples were used as the reference for normalization.

To compare He expression at the transcript and protein level, quantitative proteomic data from the SWATH analysis was extracted and He protein isoform, protein level and difference in expression between the two developmental stages of each species, SA, DS and AR, was determined. Gene expression profiles of *he* isoforms measured by QPCR in sea bream and DL samples was compared with the SWATH protein expression analysis.

4.3.6. Statistical Analysis

SPSS v23 (IBM) software was used to analyze the statistical significance of the relative protein abundance between different stages and species (n=3/experimental group). A non-parametric Mann Whitney U-test was used for the comparison between different developmental stages (BH, H) for each fish species, and an exact and asymptotic p-value is reported. Kruskal-Wallis followed by a post-hoc Dunn's test corrected by FDR and controlled using a Benjamini-Hochberg adjustment was used for multiple comparisons (between different fish species within the same stage). Differences between the relative protein concentration were considered significant at $p \leq 0.05$.

qPCR expression data are presented as the mean ± SEM of at least 6 samples composed of tRNA from 15 embryos for SA (BH and H) and DL (H). For DL significant differences in expression level between different genes were analyzed using an unpaired t test. For sea bream,

variance homogeneity was examined using Brown-Forsythe test and Bartlett's test and the normality was verified by the D'Agostino & Pearson test, Shapiro-Wilk test and Kolmogorov-Smirnov test. Significant differences in gene expression between developmental stages was analyzed using a two-way ANOVA followed by a Tukey's multiple comparisons test. All statistical analysis was performed using GraphPad Prism 8 version 8.0.2 (263) and a p-value < 0.001 was set as the significance cut-off.

4.4. Results

4.4.1. Comparison of the protein profiles between the developmental stage of the three fish species

One dimensional SDS-PAGE gel revealed the soluble protein expression profile in BH and H developmental stages of DS, AR and SA and revealed consistent high-quality protein extracts from embryo samples (Figure 4.2). Similar protein patterns for the same developmental stage between different species were found by visual inspection of Commassie blue stained SDS-PAGE gels. The protein profiles between the two developmental stages (BH and H) were different within the same fish species and for all the species under analysis (red and blue arrows, Figure 4.2). Putative stage specific candidate protein markers were observed and in general were present in the protein profile of all species (Figure 4.2, red arrows). Some exceptions were observed among the high MW proteins (≈ 250 kDa) in DS and the low MW proteins, ≈ 20 and < 15 kDa, in AR and SA, respectively (Figure 4.2, blue arrow), which may represent stage-specific protein biomarkers for species.

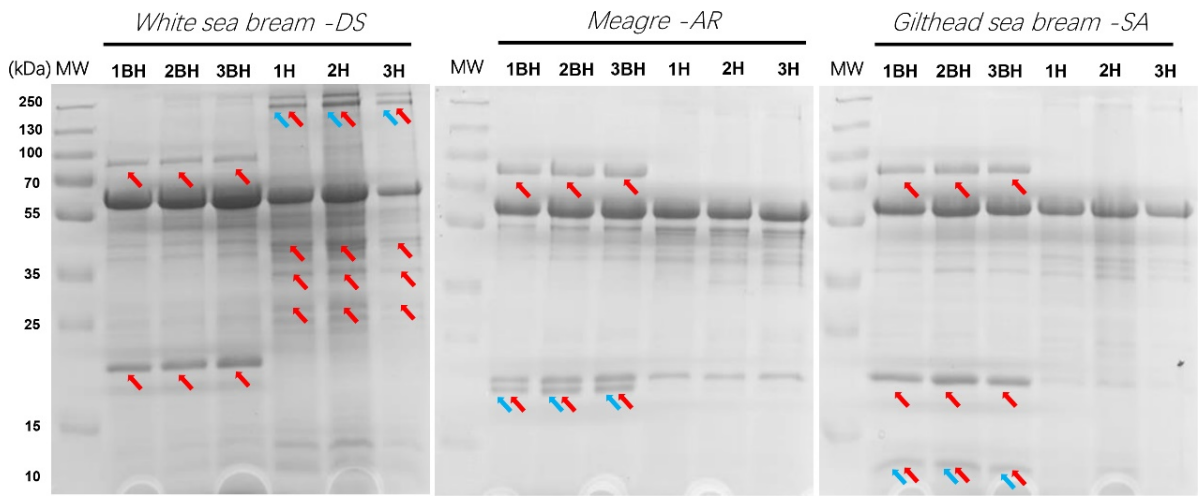


Figure 4.2. Total protein profiles of two developmental stages of embryos (BH and H) from the three

fish species under analysis (DS, AR, SA). Protein extracts (7.2 µg per lane) of 3 independent mixed embryo samples per developmental stage (BH and H) of the three fish species (DS, AR, SA) were resolved by 1D SDS-PAGE (12%) and stained with Coomassie blue. Note: arrows highlight differences in protein profiles detected by visual inspection: red – between two developmental stages within the same fish species; blue – between the 3 different fish species. MW - Molecular weight marker (kDa); BH - before hatch; H - hatch; DS - *Diplodus sargus* (white sea bream); AR – *Argyrosomus regius* (meagre); SA - *Sparus aurata.* (gilthead sea bream).

4.4.2. Proteome remodeling in the developing embryo (H and BH) of the three fish species SA, DS and AR

4.4.2.1. Comparative analysis of the global proteome (identification and quantification)

SWATH-MS based proteomics was used to generate a library of identified proteins in SA, DS and AR embryo samples (by combining pools of BH and H samples of all fish species). The protein identity in SWATH-MS data was established by comparison with the predicted proteins of the sea bass genome (Meynard *et al.* 2012; Tine *et al.* 2014). The library of identified proteins was based on the spectral library from IDA analysis of the 6 independent pools corresponding to each experimental condition ("*Library_Pools H and BH of AR_DS_SA* ", Supplementary Table 4.2 in annex II). A total of 2376 proteins supported by 13,119 peptides were confidently identified (FDR 95%) in SA, DS and AR (Table 4.1; Supplementary Table 4.2 annex II).

The SWATH quantification procedure for embryo protein extracts from individual fish species, enriched the total number of confidently quantified/identified proteins when compared with the quantification procedure obtained from the pooled results of the 3 fish species simultaneously. A total of 922, 1149 and 960 proteins were confidently quantified/identified in AR, DS and SA, respectively (Table 4.1; Supplementary Table 4.3-4.5/worksheet Proteins_ITotal annex II). The species-specific libraries generated were used to map the differentially expressed proteins (DEP) between stages within each species. The overlap between the global proteomes generated for each fish species (combining both embryonic stages) revealed greater than 60 % (694) of the proteins confidently identified/quantified (FDR < 0.05) were common, indicating their embryo proteomes were broadly similar (Figure 4.3A). The most similar embryo proteome profile was between SA and DS (180 proteins in common) and the least similar embryo proteome was between AR and SA (only 33 in common). In SA, DS and AR, 53, 188, and 108 proteins, respectively were specifically expressed in the embryo proteome, suggesting they may represent proteins associated with species specific characteristics. A further library of 622 proteins confidently quantified/identified was derived by combining simultaneously the spectral libraries of the 3 fish species (Table 4.1; Supplementary Table 4.6/worksheet Proteins_ITotal annex II) and this library was used to map the DEP's between species at the same stage.

Table 4.1 - Summary of the number of proteins and peptides in the SWATH proteomic analysis obtained from the 3 fish species embryo protein extracts at 2 developmental stages, at hatch and before hatch. The number of identified proteins/peptides is shown for the global identification library generated (IDA analysis), "Library_Pools H and BH of AR_DS_SA", prepared from pools of embryo protein extracts of AR, DS, and SA at Hatching-H and Before Hatching (BH). This library was generated by combining all the spectral files obtained from the 6 independent pools of each experimental condition and was used as the library for SWATH quantification. The number of identified/quantified proteins obtained by SWATH with confident quantitative values (local FDR < 0.05, p < 0.05) and the number of proteins whose levels changed between the indicated experimental groups (p ≤ 0.05) are presented, as well as the corresponding Tables where detailed results can be found. NA-Not applicable; Numbers within brackets (...), are the number of proteins with modified levels with a Fold change of 1.5X (0.67>ratio>1.5).

Identification (IDA)					
Library_experimental group	**Nº of Proteins**	**Nº of Peptides**	**Supported by ≥3 peptides (95%)**		**Table link**
Library_Pools H and BH of AR_DS_SA	2376	13119	1237		Supp Table 4.2_worksheet-Library_Protein Summary annex II
SWATH quantification					
Experimental group comparisons	**Nº of proteins**				**Table link**
	Identified/ Quantified	**Modified**	**UP**	**Down**	
Comparison between developmental stages within samefish species					
AR_BHvsH (H effects)	922	400	363 (159)	37 (15)	Supp Table 4.3_worksheet-ProteinsITotal_922 and 400_DE Prot annex II
DS_BHvsH (H effects)	1149	575	414 (197)	161 (49)	Supp Table 4.4_worksheet-ProteinsITotal_1149 and 575_DE Prot annex II
SA_BHvsH (H effects)	960	331	284 (148)	47 (13)	Supp Table 4.5_worksheet-ProteinsITotal_960 and 331_DE Prot annex II
Comparison between fish species within same developmental stage					
	622	246			Supp Table 4.6_worksheet-ProteinsITotal and 246_DE Prot annex II
BH group					
BH_ARvsDS	NA	144	136 (116)	8 (8)	Supp Table 4.6_worksheet-246_DE Prot annex II
BH_ARvsSA	NA	30	10 (10)	20 (16)	Supp Table 4.6_worksheet-246_DE Prot annex II
BH_DSvsSA	NA	55	3 (3)	52 (33)	Supp Table 4.6_worksheet-246_DE Prot annex II
H group					
H_ARvsDS	NA	88	76 (61)	12 (10)	Supp Table 4.6_worksheet-246_DE Prot annex II
H_ARvsSA	NA	39	6 (6)	33 (24)	Supp Table 4.6_worksheet-246_DE Prot annex II
H_DSvsSA	NA	18	3 (1)	15 (6)	Supp Table 4.6_worksheet-246_DE Prot annex II

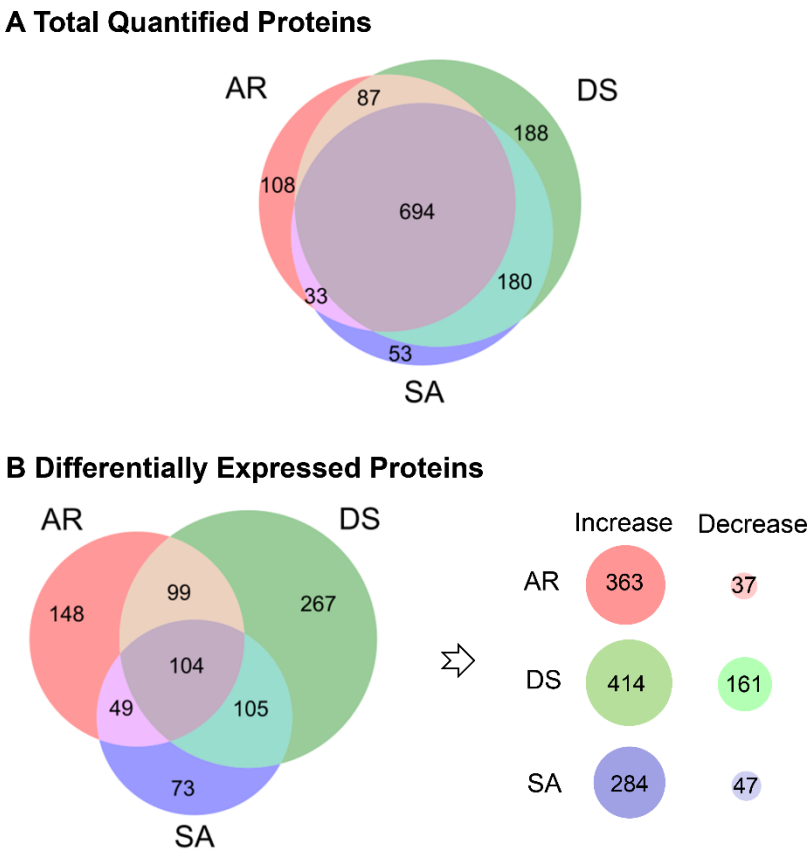


Figure 4.3. Numerical overview of the proteome changes between embryo developmental stage of the 3 fish species analysed. Venn diagrams showing: A) comparisons of the number of total confidently quantified proteins by SWATH processing among SA, DS and AR (FDR < 0.05); B) the number of overlapping differentially expressed proteins (DEPs) by comparison of the DEP libraries generated between embryo developmental stages (BH versus H) among SA, DS and AR ($p \leq 0.05$) and the number of DEPs that significantly increased or decreased at H stage (BH versus H) for the 3 fish species (the size of the circles is proportional to the number of proteins with modified expression in each experimental condition). Values inside the circles represent the number of proteins identified. DS - *Diplodus sargus* (white sea bream); AR – *Argyrosomus regius* (meagre); SA - *Sparus aurata* (gilthead sea bream).

4.4.2.2. Overview of the differentially expressed proteins (DEPs)

To understand how the proteome was modified between 2 developmental stages in the 3 fish species, 2 comparisons were established: I) between the embryo developmental stage in the same fish species and II) within the same developmental stage between fish species. The comparison of the proteome of embryonic stage (BH *vs* H) within the same species revealed that proteins with a significantly modified level ($p \leq 0.05$) was greatest in DS (575) compared to AR (400) and SA (331) (Table 4.1; Supplementary Table 4.3-4.5 annex II). The DEPs in DS, AS and SA, respectively corresponded to 50 %, 43 % and 35 % of all quantified DEPs in each species, indicating that a substantial proportion of the proteome is remodeled during the BH to H transition. Moreover, around 55 % of the total DEPs were filtered based on a minimum fold

change cut off of 1.5 (p-value ≤ 0.05, Table 4.1) compared to that only based on p-value ≤ 0.05. Of the identified DEPs in DS, 414 were up-regulated and 161 were down-regulated. In AR 363 proteins were up-regulated and 37 were down-regulated. In SA 284 proteins were up-regulated and 47 were down-regulated. Proteome remodeling in embryos at the H stage was predominantly associated with an increase in protein abundance, since a much smaller number of DEPs were identified that had decreased abundance (AR, 9.25 % and SA, 14 %) (Table 4.1, Figure 4.3B).

An overview of the DEPs between stages (BH and H) in the three fish species revealed that 104 proteins were commonly modified (Figure 4.3B) and corresponded to approximately 18% of the DEPs in DS, 26% in AR and 31% in SA. The number of differentially expressed proteins specific to each species was 73, 267 and 148 in SA, DS and AR, respectively. To assess the predominant biological processes reflected in the proteome of the 3 fish species at H, the 10 proteins (referred to as top 10 hereafter) with the biggest modification in expression were analyzed. The top 10 proteins with significantly increased levels in the H stage of DS included, myosin-related proteins (myosin light chain 3, myosin heavy fast skeletal muscle and three myosin heavy chain) and other proteins related to muscle development (tropomyosin alpha-4 chain isoform 1, tropomyosin alpha-3 chain isoform 1 and troponin skeletal muscle-like) (Supplementary Table 4.4 annex II). In SA myosin-related proteins were also predominant in the top 10 proteins (myosin-4-like, myosin heavy chain, myosin heavy fast skeletal muscle, myosin light chain 1, myosin heavy chain, Supplementary Table 4.5 annex II) indicating the importance of muscle development during the H stage for both DS and SA. In AR only one myosin (myosin heavy chain) was found in the top 10 most abundant proteins (Supplementary Table 4.3 annex II), suggesting the developmental status at hatch of AR most likely diverges from the other species, but this was not obviously reflected by morphological observation (Supplementary Figure 4.1 annex I).

The 10 proteins with the most significantly decreased levels at H in DS included hatching enzyme, vitellogenin aa and oogenesis-related proteins, and in SA was vitellogenin aa and oogenesis-related proteins. Zona pellucida sperm-binding protein 3-like, zona pellucida sperm-binding protein partial all related to the egg layers and vitellogenin aa and oogenesis-related protein were within the 10 most decreased proteins in AR. The DEPs between fish species at the same developmental stage was highest between AR and DS for BH (144) compared to the same stage in AR and SA (30) or DS and SA (55) (Table 4.1). A comparison of DEPs at H between AR and DS revealed 88 compared to 39 in AR vs SA and 18 in DS vs SA.

In addition, to the above analyses of the most modified proteins immune-related proteins were also procured in the dataset of common DEPs. Few immune-related proteins were identified and even fewer were DEPs. Of note was the immune-related protein, ELAV ((embryonic lethal-abnormal vision) - like protein 1a (elav1a)), which protects zebrafish embryos from bacterial infection was identified (Ni *et al.* 2021). A homologue of zebrafish ELAV was not identified in species common DEPs but was only significantly modified in DS from BH to H, indicating a possible vital function during DS embryo development. An important growth and development regulator, Insulin-like growth factor-II mRNA binding proteins (igf2bps) that fine-tune the expression of Insulin-like growth factor II (IGF II) via post transcriptional regulation, and thereby affects animal size was identified (Baker *et al.* 1993; Bandziulis *et al.* 1989). Significant differences in protein level of igf2bp1 was identified in the dataset of species common DEPs.

4.4.2.3. Comparative analysis of common proteins in DS, SA and AR embryos (global proteome and DEPs)

To conduct comparative analysis of the profile of common proteins identified in DS, SA and AR (Figure 4.3A) data dimensionality was reduced with PCA (Principal Component Analysis) and the pattern of protein clustering in each embryonic stage (BH and H) and in each species was established (Figure 4.4). 67.8% (PC1+PC2) in variance was found to represent the total sample dispersion. The variations between embryonic stages within the same fish species are much smaller than that of interspecies variations. The three replicated samples from each embryonic stage clearly clustered within 95% CI, indicating different fish species possessed a distinct protein pattern during embryonic development.

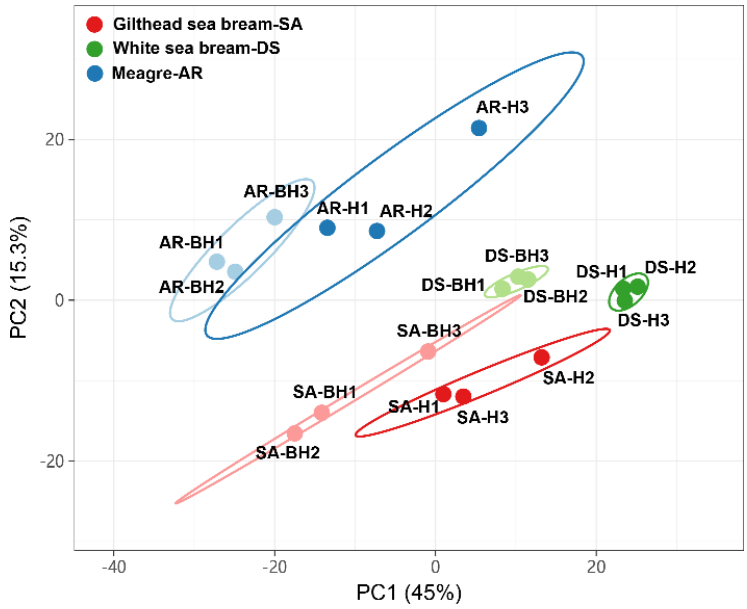


Figure 4.4. Principal component analysis (PCA) of the common protein subset to the three fish species. Principal component 1 and 2 are shown in the plot for the confidently quantified proteins by SWATH processing (FDR < 0.05) in the global embryonic proteome (combining H and BH stage/fish species) that were commonly shared among the three fish species. The elliptical area stands for 95% confidence interval. In the data processing, a 95% confidence interval was chosen to group the different samples (n=3 independent protein extracts from mixed embryonic stages) from by using different color of elliptical area. BH - before hatch; H - hatch; DS - *Diplodus sargus* (white sea bream); AR – *Argyrosomus regius* (meagre); SA - *Sparus aurata* (gilthead sea bream).

To investigate the characteristics of protein expression in different stages of the analyzed fish species, common DEPs (104) at the embryonic stage transition (BH *vs* H), (Figure 4.3B) were extracted and their expression pattern compared in a clustered heatmap (Figure 4.5). The clustering of the protein expression profile clearly separated AR from DS/SA, that were clustered by embryonic stage. According to the clustering pattern DS and SA were more similar and overall, in the BH stage proteins were down-regulated and less abundant than in the H embryonic stage. Additionally two main groups of DEPs (marked by 1 and 2) in the clustered heatmap with obviously distinct patterns of protein levels occurred (Supplementary Table 4.7 annex I). Cluster 1 contained proteins with lower protein abundance in the H stage of all species compared to the BH stage. Cluster 2 contained a group of proteins with increased abundance from BH to H. If the distinct protein expression pattern by stage had biological significance linked to embryo development was assessed by functional annotation.

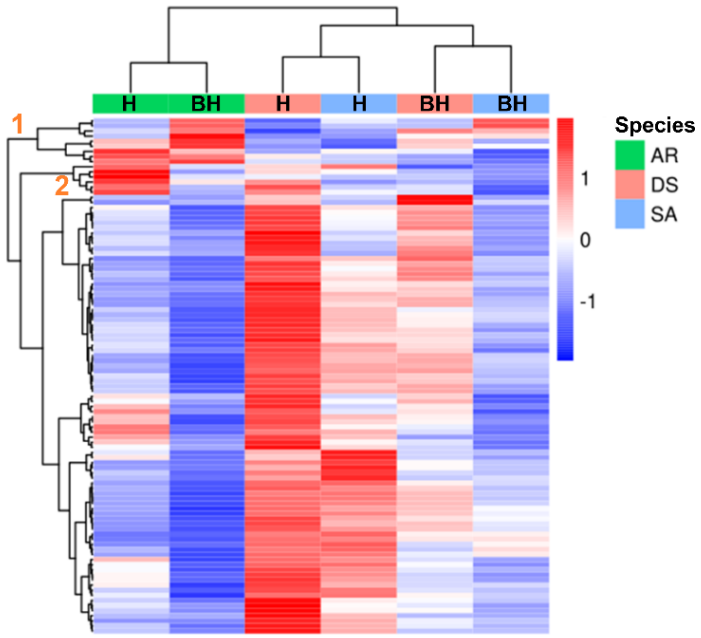


Figure 4.5. Clustering heatmap of the 104 DEPs shared between SA, DS and AR during the BH to H transition. Each column represents one embryonic stage, and each horizontal line represents a protein (based on the median of n=3 biological replicates). Colour ranges from red to blue represent the normalized protein abundance (according to the colour scale presented in the top right corner). The protein clusters marked "1" and "2" on the left hand side are listed in Supplementary Table 7 annex I. BH - before hatch; H - hatch; DS - *Diplodus sargus* (white sea bream); AR – *Argyrosomus regius* (meagre); SA - *Sparus aurata* (gilthead sea bream).

4.4.2.4. Functional analysis of embryonic proteome DEPs among the three fish species

The GOs retrieved for the 104 common DEPs between species for the BH to H comparison were compared and corresponded to a diversity of BPs (Figure 4.6A; Supplementary Table 4.8 annex I). The top five most significantly enriched GO-BP terms were *translation, peptide biosynthetic process, amine biosynthetic process, peptide metabolic process, and cellular amine metabolic process*. Ten significant KEGG pathways were enriched (Figure 4.6B; Supplementary Table 4.8 annex I) and the top five significantly enriched KEGG pathways were *ribosome, spliceosome, carbon metabolism, citrate cycle (TCA cycle), and adherens junction*. Most of the enriched BPs were related to protein synthesis and translational/transcription machinery and this was also reflected by enrichment of "Ribosome" and "Spliceosome" in the KEGGs pathways. Of note was the significant enrichment of *cellular response to estrogen stimulus* and *response to estrogen* in GO-BPs. Several core developmental signaling pathways were enriched including the *VGEF signaling pathway*, indicative of intense angiogenesis at the BH/H transition (Huang *et al.* 2022) and the enrichment of Notch signaling, indicative of organ development (Supplementary Table 4.8 annex I) (Raya *et al.* 2003; Shi & Stanley 2006).

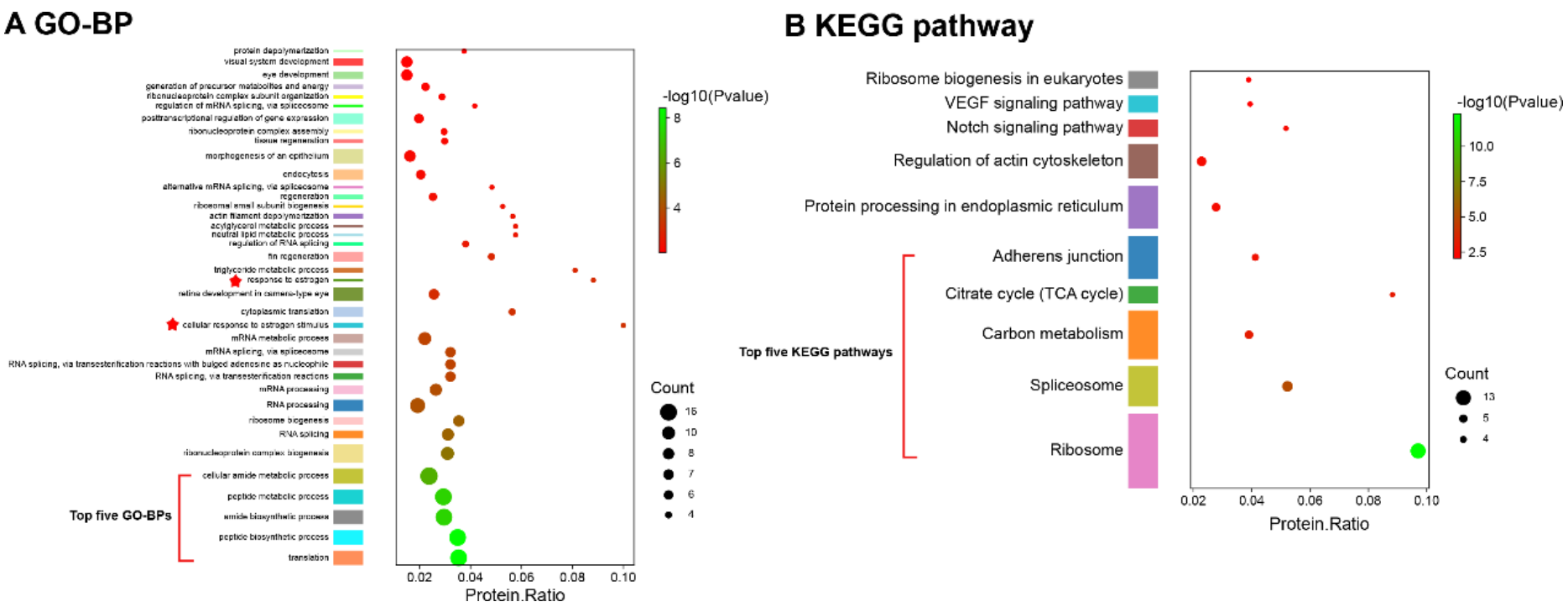


Figure 4.6. Functional enrichment analyses of the 104 common DEPs in the SA, DS and AR during the BH to H transition. Bubble plot of the enriched Gene Ontology Biological Processes (GO-BP) retrieved by the 104 DEPs (A), and KEGG pathway (B) that were overrepresented. GO and KEGG terms are shown on the left of the bubble plot. BPs and Pathways were selected to display based on enriched p-value < 0.01 and color ranges from green to red represents the significant difference of each BP from highest to lowest. Number of proteins from proteomic analysis that contributed to the enriched BP are shown in dark circle with different size on the right side of the bubble plot. Top five BPs and pathways were displayed by red square brackets. Two endocrine-related BPs were marked by red star. For detailed lists of all significantly enriched GO terms and groups and associated KEGG pathways consult Supplementary Table 4.8 annex I.

Enrichment of Ssr1 (signal sequence receptor, alpha), Vtg2 (vitellogenin 2), and Hdlbpa (high density lipoprotein binding protein a) are DEPs in the present study and are functionally associated to the aforementioned endocrine-related pathways, indicating the activation of these specific processes during early development (Supplementary Table 4.8 annex I).

The GO-BPs identified were coherent with the embryonic morphological and functional alternations at the studied developmental stages. GO-BPs like *eye development* and *visual system development* were significantly enriched in the species common DEPs between BH and H including ctsd, hdac1 (probable histone deacetylase 1-b-like), smarca5 (swi snf-related matrix-associated actin-dependent regulator of chromatin subfamily a member 5), cdc42 (cell division control protein 42 homolog isoform 1), mvp (major vault), igf2bp1, top2b (dna topoisomerase 2-beta), sf1 (splicing factor 1 isoform 1) identified in all species and all significantly up-regulated during the BH to H transition (Supplementary Table 4.8 annex I), indicating histogenesis of the visual system in embryos between BH and H. Muscle-related BPs (*skeletal muscle organ development*, *skeletal muscle tissue development*, *muscle organ development*, *striated muscle tissue development*, *muscle tissue development*, *muscle structure development*, *muscle cell development* and *striated muscle cell differentiation)* were significantly enriched in the DEPs identified during the BH to H transition in DS (Supplementary Table 4.9 annex I), indicative of up-regulated muscle development and in accordance with the highly abundant muscle-related DEPs in DS. BP and KEGG pathways obviously associated with development-linked morphology were not identified when DEPs only expressed in SA and AR were analyzed.

The results of the study extend the understanding of species-specific molecular mechanisms that govern the BH to H transition in each species. The 73 DEPs only expressed in SA generated a range of GO-BPs and KEGG pathways (Supplementary Table 4.10 annex I) and included in the topmost 5 BPs, *ribosomal subunit export from nucleus, ribosome localization, rRNA-containing ribonucleoprotein complex export from nucleus, intracellular protein transport, chromosome organization*. The top five KEGG pathways included g*lutathione metabolism, carbon metabolism, RNA transport, ribosome biogenesis in eukaryotes, RNA degradation*. The 267 DEPs only expressed in DS identified as the five top GO-BPs, *Carboxylic acid metabolic process, oxoacid metabolic process, organic acid metabolic process, cellular amide metabolic process, monocarboxylic acid metabolic process* (Supplementary Table 4.9 annex I). The five top KEGG pathways were *Carbon metabolism, valine, leucine and isoleucine degradation, biosynthesis of amino acids, glycolysis / gluconeogenesis, pyruvate*

metabolism (Supplementary Table 4.9 annex I). In AR, the GO-BPs most significantly enriched in the 148 species-specific DEPs were *translation, peptide biosynthetic process, peptide metabolic process, amide biosynthetic process, and cellular amide metabolic process*. The five most significant KEGG pathways were *ribosome, protein export, aminoacyl-tRNA biosynthesis, protein processing in endoplasmic reticulum, and RNA transport* (Supplementary Table 4.11 annex I). In fact, different biological processes, and pathways (transcription/translation machinery-SA; organic metabolic process and energy metabolism- DS; protein synthesis-AR) are associated with the embryonic stage transition in DS, SA and AR and may be a consequence of species-specific differences in developmental stage or the hatching process. The GO-BPs and KEGG pathways enriched when species specific DEPs identified between BH and H were compared were more similar between SA and AR compared to DS. Nonetheless, the proteins in the common KEGG pathways found between species were not always identical and this explains in part the existence of DEPs between BH and H that were only found in a single species.

The number of common GO-BP and KEGG pathways enriched in the three fish species were identified and are represented in Figure 4.6 (Supplementary Table 4.8-4.12 annex I). Six GO-BPs were enriched in all the fish species (*translation, peptide biosynthetic process, amide biosynthetic process, peptide metabolic process, cellular amide metabolic process, generation of precursor metabolites, and energy)* when DEPs expressed during the BH and H transition were analyzed (Supplementary Figure 4.2A; Supplementary Table 4.12, both in annex I). It is noticeable that the GO-BPs *cellular response to estrogen stimulus* and *response to estrogen* were represented within the 104 DEPs common to the three fish species (Supplementary Figure 4.2A annex I), indicating their involvement in the hatching process of all species. For the KEGG pathways, carbon metabolism is the only pathway to be shared in all species (Supplementary Figure 4.2B; Supplementary Table 4.12, both in annex I). Four additional pathways were significantly enriched: *adherens junction, regulation of actin cytoskeleton, notch signaling pathway, and VEGF signaling pathway* in the 104 common DEPs, and were not enriched in other species-specific DEPs.

In the functional analysis of the DEPs between BH and H, the protein He, that drives hatching was identified. A single He protein isoform was found in SA and DS and the levels were of the identified both decreased from BH to H. Moreover, none of the He proteins were enriched in any of the processes or pathways identified, suggesting a relatively specific role for this protein unlike the other proteins that clustered by function.

4.4.3 Hatching enzyme (He) expression analysis

To thoroughly study the He protein, SWATH quantitative proteomics analysis revealed that He2 protein was significantly down regulated (p-value < 0.05) at H compared to BH in the embryonic proteome of both SA and DS (Figure 4.7A; Supplementary Tables 4.4-4.5 annex II). The reduction of He in the hatching stage was unexpected as these enzymes are important for successful hatching (Carballo *et al.* 2019). To assess if the proteomic pattern observed in DS, SA and AR was common to SA and DL during the BH to H transition *he* gene transcripts were analyzed by QPCR, since specific antisera were unavailable for Western blotting.

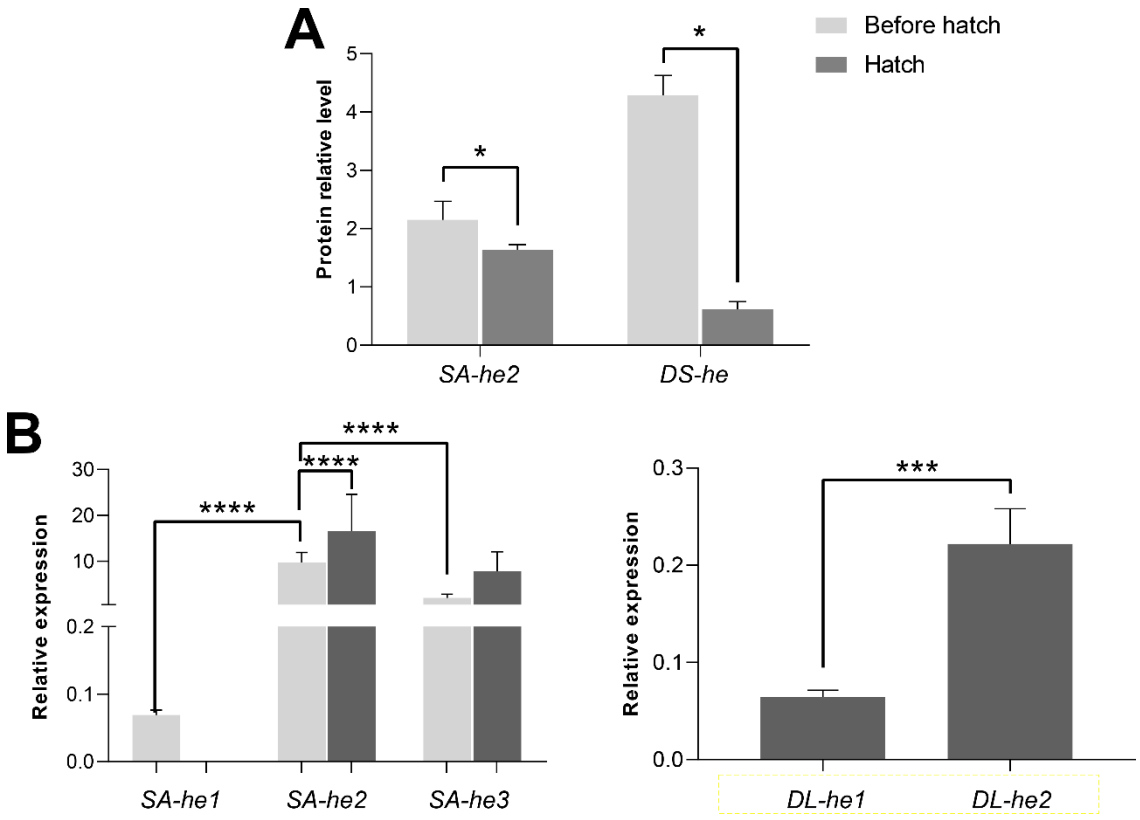


Figure 4.7. Hatching enzyme (He) protein content and relative expression analysis by qPCR during fish early development. A) Graphical view of the relative expression of He2 protein, extracted from the data of the quantitative SWATH proteomic analysis of the embryonic stages from SA and DS. **B)** Relative expression of *he´*s transcripts in SA and DL at two embryo developmental stages (BH and H). Transcript levels were normalized by dividing the calculated transcript copy number by the expression of the reference genes 18s (DL) and *fnta* (SA). See Supplementary Table 4.1 annex I for the complete list, accession numbers and abbreviations of all analysed transcripts. Data are presented as the mean ± SEM (n=6 and 12 / SA group; n=15 / DL group). *, *** and **** represent statistically significant differences, for p < 0.05; 0.001; 0.0001, respectively. BH - before hatch; H - hatch; DS - *Diplodus sargus* (white sea bream); SA - *Sparus aurata* (gilthead sea bream); DL - *Dicentrarchus labrax* (European sea bass).

The number of He isoforms was determined by blast searches in the SA and DL genome and 3 *he* genes were retrieved from each species. The gene *he2* encoded the DEP identified in

SA (*SA-he2*) and quantified by SWATH-MS (Supplementary Table 4.1 annex I). The expression of the three *he* transcripts (*he1, he2, he3*) in SA was detected in BH and H embryos and in DL only *he1* and *he2* were identified at H (Figure 4.7B). The results revealed that *he3* was only expressed during embryonic development in SA. The expression of *he1* was detected in SA at the BH stage and in DL at the H stage and *SA-he1* and *DL-he1* had lower expression levels compared to the other *he* genes ($p < 0.001$) at the same embryonic stages. *he1* transcript was not amplified in SA at H stage, reflecting the stage specific expression. *SA-he2* transcripts were significantly up-regulated at H compared to BH ($p < 0.0001$). He2 protein increased from BH to H in SA and DS and *he2* gene expression increased from BH to H embryonic stages, suggesting He2 as the enzyme driving the hatching event in fish embryo. No significant differences in expression were found for *SA-he3* between BH and H stages. The abundance of *he* gene isoforms in BH embryos of SA was significantly different (p-value < 0.0001).

4.5. Discussion

Differential cellular functions are associated with the morphological and physiological changes that occur during the early developmental stages of fish, and is reflected in variations in the associated global proteome. Therefore, understanding the cellular mechanisms underlying fish development by mapping the proteome during early stages is important and of relevance for tracing the species phylogenetic origin and finding molecular markers that explain species differences. A comparative proteomic approach was established to better understand the molecular dynamic and protein modeling between BH and H stages during fish ontogeny of the three fish species. Quantitative proteome probing using a data independent acquisition method, SWATH, was shown to generate robust quantification of the global proteome for the three fish species. Three independent fish specific embryonic protein libraries were generated, and the quantified proteins revealed a generally conserved protein landscape during fish early ontogeny. The conserved and common proteins in the embryos of the different fish species had a distinct clustering pattern, probably due in part to variable protein levels. Of note was the very limited number of DEPs that were classified as immune or endocrine during embryonic development. However, the identification of proteins with a species-specific pattern of expression suggests they may be useful as species-specific molecular markers, as well as suggesting divergence in the ontogenetic timing of biological processes between species something that will require further investigation in the future. To test the robustness of the proteins identified in the BH to H transition, we considered a pivotal molecule in hatching namely hatching enzyme (He). Three

hes transcripts were identified in SA and DL genomes, respectively. The expression profile of the three *hes* suggests the different isoforms may have different functional importance during the hatching process and that this may vary according to species.

The pattern of total extractable protein in embryos when fractionated by SDS page was similar for the same stages of SA, DS and AR. Myriad proteins contribute to the conserved profile across different species between the two developmental stages analyzed (> 60% of the quantified proteins were common to all fish species). Similar results have previously been reported in the olive flounder (*Paralichthys olivaceus*) where conserved protein profiles (SDS-PAGE analysis) were identified among different embryonic developmental stages, despite the slight difference in protein content (Kim *et al.* 2018). Moreover, several vitellogenin proteins belonging to the vitellogenin 2 family were found from blastula to the tail bend stage and higher-molecular-weight vitellogenin proteins were degraded and may be important for morphogenesis in early embryos. Similar studies were also carried out on zebrafish, and more notable differences were reported between different development stages on SDS-PAGE (Yilmaz *et al.* 2017; Purushothaman *et al.* 2019). However, these studies did not include a comparative analysis among different species to investigate and explain species heterogeneity during early development.

The three species specific embryonic protein libraries (combining BH and H stage) generated include a large number of proteins that were confidently identified/quantified in three fish species. The numbers of proteins identified in the present study was substantially higher than previous proteome studies of fish embryo (Niksirat *et al.* 2017; Yilmaz *et al.* 2021; Kohn *et al.* 2015). Despite the limited number of proteomics studies that exist for fish embryos those from zebrafish suggest that for different developmental stages, despite the drastic morphological changes during the first week of development, the patterns of abundant proteins are largely conserved (Tay *et al.* 2006). By analogy, the conserved developmental ontogeny during fish oogenesis and early embryogenesis led us to hypothesis that this may be linked to conserved functional molecules, such as proteins, of autogenous and maternal origin. In a quantitative (protein numbers) comparison between SA, DS and AR in addition to the higher number of common proteins shared between the three fish species, a higher number of common proteins were shared between SA and DS than with AR, reflecting the phylogenetic proximity of the species. In fact, phylogenetic approaches based on teleost mitochondrial and nuclear genes have shown SA and DS are more proximate than AR (Meynard *et al.* 2012). Other important observation, common across the fish analyzed was the tendency for protein abundance to increase rather than decrease during the BH to H transition. Similar observations

were made in relation to the abundance of proteins in a previous proteomics study of zebrafish embryos (Purushothaman *et al.* 2019), where the number of proteins increased from MZT to the bud stage in both non-deyolked zebrafish embryos, and this was even more evident in deyolked embryos. Specific comparisons of the proteins that changed in the present study and the previously published zebrafish study are not possible as proteins that significantly increased or decreased were not identified.

To further explore the proteomic data, PCA was used to assess the 694 common proteins and showed the data dispersion in non-deyolked embryo samples between different fish species. The results revealed a distinct pattern in proteins between stages and species, although interspecies differences were more obvious than those between embryonic stage (intraspecies). This is the first time that developmental differences were reported based on protein category and level among phylogenetically proximate species and presumably reflects phenotypic and physiological differences. Studies revealing species divergence in gene expression have been proposed to result from, and drive phenotypic plasticity and divergence between species (Wittkopp 2007). Moreover, since proteins are the outcome of gene expression they can be considered to represent a snapshot of genome expression and more directly influence phenotype (Lopez 2005; Shen *et al.* 2018). As a product of gene expression, proteins can also be regarded to be important for species differentiation. In the present study, quantitative proteomics analysis identified core proteins that may explain interspecies and stage-specific differences and future work should be centered on identifying the functional importance of the proteins identified. It is intriguing that a small subset of DEPs in the clustered heatmap of AR showed a striking alteration in protein level, which may be suggestive of an important role for these proteins during the AR BH to H transition.

Genomic sequencing has made it clear that a large fraction of the genes specifying core biological functions are shared by all eukaryotes. Knowledge about the biological role of such shared proteins in one organism can often be transferred to other organisms (Ashburner *et al.* 2000). Furthermore, it is well recognized that embryonic development is governed by highly coordinated changes in the expression of large protein sets (Lucitt *et al.* 2008). Several studies indicate that numerous genes expressed in oocytes are conserved across species, for instance, mouse (*Mus musculus*), bovine (*Bos taurus*), Xenopus (*Xenopus laevis*) and vase tunicate (*Ciona intestinalis*) due to their conserved and essential functions during development (Evsikov *et al.* 2006; Vallee *et al.* 2008). Proteomic studies of unfertilized eggs/oocytes in sea bream identified chaperonin containing (TCP1), serpin A1, and importin alpha1 as putative maternal factors (Ziv *et al.* 2008).

In the present study, the GO-BP and KEGG pathways of total quantified proteins revealed the scale of common biological functions among the three species analyzed. The results showed that the GO-BPs identified by the total quantified proteins were common in the three species, especially those for protein synthesis (*translation, peptide biosynthetic process, amide biosynthetic process, and peptide metabolic process*). Proteomics studies of fish non-deyolked embryos, aimed at cataloguing the complete set of proteins, have revealed that enriched molecular processes in zebrafish embryos are cellular organization, cell cycle, control of replication and translation, and mitochondrial functions (Purushothaman *et al.* 2019). We also noticed that many regulatory processes in fish embryos that were significantly enriched shared similarities to embryos of higher vertebrates (e.g. human) that have a large fraction of the protein mass dedicated to regulatory functions (Beck *et al.* 2011). Enrichment analysis of zebrafish embryos revealed that the proteins in high abundance were mainly related to GO-BPs of basic metabolic functions (such as primarily translation related processes at 24 hours post fertilization) and low abundance proteins were related to embryonic morphogenesis and development (Alli Shaik *et al.* 2014). The overlap and commonality of GO-BPs identified in proteomics studies of early vertebrate development is presumably explained by the shared general morphological characteristics of embryo development (Arezo *et al.* 2005), and is coherent with the conserved patterning of gene expression in embryos of *M. musculus*, *G. gallus*, *X. laevis*, and *D. rerio* (Irie & Kuratani 2011).

Based on our findings, several metabolic processes and pathway were enriched in all species (SA, DS and AR), indicating that during fish embryogenesis similar biological processes are activated. In a proteomic study of zebrafish embryogenesis the GO-BPs identified were mainly involved in cellular organization/biogenesis, metabolism/energy, transport, and signal pathways, which were also important processes in our study (Lin *et al.* 2009). A proteomics study of European whitefish (*Coregonus lavaretus*) larvae revealed their responsiveness to the environment since osmoregulation related processes were enriched under high salinity, and a divergent molecular response of larvae from different fish populations was identified (Papakostas *et al.* 2012). Taken together the proteomics studies of fish embryos reveals relatively good conservation of BPs between different fish species, and that the proteome is dynamic and indicates the potential of proteomics studies for uncovering the effect of environmental factor during early development. To sum up, these functional analyses have unveiled different couples of proteins and their dynamic and features in different fish species, although insights into immune- and endocrine-related processes remain limited during fish embryogenesis.

In the present study the proteome and transcriptome generated for the same SA samples were integrated and revealed at total of 543 proteins only present in the proteome, which presumably represent candidate maternal factors (*in Chapter 5*). There is similarity between the protein repertoire of fish oocytes and a diversity of phyla, such as mammals and insects, due to the conservation during evolution of oocyte maturation (Ziv *et al.* 2008). Six immune-related DEPs were common in SA, DS and AR and may be potential biomarkers for immune status and the developmental performance of non-deyolked embryos. Among the identified proteins was Ctsd that has previously been proposed as a potential indicator of teleost fish egg and embryo quality in DL and SA (Carnevali *et al.* 2001a; Carnevali *et al.* 2001b). Common enriched KEGG pathways were identified in the starlet sea anemone (*Nematostella vectensis*) and the mouse MII oocyte, indicating there is substantial conservation in oocyte functional pathways (Lotan *et al.* 2014). In the present study similar protein profiles and common KEGG pathways were identified in three fish species, and presumably reflects a similar molecular basis for embryonic development in these phylogenetically proximate fish.

Developmental proteome reveals the dynamics of protein changes that have reported in several species during oocyte maturation or embryogenesis, such as Drosophila (Kronja *et al.* 2014), Drone Honeybees (*Apis mellifera ligustica* Spinola, 1806) (Kronja *et al.* 2014) and Xenopus (*Xenopus laevis*) (Sun *et al.* 2014). Several studies of fish ploidy have provided a molecular clue for indicating embryo developmental performance. A comparative proteome was conducted on eye formation stage of gynogenetic haploid and diploid goldfish (*Carassius auratus*) embryos (Huang *et al.* 2004). Half of total differentially expressed proteins only found in diploid embryos (sixteen proteins) but not in haploid embryos, including specific protein candidates that are correlated with eye development, nerve development, developing regulation, cell differentiation, and signal transduction. The authors speculated that the differences in the proteome could explain the abnormality of haploid embryos. Our study links protein dynamics to the morphological alterations between BH and H of normal ploidy embryos of DS, SA and AR, and revealed the proteins of wild type fish embryos that are associated with the profound changes in morphology and physiology during this developmental transition. A study of the proteome of rainbow trout diploid and triploid embryos and fry (Babaheydari *et al.* 2017) revealed structural and metabolic proteins were significantly modified in abundance in triploids and it was proposed this may be related to their increased cell size and lower growth performance compared to diploid fish. Other studies have focused on the development of SA germ cells using comparative proteomics (Ziv *et al.* 2008; Zilli *et al.* 2014) and reveal again the high conservation of the fish oocyte proteome presumably due to evolutionary conservation of

oocyte maturation across species.

As the final step in embryogenesis, hatching is a crucial milestone and is the transition between embryogenesis and larvogenesis. To study the mechanism of this embryonic stage, we focused on *hes* and exploited this pivotal molecule in two of the studied teleost species (SA and DL) using a comparative approach at the transcript and protein level. Three *he* genes were retrieved by *in silico* analysis from the SA and DL genomes. Only two *he* genes were expressed in DL embryos and three *he* (*he1, he2* and *he3*) genes were expressed in SA. Previous studies have reported that the expression pattern of three *he* genes (*he1a*, *he1b* and *he2*) in zebrafish and crucian carp embryos, *he1a*, as well as *he1b* were highly expressed before hatch and had low expression levels until hatch, but *he2* was only expressed in the zebrafish embryo (Fu *et al.* 2021). In contrast, the four *hes* identified in the flatfish, Senegalese sole, had a similar expression pattern during development (Carballo *et al.* 2019) and all *he* genes were detectable from gastrula but the transcript abundance increased significantly 8h before hatching and remained elevated until larval hatching. In the present study, the transcript abundance in SA of *he3* was high in BH and H, while *he2* was significantly up regulated from BH to H and *he1* was very low abundance in the BH stage. The proteomics results showed that He2 was a DEP between embryonic stages in SA and DS. However, comparison of gene transcripts and proteins revealed an inverse relationship with decreased He2 protein content from non-deyolked SA and DS embryos at hatch compared to the gene transcripts. The divergent results between the proteome and transcriptome results reveal the importance of caution when interpreting such data. Overall, there was a divergent abundance of hatching enzyme isoforms during development in different fish species indicating divergence in the hatching mechanism.

4.6. Conclusion

A comparative proteomics analysis among three fish species during early development was performed to dissect the molecular mechanisms that drive fish ontogeny. A conserved protein landscape was found in total quantified proteins across the species. In contrast, few common DEPs were identified between the species at the BH to H transition. The identification and quantification of species-specific DEPs provides a promising resource for species specific identification during ontogeny and candidate proteins for understanding species-specific phenotypes. The identification in all species of two significant endocrine-related BPs highlights the importance of endocrine factors in the regulation of normal embryo development. A protein-poor immune repertoire existed in the embryos, suggesting in early stages intrinsic barriers and physicochemical factors may be more important. For the functional potential of pivotal

molecules in prompting hatch and embryo quality, hatching enzyme was investigated to determine the levels of their transcripts and to compare this with their protein levels, and revealed proteins and transcripts had a divergent expression pattern. Overall, the outcome of this study provides a large protein dataset for fish developmental biology and quality assessment.

CHAPTER 5

Screening for potential molecular markers for gilthead sea bream quality performance in commercial hatcheries

Lisen Li[1], João R Cardoso[1], Adelino VM Canario[1,2] and Deborah M Power[1,2]

[1]*Centro de Ciências do Mar (CCMAR), Universidade do Algarve, Faro, Portugal.*
[2]*International Institution of Marine Science, Shanghai Ocean University, Shanghai, China.*

Manuscript in revision

CRediT statement; LL- Investigation and experimental methodology, data analysis, writing - original draft, DMP – conceptualization, funding acquisition, project management, supervision, data analysis, writing – drafting, review & editing, JRC – data analysis, drafting, review & editing, AVMC – data analysis, drafting, review & editing.

Acknowledgements

This work was supported by funding from the European Union Horizon2020 Programme (PERFORMfish, grant nº 727610). The authors are especially grateful to Soraia Santos for her support in maintaining the sample biobank and RNA extractions.

5.1. Abstract

Effective classification of fish embryonic quality is a major bottleneck in commercial hatcheries as it limits productivity and restricts aquaculture sustainability. Current parameters for embryo quality assessment lack common consensus criterion as they are largely dependent on environmental rearing conditions at hatchery production sites and biotic factors, which diverge between different fish species and with broodstock. Here we have generated a large-scale transcriptomic analysis of batches of gilthead sea bream (*Sparus aurata*) non-deyolked embryos of distinct quality and collected from different hatcheries across Europe. We provide for the first time a comprehensive insight into the molecular events that regulate embryo development and quality. Analysis of the transcriptome between good and poor-quality non-deyolked embryos at the same stage but from different European hatcheries revealed large variations exist between sites and only 42 DE transcripts were common and may be considered genes potentially linked to quality. Functional enrichment of the DE genes between good and poor-quality embryos revealed enrichment of metabolic, biosynthetic, and developmental related biological processes and pathway analysis indicated some genes are associated with immune-related pathways. Non-coding RNAs were also found to be DE and we hypothesize that they are likely to be regulators of gene transcription during embryo development. Integration of gilthead sea bream embryo transcriptome data with the embryo proteome identified 543 maternally inherited proteins in the sea bream non-deyolked embryo some of which were related to immunity. This study contributes to improve understanding of sea bream developmental biology and plasticity and identifies novel regulatory factors that may be linked to embryo quality in hatcheries as well as candidate embryo quality markers.

Keyworks: Commercial hatchery; embryo quality biomarkers; maternal transcripts; sea bream non-deyolked embryo; transcriptomes.

5.2. Introduction

Aquaculture is the fastest growing food-production sector in the word. It currently provides 52% of the total fish consumed by humans and this industry is valued at USD 250 billion. Moreover, aquaculture production is expected to expand and diversify so it can deliver the future global demands for food (FAO, 2020). One of the major bottlenecks in aquaculture is the unpredictable quality of fish eggs that can affect the production volume and management of production cycles. Fish egg quality is commonly defined as the "*ability of the egg to be fertilized and subsequently develop into a normal embryo*" (Bonnet *et al.* 2007b). To achieve sustainable fish production, appropriate husbandry and management approaches are essential. This requires a good understanding of how intrinsic factors such as broodstock genetics, fish age and diet affect egg and embryo quality (Ma *et al.* 2019a) as well as the impact of external environmental factors such as water quality, salinity and photothermal cycle or other potential stressors such as the microbiota (Contreras-Sanchez *et al.* 1998; Campbell *et al.* 1992; Bonnet *et al.* 2007a; Aegerter & Jalabert 2004).

Currently, monitoring of egg and embryo quality in aquaculture is based on the qualitative evaluation of embryo morphology, embryo symmetry and buoyancy, fertilization rate and development and growth performance (see review Li et al., chapter 1). However, high species variability and the diversity of egg and embryo characteristics make the establishment of common criteria for quality assessment in aquaculture difficult (Kjorsvik *et al.* 2003). While in salmonids fertilization rate is a common reference indicator for egg and embryo quality (Kjorsvik *et al.* 2003; Ienaga *et al.* 2021) in the Atlantic cod (*Gadus morhua*), Atlantic halibut (*Hippoglossus hippoglossus*), turbot (*Scophthalmus maximus*) (Shields *et al.* 1997; Kjorsvik *et al.* 2003), and other marine fish, cell symmetry at early cleavage stages (normal blastomeres) is proposed as a quality indicator (Kjorsvik *et al.* 1990). New approaches for assessing egg and embryo quality based on molecular data and the identification of regulatory factors for normal fish egg and embryo development may be a source of quality markers associated with healthy offspring that perform well in aquaculture (Ma *et al.* 2019a).

Fertilization of vertebrate eggs trigger a conserved developmental program that starts from a single egg cell stage and culminates in a multicellular organism and passes through several well recognized stages, zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching (Mishima & Tomari 2016). During oogenesis uptake and accumulation of maternal nutrients and factors essential for the development of the zygote occurs. After fertilization the maternal-to-zygotic transition (MZT) (Tadros & Lipshitz 2009) marks the suppression of

maternally inherited transcription and the activation of zygote gene expression and the concurrent formation of the yolk syncytial layer (YSL) (Lubzens *et al.* 2017). The decay of maternally inherited mRNA prevents inappropriate expression of maternal transcripts during the zygote stages (Ma *et al.* 2019a; Jukam *et al.* 2017; Bashirullah *et al.* 1999). Maternal RNAs are synthesized and accumulate in growing oocytes and have an essential role during early embryogenesis (Schisa 2012; Lubzens *et al.* 2017) and in fish may affect egg and embryo quality (Abrams & Mullins 2009; Aanes *et al.* 2011; Lanes *et al.* 2013; Howley & Ho 2000; Lyman & Pelegri 2007).

In the Atlantic cod (*Gadus morhua*) maternal transcripts related to fructose, fatty acid, glycerophospholipid metabolism, and oxidative phosphorylation were proposed as egg and embryo quality biomarkers (Lanes *et al.* 2013). In zebrafish, maternal RNAs of *ski7*, *otulina* and *slc29a1a* genes were proposed to be linked with egg and embryo quality (Cabrera-Quio *et al.* 2021; Cheung *et al.* 2019). Moreover, transcriptome analysis of low- and high-quality European sea bass embryos using a microarray identified six differentially expressed genes (*rnf213, irf7, usp5, mem-prot, plec* and *cenpf*) as potential candidate quality biomarkers for European sea bass development (Zarski *et al.* 2017). Recently the role of non-coding RNAs (ncRNAs) on the regulation of gene expression and their association with embryo quality has been suggested but remains to be demonstrated (Li et al., 2022).

Despite the advances that have been made in fish biology using genomics, transcriptomics and proteomics the diversity of fish species used for aquaculture means resources are only available for a few species. A few notable exceptions exist, for example, the salmonids although most studies of embryonic development are based on non-aquaculture fish models such as the zebrafish (*Danio rerio*). Relatively few transcriptome and proteome studies exist for commercial aquaculture species, which makes the selection of putative quality biomarkers and the assessment of their suitability for application in non-target species difficult. The gilthead sea bream (*Sparus aurata*) is a valuable species for fisheries and aquaculture and important for the economy of the Mediterranean region (Grigorakis *et al.* 2002; Llorente *et al.* 2020). This marine teleost is a Perciform of the Sparidae family, a hermaphrodite (Pavlidis & Mylonas 2011), its genome is sequenced (Pauletto *et al.* 2018) and its production in aquaculture is built on a solid knowledge base about its biology (Zorrilla *et al.* 2003; Aly *et al.* 2021a; Deguara *et al.* 2003; Aly *et al.* 2021b; Ferosekhan *et al.* 2021; Bessonart *et al.* 1999; Ibarz *et al.* 2010; Perera *et al.* 2021; Pulido-Rodriguez *et al.* 2021; Carnevali *et al.* 2001b).

In the present study we developed transcriptomes for gilthead sea bream non-deyolked

embryos that were obtained from multiple commercial hatcheries with different broodstock and production regimes and different geographical locations in southern Europe. The multiple transcriptomes generated were used to investigate and characterize the mechanisms and factors associated with embryo quality during embryo development by comparing production batches classified as good or bad quality. The hatchery-based classification of quality was based on hatching rate, buoyancy, and the morphology at first cleavage. Taking multiple data analysis approaches for transcriptome comparisons, candidate genes of embryo quality for gilthead sea bream were identified. The data collected includes coding gene transcripts, as well as putative inherited maternal and ncRNA transcripts and forms the basis for future studies aimed at understanding their fundamental role in sea bream embryo development and quality.

5.3. Materials and methods

5.3.1. Sampling and *Ethics statement*

Gilthead sea bream embryos were obtained from three European commercial hatcheries from France (***F***), Greece (***G***) and Italy (***I***). Replicate embryo samples (n= 4 to 6 per site) from different quality egg batches [good quality (g) and poor quality (p)] were collected for analysis. Quality assessment of the different egg batches was performed by the companies and was based on hatching rate, buoyancy, and morphology of the fertilized egg at first cleavage. Replicate samples were collected from the same tanks (Figure 5.1) using a similar sampling protocol between different companies. Samples contained embryos from mixed broodstock at each hatchery ***F***, n = 6; ***G***, n = 4, ***I***; n = 4) and the broodstock of each hatchery were genetically distinct. Samples of embryos were collected as part of the routine sampling at hatcheries to investigate production performance and was carried out in compliance with 2009/58/EC (protection of animals kept for farming) supervised by a veterinary doctor. The collaborating commercial hatcheries had GLOBAL G.A.P. (Good Agricultural Practice) certification, which includes demonstration of compliance with good animal welfare practices.

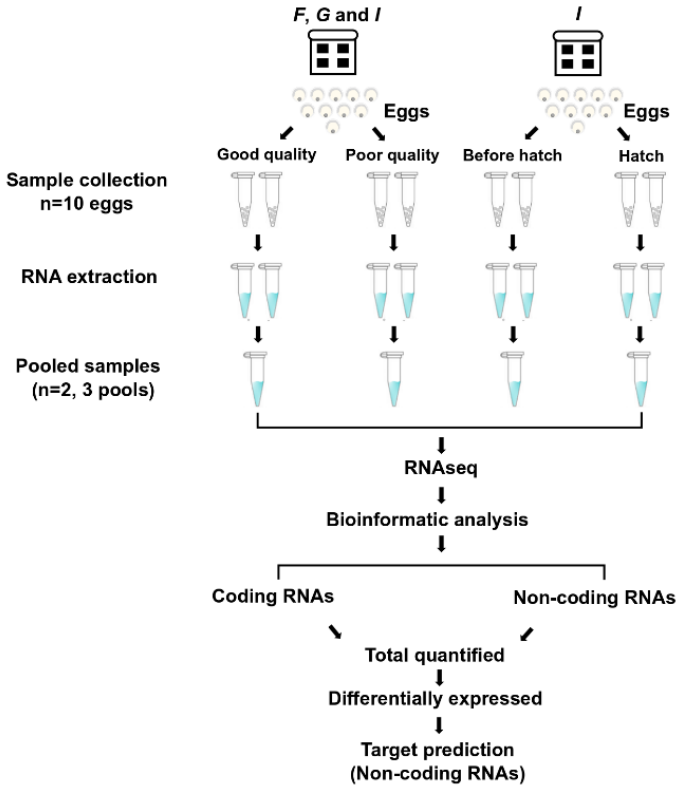


Figure 5.1. Workflow of sampling for transcriptome analysis.

5.3.2. RNA isolation

Total RNA from embryos was extracted using an E.Z.N.A kit (VWR, USA) following the manufacture's protocol. Briefly, 10 embryos per sample were placed in 700 µl of lysis buffer in a microcentrifuge tube and homogenized by physical disruption using two iron beads (5 mm) and 4 cycles of 30 s at room temperature in a Tissue lyser II (Qiagen, Germany) and column purified. DNase I digestion (Omega-Biotec Tek) was performed directly on the column containing the purified RNA. The quality and integrity of the extracted total RNA was checked by gel electrophoresis on a 1% agarose gel in 1x TAE (40mM Tris, 20mM acetic acid, 1mM EDTA disodium salt pH=8.0) buffer and by measuring the concentration using using a NanoDrop One (Thermofisher, Spain).

5.3.3. Library preparation and sequencing

Each sample sent for sequencing was composed of a pool of two independent total RNA extracts present in equal amounts to give a final concentration of at least 1.5 µg and were stored in TE (10mM Tris + 1mM EDTA) buffer at -20 ºC. Library preparation and sequencing of total RNA extracted from embryos was performed by Shanghai Majorbio Bio-Pharm Technology

Co., Ltd (Shanghai, China). Embryos transcriptome libraries (***F*** **-** n = 3 good, n = 3 poor, ***I*** - n = 2 good, n = 2 poor, ***G*** - n = 2 good, n = 2 poor, ***I*** - before hatch n = 4, hatch n= 2) were prepared using an Illumina TruSeqTM RNA sample preparation Kit (San Diego, CA) using 1µg of total RNA and following the manufacturers protocol. Before total RNA library preparation, the quality was determined using a 2100 Bioanalyser (Agilent, USA) and quantified using a NanoDrop ND-2000 (NanoDrop Technologies, USA) and only samples with an OD 260/280 = 1.8 - 2.2, OD 260/230 ≥ 2.0, RIN ≥ 6.5, 28S: 18S ≥ 1.0, > 2µg were used. Libraries were size selected for cDNA target fragments of 200 – 300 bp on 2% Low Range Ultra Agarose followed by PCR amplification using Phusion DNA polymerase (NEB, Germany) for 15 PCR cycles. Paired-end RNA-seq sequencing libraries were sequenced (2 × 150 bp read length) with an Illumina HiSeq xten (Model: PE151, 8, 151).

5.3.4. Data analysis

Data analysis and annotation was performed following standard methods using Galaxy (https://usegalaxy.eu/) (Batut *et al.* 2018; Afgan *et al.* 2018). FastQC was performed on the forward (F) and reverse (R) raw reads to assess quality and the cutadapt tool was used to trim off low-quality bases to increase mapping efficiency. The FASTQ files (R1/first of pair and R2/second of pair) were submitted for data running with the options "20" for Minimum length and Quality cutoff and "*Read Modification Options*" command. RNA STAR was used for alignment and mapping of reads against the gilthead sea bream annotated genome, available from NCBI (https://www.ncbi.nlm.nih.gov/genome/?term=sea+bream). The "*Featurecounts*" tool was used for counting the number of reads per annotated gene and the DEseq2 tool was employed for differential gene expression analysis. Differentially expressed genes were selected based on an adjusted p-value < 0.001. Quality assessment and clustering of the transcriptomes of all samples (hatchery and embryo quality) was performed using PCA (Principal component analysis), pair-wise heatmap, dispersion estimates, histogram of p-values and M (log ratio) A (mean average)-plots within the "*DESeq2*" tool of Galaxy (https://usegalaxy.eu/).

5.3.5. Annotation of differentially expressed (DE) genes

The accession numbers of NCBI-annotated DE genes were converted to Ensembl accession numbers for subsequent annotation and enrichment analysis. The conversion was performed using bioDBnet online software (https://biodbnet-abcc.ncifcrf.gov/) and using the

gilthead sea bream genome as the reference. To increase the number of DE genes assigned an Ensembl accession number, NCBI transcripts without a homologue sequence in Ensembl were manually analyzed using local BLAST against the sea bream Ensembl genome annotation in Tbtools software v1.098669 (Chen *et al.* 2020). DE ncRNA were identified and classified using the linc2function website (https://bioinformaticslab.erc.monash.edu/linc2function) (Ramakrishnaiah *et al.* 2020; Ramakrishnaiah *et al.* 2021). The nucleotide sequence of DE ncRNA was extracted and uploaded to the linc2function website using "Species Agnostic" and "Full Model" options as the default to obtain the potential protein targets (Protein Interactome column).

5.4.6. Enrichment analysis

Comparative analysis between DE genes identified in samples was performed using Venn diagrams and the online software (https://bioinfogp.cnb.csic.es/tools/venny/) (Oliveros 2015). DE genes were classified using Gene Ontology (GO) analysis under three terms, Biological process (BP), Cellular component (CC), and Molecular function (MF). GO enrichment analysis was performed on the Profiler website (https://biit.cs.ut.ee/gprofiler/convert) using the Gilthead sea bream as the reference species. To identify enriched pathways represented by the DE genes the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed using zebrafish (*Danio rerio*) as the reference species. Functional terms and pathways were considered significant when a p-value < 0.05 was obtained.

5.3.7. Integrative omics analysis

A sea bream embryo proteome (dataset identifier PXD033417) of two developmental stages (Before hatch (BH) and Hatch (H)) was compared with the sea bream embryo transcriptome at similar developmental stages (*I*-BH *vs* H, chapter 4). All identified proteins from the proteome and significantly modified transcripts ($p < 0.001$) were used in the analysis. All sea bream protein and gene transcript sequences were assigned an Ensembl zebrafish gene accession number using bioDBnet online software (https://biodbnet-abcc.ncifcrf.gov/). GO-BP and pathway analysis were carried out on gene transcripts also found in the proteome to specifically obtain putative maternal proteins.

5.4. Results

5.4.1 Sequencing parameters for the sea bream embryo samples

Transcriptome data from 18 embryo samples were successfully obtained (Supplementary Table 5.1 annex I). The total number of trimmed and cleaned reads per sequencing library ranged from 4,020,758 to 25,227,997 base pair (bp) and 86.06 to 94.31 % unique transcript were obtained. PCA and heatmaps were performed using the total quantified transcripts (TQTs) and using all the sequencing data irrespective of hatchery or quality of the samples; this analysis failed to generate coherent clustering. For this reason, PCA analysis was performed using the sequencing data for embryos from each hatchery irrespective of quality and this analysis revealed transcriptome clustering by quality (Supplementary Figure 5.1 annex I).

The PCA sample clustering was coherent with the quality classification given by the companies except in the case of one group of samples in hatchery ***I***, which was classified as poor but clustered with the good quality embryos and so the classification was reassigned to good for the analyses. In the case of hatchery ***G*** two samples that were classified as good quality embryos had ambiguous clustering in the PCA and were excluded from the transcriptome analysis (data not shown). To identify DE genes between good and poor-quality embryos in samples from each hatchery two comparisons were performed, 1) to identify biomarkers related to embryo quality gene transcripts of good versus poor quality embryos were compared and 2) to characterize gene transcripts associated with embryo development and hatching, transcriptomes of embryos before hatch (BH) and at hatch (H) were compared (Supplementary Table 5.1 annex I). The number of TQTs between the four groups analyzed (quality comparison of ***F***, ***G*** and ***I*** and stage comparison of ***I***) was similar, ranging from 20123 to 24539, the number of DE genes identified ranged from 146 in good versus poor-quality embryos for ***F*** to 5130 in BH versus H in ***I*** ($p < 0.001$).

5.4.2 Differentially expressed protein coding and non-coding transcripts in sea bream embryos

The Venn diagram of DE genes between *g* versus *p* quality embryo samples from the different hatcheries showed a variable number of overlapping transcripts with site. Between ***F*** and ***I*** hatcheries 86 DE genes were common and between ***F*** and ***G*** hatcheries 73 DE genes were common and between ***I*** and ***G*** hatcheries 515 common DE genes were identified (Figure 5.2A). Comparisons of DE genes between *g* and *p* quality embryos at the three hatcheries identified only 42 common DE genes indicating that the embryo transcriptome at each hatchery was specific presumably due to the differing broodstocks and / or environmental conditions. Most

of the DE ncRNA gene transcripts were site-specific, but three were common across all hatchery samples (XR_003981779.1, XR_003981339.1, XR_003981340.1) and are candidate factors associated with embryo quality (Figure 5.2B). To be more specific, FUS RNA Binding Protein, Aconitase 1, KH RNA Binding Domain Containing, Signal Transduction Associated 3 are potential protein targets of XR_003981779.1. XR_003981339.1 was predicted to have protein target including FUS RNA Binding Protein, Pumilio RNA Binding Family Member 2, Muscleblind Like Splicing Regulator 1, KH-Type Splicing Regulatory Protein, Serine and Arginine Rich Splicing Factor 10, Serine and Arginine Rich Splicing Factor 1, KH RNA Binding Domain Containing, Signal Transduction Associated 3, ELAV Like RNA Binding Protein 1. For DE ncRNA XR_003981340.1, FUS RNA Binding Protein, KH-Type Splicing Regulatory Protein, Muscleblind Like Splicing Regulator 1, Y-Box Binding Protein 1, RNA Binding Motif Protein 4, ELAV Like RNA Binding Protein 1, KH RNA Binding Domain Containing, Signal Transduction Associated 3, Serine and Arginine Rich Splicing Factor 10, Serine and Arginine Rich Splicing Factor 1 are predicted protein targets (Table 1).

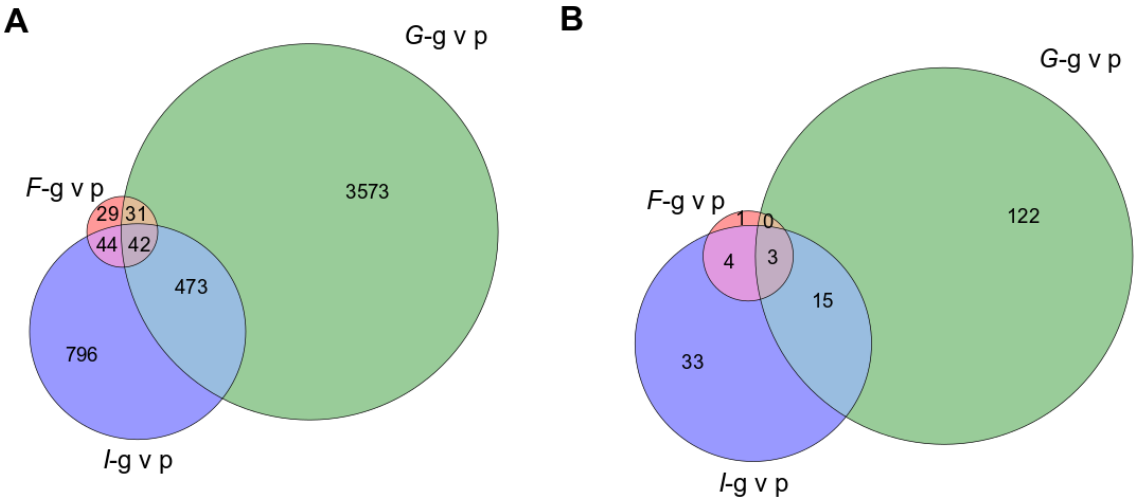


Figure 5.2. Venn diagram analysis of total DEGs (A) and specific DE ncRNAs (B) between the different quality classification collected from the different sample sites (*F*, *G*, *I*). The "g" and "p" indicate "Good" and "Poor" quality embryo batches. Numbers represent the number of DEGs (p-value < 0.001).

Comparison of the DE genes linked to quality (g versus p) or stage of development (BH *vs* H) (Figure 5.3) revealed that 781 protein coding genes (Figure 5.3A) and 38 DE ncRNA (Figure 5.3B) gene transcripts were common. The single DE ncRNA from the quality comparisons (g v p) in ***F*** and the 15 DE ncRNAs that were identified in quality comparisons (g versus p) and were common in ***I*** and ***G*** were selected for RNA target prediction (Table 5.1).

The results revealed that ELAV-like protein 1, Serine/arginine-rich splicing factor 1, KH domain-containing, RNA-binding, signal transduction-associated protein 3, YTH domain-containing protein 1, and RNA-binding motif protein X chromosome are likely to be embryonic quality associated targets for regulation by ncRNAs.

Table 5.1. List of DE ncRNAs and protein target prediction.

LncRNA	Potential protein target
Unique from F-g v p	
XR_003986855.1	Y-box-binding protein 1, Serine/arginine-rich splicing factor 9, ELAV-like protein 1
Common between G-g v p and I-g v p	
XR_003982457.1	Eukaryotic translation initiation factor 4B, Serine/arginine-rich splicing factor 1, Serine/arginine-rich splicing factor 10, KH domain-containing, RNA-binding, signal transduction-associated protein 3, ELAV-like protein 1
XR_003981066.1	RNA-binding motif protein, Y chromosome, family 1 member A1, Eukaryotic translation initiation factor 4B, Serine/arginine-rich splicing factor 9, RNA-binding protein VTS1, YTH domain-containing protein 1, Serine/arginine-rich splicing factor 10
XR_003984491.1	Y-box-binding protein 2-A, Muscleblind-like protein 1, RNA-binding motif protein, X chromosome, Serine/arginine-rich splicing factor 10, KH domain-containing, RNA-binding, signal transduction-associated protein 3
XR_003981080.1	Non-POU domain-containing octamer-binding protein, RNA-binding motif protein, X chromosome, KH domain-containing, RNA-binding, signal transduction-associated protein 3
XR_003984405.1	RNA-binding protein VTS1, YTH domain-containing protein 1, RNA-binding motif protein, X chromosome, ELAV-like protein 1, KH domain-containing, RNA-binding, signal transduction-associated protein 3
XR_003983750.1	Far upstream element-binding protein 2, Muscleblind-like protein 1, YTH domain-containing protein 1, KH domain-containing, RNA-binding, signal transduction-associated protein 3, ELAV-like protein 1
XR_003981604.1	Pumilio homolog 2, Serine/arginine-rich splicing factor 9, Muscleblind-like protein 1, YTH domain-containing protein 1, RNA-binding motif protein, X chromosome
XR_003986252.1	Polyadenylate-binding protein 1, Y-box-binding protein 1, RNA-binding motif protein, X chromosome

XR_003982440.1	Pumilio homolog 2, Far upstream element-binding protein 2, YTH domain-containing protein 1, RNA-binding motif protein, X chromosome, KH domain-containing, RNA-binding, signal transduction-associated protein 3
XR_003982247.1	Eukaryotic translation initiation factor 4B, Muscleblind-like protein 1, RNA-binding protein VTS1, YTH domain-containing protein 1, Serine/arginine-rich splicing factor 10, Serine/arginine-rich splicing factor 1
XR_003984464.1	Protein quaking, RNA-binding motif protein, X chromosome, ELAV-like protein 1, KH domain-containing, RNA-binding, signal transduction-associated protein 3
XR_003986536.1	Y-box-binding protein 2-A, Serine/arginine-rich splicing factor 9, Eukaryotic translation initiation factor 4B, Far upstream element-binding protein 2, RNA-binding motif protein, X chromosome, RNA-binding protein 4, Serine/arginine-rich splicing factor 1
XR_003986086.1	Pumilio homolog 2, Serine/arginine-rich splicing factor 9, Muscleblind-like protein 1, YTH domain-containing protein 1, Serine/arginine-rich splicing factor 1, ELAV-like protein 1
XR_003985850.1	RNA-binding protein FUS, Pumilio homolog 2, Serine/arginine-rich splicing factor 9, Eukaryotic translation initiation factor 4B, Far upstream element-binding protein 2, Serine/arginine-rich splicing factor 1, RNA-binding motif protein, X chromosome
XR_003986739.1	RNA-binding motif protein, Y chromosome, family 1 member A1, Pumilio homolog 2, Eukaryotic translation initiation factor 4B, RNA-binding protein VTS1, RNA-binding motif protein, X chromosome, Serine/arginine-rich splicing factor 1
Common between F-g v p, G-g v p and I-g v p	
XR_003981779.1	FUS RNA Binding Protein, Aconitase 1, KH RNA Binding Domain Containing, Signal Transduction Associated 3
XR_003981339.1	FUS RNA Binding Protein, Pumilio RNA Binding Family Member 2, Muscleblind Like Splicing Regulator 1, KH-Type Splicing Regulatory Protein, Serine And Arginine Rich Splicing Factor 10, Serine And Arginine Rich Splicing Factor 1, KH RNA Binding Domain Containing, Signal Transduction Associated 3, ELAV Like RNA Binding Protein 1
XR_003981340.1	FUS RNA Binding Protein, KH-Type Splicing Regulatory Protein, Muscleblind Like Splicing Regulator 1, Y-Box Binding Protein 1, RNA Binding Motif Protein 4, ELAV Like RNA Binding Protein 1, KH RNA Binding Domain Containing, Signal

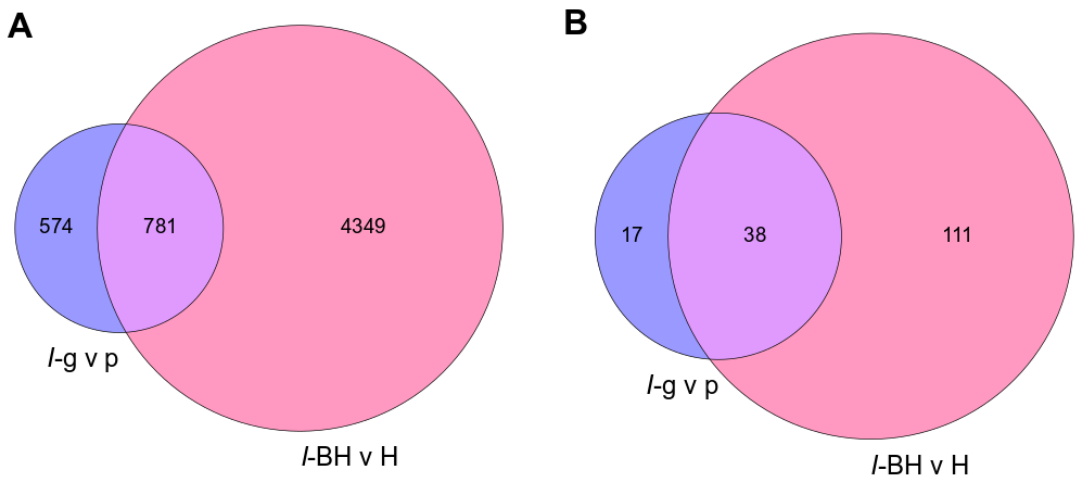


Figure 5.3. Proportional Venn diagram of total DETs (A) and DEncRNAs (B) in comparison between *I*-g v p and *I*-BH v H. g: Good quality; p: Poor quality; BH: Before hatch; H: Hatch; all DETs included in venn diagram were with p-value < 0.001.

5.4.3 Functional enrichment analysis of embryo transcripts

The GO enrichment analysis indicated that most DE genes were significantly enriched in metabolic, biosynthetic, and developmental related BPs (Supplementary Table 5.2-5.5 in annex I). A Venn diagram of g and p comparisons across the different hatcheries revealed that 120 enriched BPs overlapped between ***I*** and ***G*** hatcheries (Figure 5.4A). 25 BPs were common between the three hatcheries (***I***, ***G*** and ***F***) suggesting a large heterogeneity in the embryo transcriptome existed between each production site. Analysis of enriched BPs in embryos of the ***I*** hatchery (Figure 5.4B) revealed 144 common BPs between BH v H and g versus p suggesting that most BPs were related to development although 32 BPs were specifically linked to quality.

KEGG pathway analysis also revealed that low similarity existed between the quality-related DE genes from embryos of different hatcheries (***I***, ***G*** and ***F***, Figure 5.5). Four main KEGG pathways were significantly enriched and common in the quality-linked DE genes across the three hatcheries and included: common metabolic pathways, ribosome, spliceosome, and oxidative phosphorylation.

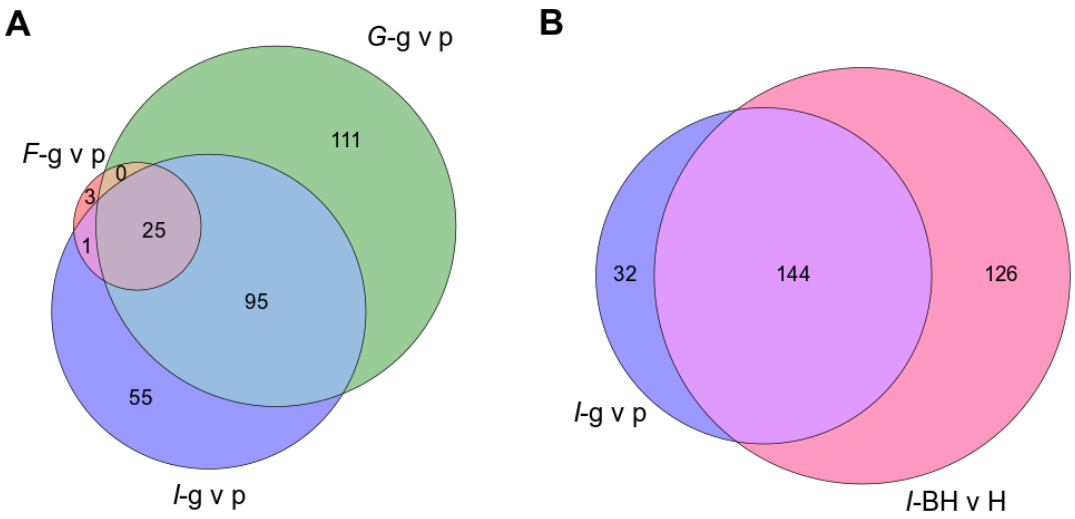


Figure 5.4. Proportional Venn diagram of GO-BPs in comparisons between embryo qualities from *F*, *G* and *I* regions (A) and between quality and stage from *I* region (B). Values within the circles represent the number of significant GO-BPs (adjusted p-value < 0.05).

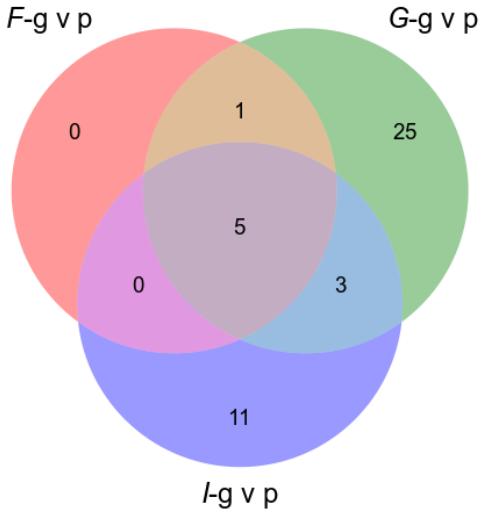


Figure 5.5. Venn diagram of KEGG pathways. Values represent the number of significant KEGG pathway (adjusted p-value < 0.05).

5.4.4 Embryo transcriptome and proteome integrated analysis

The DE genes from the embryo transcriptome and differentially expressed proteins available from sea bream proteome data [from chapter 4, data deposited to the ProteomeXchange Consortium via the PRIDE (www.ebi.ac.uk/pride, (Perez-Riverol *et al.* 2022) partner repository with the dataset identifier PXD033417] were compared. Comparison of quality-linked DE genes and the embryo proteome revealed that 543 proteins were unique to the proteome suggesting that they might correspond to maternal proteins (Figure 5.6;

Supplementary Table 5.6 annex I). Among them, ATPase H+ transporting V1 subunit Ab (*atp6v1ab*), heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 1 (*hsp90aa1.1*), clathrin, heavy chain b (Hc) (*cltcb*), ATPase, H+ transporting, lysosomal, V1 subunit B, member a (*atp6v1ba*), adaptor related protein complex 2 subunit alpha 1 (*ap2a1*), ATPase H+ transporting V1 subunit H (*atp6v1h*), and ATPase H+ transporting V1 subunit E1b (*atp6v1e1b*) were found to be DE proteins and these are potential immune-related proteins and may possess more crucial function during embryo development (Supplementary Table 5.6 annex I). GO analysis shows that multiple metabolic and catabolic processes were significantly enriched (Supplementary Table 5.7 annex I). It is noticeable that 543 maternal candidate proteins are involved in developmental process including eye-related retina development, as well as tissue and organ development, indicating the role of maternal regulation in the development of some organs during embryogenesis. In addition, pathways of different types of protein complexes such as Spliceosome, Proteasome, Ribosome were most significant, and based on the results of the proteoms are likely to be regulated by maternal signals.

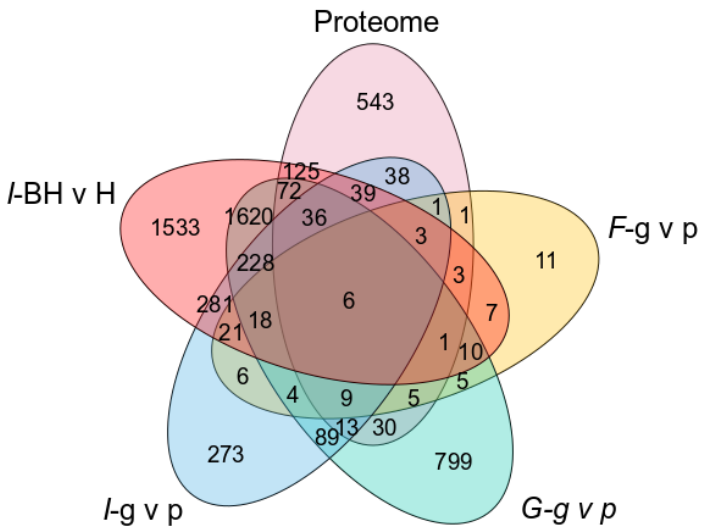


Figure 5.6. Venn diagram of transcriptomics and proteomics sea bream embryo samples. Values represent the number of significant transcripts (p-value < 0.001) and total quantified proteins.

5.5. Discussion

This study exploited high-throughput RNAseq to understand the molecular dynamics of gilthead sea bream embryo development and identified potential biomarkers related to embryonic quality. The number of large-scale RNAseq studies targeting fish embryogenesis are scarce and even fewer are available for teleosts used in aquaculture, and none exist for sea

bream. Herein we produced and analyzed eighteen gilthead sea bream cDNA libraries prepared from embryos with different quality characteristics and developmental stages. The samples analyzed were obtained from standard production protocols applied at different hatcheries in the Mediterranean and were classified as good or poor quality based on standard performance characteristics, viability, survival, hatching rate. Large differences in the expression profile of transcripts were identified for embryos coming from different hatcheries both in relation to protein coding and non-coding gene transcripts. Assessment of functional analysis enrichment corroborated the differences seen in relation to gene transcript expression and specific analysis of genes associated with immune-related processes revealed they were very low abundance/absent from the transcriptome of embryonic stages. Comparisons of the gene transcripts that were differentially expressed (DE) between embryos classified as good or poor quality and the sea bream embryo proteome permitted identification of candidate maternal transcripts. Targets for regulation by ncRNAs identified during embryo development and related to embryonic quality were identified.

The transcriptome of embryos from different hatcheries was very different and this meant that data analysis had to be carried out for the samples from each hatchery and a large variation in the number DE gene transcripts related to embryonic quality was found between hatcheries (***F***, 146; ***I***, 1355; ***G***, 4119). Comparison of the quality-linked DE genes in each hatchery identified a common set of core gene transcripts (n = 42) shared between the three sites and these were considered to represent universal gene transcripts linked to quality and putative candidates for embryonic quality monitoring in sea bream aquaculture production. The large differences in identified DE genes between the hatcheries that provided the samples is intriguing and revealed that rearing conditions and presumably broodstock have a significant effect on the changes in gene expression during development and revealed the complexity of trying to identify quality-linked gene transcripts. The present large-scale meta-analysis of embryos at a similar stage but from multiple production regimes may contribute to explain why the identification of genes that determine embryonic quality have been elusive. Here, it is worth noting that it was not yet possible to obtain environmental parameters including water quality and broodstock nursery regime due to commercial confidentiality issues and clearly it will be of interest to correlate such parameters with the transcriptome data. Nonetheless, since the hatcheries have a common procedure for determining embryonic stage and for quality analysis, which includes viability, floatability and hatching rate this provided a solid basis for the analysis and was coherent with the sample transcriptome-based clustering in the PCA analysis.

A previous transcriptome study on the European sea bass (*Dicentrarchus labrax*) (Class Eupercaria), which is phylogenetically proximate to the sea bream, identified 461 gene transcripts proposed to be potential biomarkers of egg and embryo quality (Zarski *et al.* 2017). In the rainbow trout (*Oncorhynchus mykiss)* (Class Salmoniformes) a much higher number of potential embryo quality markers were reported and varied between 900 to 1100 DE transcripts (Ma *et al.* 2019a). Validation studies are needed to demonstrate indisputably the predictive value of the identified quality-related genes in both species. Furthermore, the significance of the difference in the number of quality-linked DE genes between the sea bass (461) and rainbow trout (900 -1100) is difficult to establish particularly considering the variability in the number of quality-related DE genes in the present study of sea bream from different hatcheries. The variation in the number of gene transcripts associated with embryo quality between different teleost species and within the same species of different origin suggests that external environmental factors and the genetic background of the broodstock influences the embryo transcriptome as well as embryo quality and indicates the need for caution when assigning genes as fish embryo quality markers.

In the model organism, the zebrafish, which is a non-marine teleost, molecular studies of egg and embryo development exist (Mathavan *et al.* 2005; Vesterlund *et al.* 2011; Aanes *et al.* 2011), GO enrichments analysis between good- and poor-quality zebrafish embryos revealed that the top five most significantly represented processes were cellular process, metabolic process, biological regulation, response to stimulus, and cell component organization and biogenesis (Cheung *et al.* 2018a) and the KEGG pathways for protein synthesis (Cheung *et al.* 2019). Genes related to the immune system were also significantly enriched between zebrafish fertilized eggs of good and poor quality. The significantly enriched processes between good- and poor-quality fertilized eggs in an aquaculture species the rainbow trout differed from those found in zebrafish, and were ribosome biogenesis, translation, metabolic process, oxidation-reduction process, and regulation of transcription and DNA-template (Ma *et al.* 2019a). In the Atlantic cod (*Gadus morhua*) metabolism-related pathways such as fructose metabolism, fatty acid metabolism, glycerophospholipid metabolism, and oxidative phosphorylation were significantly enriched when fertilized eggs from farmed (normal quality) and wild broodstock (superior quality) were compared (Lanes *et al.* 2013). In the present study of the gilthead sea bream DE genes related to embryonic quality at different hatcheries in Italy, France and Greece fell into several different categories and the most significant were within metabolic and biosynthetic process pathways. Immune related genes/pathways were only identified by manual

scrutiny of the DE gene transcripts between good and poor-quality embryos and were restricted to specific hatcheries and included transcripts related to the MAPK signaling pathway and Salmonella infection pathway. The results obtained raise doubts about the link between immune-related gene transcripts and quality and raise questions about the hatchery-specific factors that stimulated their upregulation.

Non-coding RNAs are a specific category of RNA with non-protein coding potential that function by regulating gene expression and chromatin modifications (Alexander *et al.* 2010). Interestingly in our study several DE non-coding transcripts were found to be related to embryonic quality, suggesting they may regulate factors with a key role in embryogenesis. Analysis of the putative target genes of the quality related DE non-coding RNA revealed three that were common across all hatcheries and others that were shared between two hatcheries and others that were hatchery specific. Some of the putative target genes for regulation by the non-coding RNA included elav-like protein 1, Y-box-binding protein 1 and 2A (*yb-1* and *2a*) and Serine/arginine-rich splicing factor 9 (*srsf9*), all genes that are involved in zebrafish development (Li *et al.* 2014; Huang *et al.* 2021a; Joris *et al.* 2017; Kleene 2016; Sun *et al.* 2018). For example, two Elav1-like genes, *elavl1a* and *elavl1b*, have been described in zebrafish (Li *et al.* 2014) and elavl1a was mainly expressed during embryogenesis and is required for erythropoiesis (Li *et al.* 2014; Zhou *et al.* 2019a). The Elav1-like genes encode proteins that are abundant in eggs/embryos and are lipoteichoic acid (LTA)- and LPS-binding proteins and presumably have immune-related functions in the zebrafish (Ni *et al.* 2021). The genes products of *yb-1* and *srsf9* are important in maintaining zebrafish embryonic development, and *yb-2A* coordinates the storage of translationally repressed mRNAs in round spermatids of zebrafish (Huang *et al.* 2021a; Kleene 2016; Joris *et al.* 2017; Sun *et al.* 2018). However, the role of these proteins in marine aquaculture species remains to be established and also their involvement and mechanisms of action in embryonic quality.

It is known that the early stages of fish development are dependent on transcription of maternal RNA and on maternal proteins but their impact on sea bream development has not been described (Lubzens *et al.* 2017). In other teleost species the role of maternally inherited transcripts has been described and characterized and include some immune factors although this varies among the teleost species that have been studied. For example, in the rainbow trout unfertilized eggs maternally inherited transcripts were important in the cytoskeleton microtubules (*β-tubulin*) and is one of the mitogenic peptides that regulate vertebrate growth, insulin-like growth factor II (*igf2*) (Aegerter *et al.* 2005; Aegerter *et al.* 2004; Pierce *et al.* 2011).

In the Atlantic cod maternal transcripts *phb2* and *ddc* play an important role in development (Rise *et al.* 2014; Shih *et al.* 2013; Bonnet *et al.* 2007b) and in the Japanese eel (*Anguilla japonica*) maternal transcripts *dnajb4*, *gnpat*, *card14*, *pdp1*, *fcgbp*, *ttn*, were associated with embryogenesis while *b4galnt1*, *acsl6*, *rtkn*, and *trim24* were related to embryo hatchability (Izumi *et al.* 2019). At the protein level, ZRANB2 is a maternal immune factor in zebrafish that protects embryos/larvae against bacterial infections (Wang *et al.* 2016b) and another immune factor, lysozyme, was maternally inherited in coho salmon (*Oncorhynchus lusutch*) eggs (Yousif *et al.* 1994). In the present study by comparing the DEGs in the transcriptome with total quantified proteins in a proteomics study that compared BH and H embryos, putative inherited maternal proteins were identified in the gilthead sea bream embryos (chapter 4). This led to the identification of 543 potential maternal proteins, which included some that were immune-related (*atp6v1a*, *hsp90aa1.1*, *cltcb*, *atp6v1ba*, *ap2a1*, *atp6v1h*, *atp6v1e1b*), although no similarity with studies in other species were found. Further studies will be required to look in greater detail at the function and importance of these maternally inherited factors in determining embryonic quality.

The contribution of maternal factors to embryo development now includes a catalogue of factors such as transcripts that regulate cell division and determine oocyte polarity, embryo patterning, and the transition from maternal to zygotic gene expression (Lubzens *et al.* 2017). Previous studies have reported that different quality-related maternal transcripts were identified in Japanese eel (*Anguilla japonica*) and Atlantic salmon (*Salmo salar*) by comparative analysis of transcriptomes between fertilized egg samples classified as “good” and “poor” (Izumi *et al.* 2019; Bizuayehu *et al.* 2019). The identified candidate maternal mRNA transcripts in aquaculture species are promising and provide insight into important candidate genes linked to the quality of fish embryogenesis. Moreover, maternal RNAs have been shown to carry information about the environmental conditions experienced by the mother and in the Round Goby (*Neogobius melanostomus*) embryo at 32-cell stage the maternal RNA expression levels were correlated with the water temperature experienced by the mother before oviposition (Adrian-Kalchhauser *et al.* 2018). The temperature-responsive genes included core nucleosome components or the microtubule cytoskeleton, which are potentially relevant pathway for non-genetic inheritance. Another study on the association between fertilized egg size (prior 2-cell stage) and maternal RNA input in 15 different mouthbrooding cichlid fish species adapted to different trophic niches from different lakes revealed correlations between fertilized egg size and maternal mRNA abundance of two growth-related genes *igf2* and *ghr2*. Furthermore,

distinct clustering of these cichlids was found based on their trophic specialization using and maternal mRNA abundance of five genes (*ghr1*, *ghr2*, *igf2*, *gr* and *sgk1*), indicating that variations in fertilized egg size in evolutionary close cichlid species could be associated to differences in maternal RNA deposition of key growth-related genes (Ahi *et al.* 2018). Of note is the recent study of maternal RNA localization in oocytes that revealed poor correlation between the vegetally localized transcripts but a relatively good correlation between the animal pole localized transcripts in comparisons between amphibians (*Xenopus laevis* and *Ambystoma mexicanum*) and fishes (*Acipenser ruthenus* and *Danio rerio*) using RNA-Seq approach (Naraine *et al.* 2022). This suggests that maternal gene localization and species specificity appears at pre-fertilization stage and the regulation of embryonic development within the animal kingdom is highly diverse (Naraine *et al.* 2022).

5.6. Conclusion

In this study 18 sea bream embryo transcriptomes were produced and annotated and revealed that the number of DE genes between samples varied between hatcheries indicating that external and biological factors play a key role in the embryo transcriptome and embryogenesis. Despite the high variability of the transcriptome of embryos from different hatcheries, 42 common quality-related DE genes were shared between all hatcheries and may be good candidates for biomarkers of embryonic quality in aquaculture. Several maternally inherited transcripts were identified, as well as ncRNA, which were common in the DE genes of embryos from all hatcheries and may play a key role in the regulation of proteins crucial for fish embryogenesis. Several immune related genes were found highlighting the presence of protective immune milecules during embryogenesis and their potential as biomarkers to assess embryo performance. Our study provides for the first-time fundamental datasets of the sea bream embryo transcriptome that can be used as a guide for understanding the molecular basis of development and established potential molecular quality markers that can be applied in the future after validation in aquaculture hatcheries.

CHAPTER 6

General discussion and future perspectives

6.1 General discussion

The expansion of the World aquaculture industry is highly dependent on the production of high-quality fish embryos by hatcheries to guarantee successful larvae development to adulthood. Most fish release eggs into an aquatic environment, which is rich in microbes including pathogenic bacteria and viruses that can compromise embryo development (Wang *et al.* 2008). Fish offspring, as for other vertebrates, depend on maternally inherited nutritional and immune factors for survival during the early stages of development before independent feeding (Swain & Nayak 2009). Maternal immune protection is presumed to be important in early larval stages due to the delayed maturation of their immune system (Koumansvandiepen *et al.* 1994; Zapata *et al.* 2006), and the limited ability of fish embryos and larvae to produce specific antibodies until many weeks after hatching (Ellis 1988; Swain *et al.* 2002; Magnadottir *et al.* 2004). A study in zebrafish, a non-marine and non-aquaculture teleost fish, revealed that the only immune cells in the yolk sac at 24 hpf are macrophages (Ito *et al.* 2008). To date, IFN-γ (Sieger *et al.* 2009), LYZs (Li *et al.* 2021), hatching enzymes (Saleh *et al.* 2019), and Elavl1a were reported to serve as immune regulators in early life stages of fish (Ni *et al.* 2021). Nonetheless, even though most fish produce vulnerable, poorly developed larvae they are the most specious, diverse and successful of the vertebrates and 85% of all fish belong to the teleosts, which include most species of economic importance (Lasram *et al.* 2009).

Freshwater teleost fish species belonging to the carp (family Cyprinidae), tilapia (family Cichlidae) and catfishes (family Siluriformes) families are the most exploited fish worldwide for aquaculture (Cai & Galli 2021). In the Mediterranean area the European sea bass and gilthead sea bream (*Sparus aurata*) are the two most exploited marine aquaculture species (Morretti *et al.* 1999). Currently, further expansion of aquaculture in the Mediterranean is limited by sustainability issues, which is related in part to market demand and economics and biological challenges linked to the hatchery production stage, such as unpredictable spawning and highly variable egg and embryo quality, which affects production sustainability (Forniés *et al.* 2001; Lahnsteiner & Patarnello 2004b; Lahnsteiner & Patarnello 2004a). Studies of the basic biology of hatchery stages were common during the establishment of sea bream and sea bass aquaculture, but more recently relatively few studies have investigated embryonic development and quality markers in these and other marine aquaculture fish species (Table 1.5; Chapter 1) and the studies that exist tend to be species-specific.

This thesis aimed to identify and characterize molecular factors related to fish embryo development and quality performance for the sea bream and sea bass by exploring existing data

and generating large molecular datasets using omics approaches (proteomics and transcriptomics) coupled to expression analysis and functional assays. Samples from different commercial hatcheries across Europe and egg/embryo batches with different quality performance were used as a source of quality markers for hatchery production of these fish species. The main achievements reported in this thesis are shown in Figure 6.1. From the study of lysozyme (Chapter 2), bioinformatics analysis, enzymatic assay and gene expression analysis confirmed that lysozyme is present in embryos and the activity may vary between species. Importantly, we found the differential expression of *lyg2* and *lalba* genes between different quality embryos, providing reliable quality markers at least for classifying different quality embryos of sea bream. Then we focused on C5 as another factor of the humoral innate immune response to understand the biological significance of this molecule in fish (Chapter 3). The use of bioinformatics for evolutionary, biological process, functional enrichment and structural analysis were conducted and a C5, C5R (C5 receptor, CD88) were identified among different fishes and the identification of C5R in the genome of only some species suggested it evolved in a species-specific manner. A further aspect of this work was the successful construction of the structure of C5a and its receptor by homology modeling. Molecular docking studies with the established models identified critical docking site for theoretical support of the interactions between C5 and C5R in the fish immune system. To gain knowledge about changes in the spatio-temporal use of regulatory molecules, as well as changes in the specificity of regulatory proteins correlated with biological differences between phylogenetically distant species during fish embryonic development, total protein was extracted from fertilized egg samples at different embryonic stages from white sea bream (*Diplodus sargus*), meagre (*Argyrosomus regius*) and gilthead sea bream and used to generate proteome libraries. The proteome was established for the three species under study and a large number of proteins were found to be significantly modified between yolked embryos before hatch (BH) and at hatch (H). Of the enriched biological processes between the two stages compared (BH *vs* H), two endocrine-related processes were enriched indicating a possible activation and involvement of the endocrine system in the regulation of embryo development. Moreover, six immune-related proteins (Sodium potassium-transporting atpase subunit alpha-2-like (atp1a2a), sodium potassium-transporting atpase alpha-1 subunit (atp1a1), heat shock protein hsp 90-alpha 1-like (hsp90aa1.1), heat shock cognate 70 (hsp70), ap-2 complex subunit alpha-2-like isoform 1 (ap2a1), cathepsin d (ctsd)) were found to be significantly modified at a protein level between the comparison of BH and H in all three fish species. Gene expression analysis was conducted

by qRT-PCR on *Hatching enzyme* (*He*) and combine to the protein level from proteome and expression profile by qRT-PCR suggests that a characteristic of species-specific and stage-specific of He and He2 may serve as the functionally important enzyme driving the hatching event in the fish embryo. As an extension, transcriptomics was used to study different quality sea bream yolked embryos obtained from different production sites om Europe (Chapter 5). Using site specific transcriptomics to compare good and bad quality embryos and then comparing the differentially expressed genes (DEGs) obtained for each site, 42 common embryo associated DEGs were identified. Proteomics analysis identified 543 total proteins in the yolked embryo proteome between before hatch and hatched embryos (Chapter 4). A number of putative immune-related DE proteins most likely of maternal origin were identified, *atp6v1ab*, *cytosolic*, *hsp90aa1.1*, *cltcb*, *atp6v1ba*, *aap2a1*, *atp6v1h*, and *atp6v1e1b*, suggesting maternal transfer of immune protection to the embryos seems to occur.

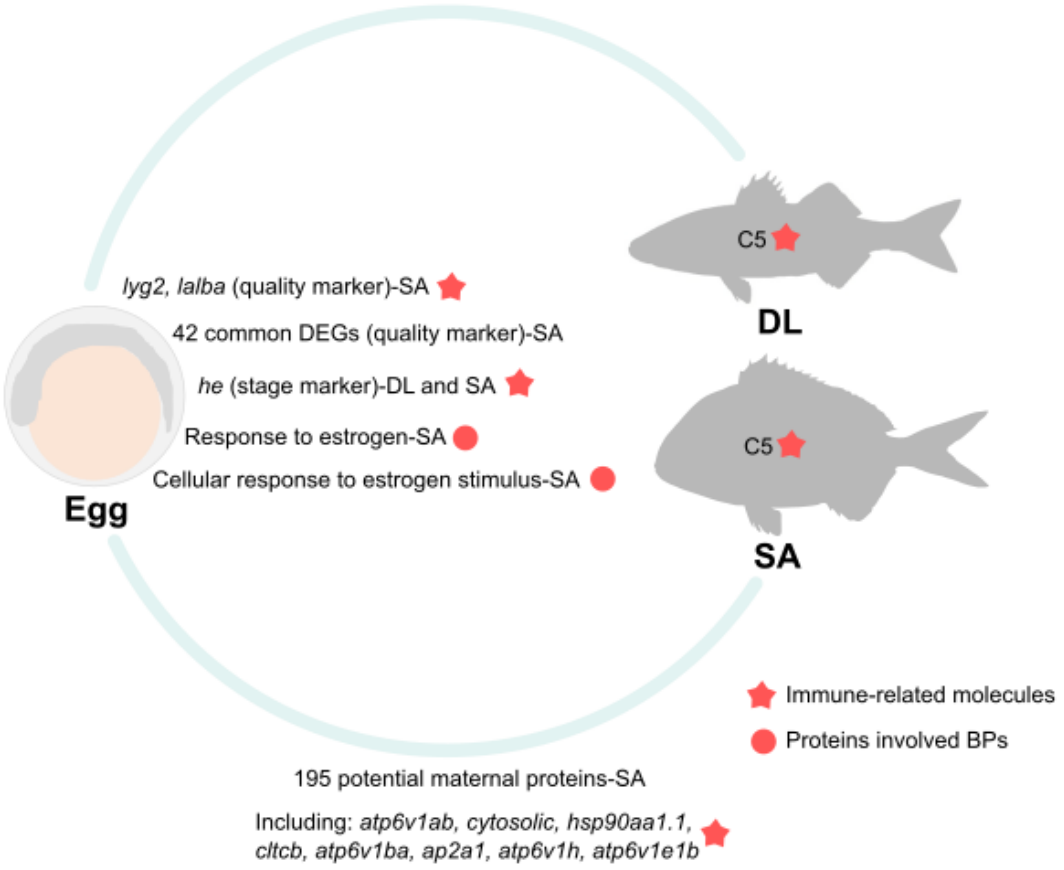


Figure 6.1. **Schematic diagram summarizing putative quality markers obtained from the thesis work. DL-European sea bass; SA-gilthead sea bream.** DEGs: Differential expressed genes; BPs: Biological processes. Lysozyme g2 (*lyg2*), lactalbumin (*lalba*), hatching enzyme (*he*), ATPase H+ transporting V1 subunit Ab (*atp6v1ab*), heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 1 (*hsp90aa1.1*), clathrin, heavy chain b (Hc) (*cltcb*), ATPase, H+ transporting, lysosomal, V1 subunit B, member a (*atp6v1ba*), adaptor related protein complex 2 subunit alpha 1 (*ap2a1*), ATPase H+ transporting V1 subunit H (*atp6v1h*), and ATPase H+ transporting V1 subunit E1b (*atp6v1e1b*), Complement C5 (*C5*).

The main outcomes of the thesis were described in the four studies reported in the four results chapters (2 – 5) of this thesis. At the start of my thesis, the understanding of immune factors and their function during fish development was relatively limited, and the studies reported in this thesis contributed to increase knowledge about early stages of fish and immunity. The first two chapters of the thesis were focused on the characterization of individual gene families that have been linked to fish innate immunity. The last two chapters of the thesis were dedicated to omics analysis (proteome and transcriptome) to characterize the molecular dynamics related to sea bream embryo development (proteins and gene transcripts). The first results chapter (Chapter 2) described the characterization of a class of innate immune humoral factors, the lysozymes, in fish and provide a thorough analysis of their presence and species-specific divergent evolution and function focusing on marine aquaculture teleosts. Chapter 3 described the evolution and functional classification of complement system proteins focusing on C5 and its receptor C5AR (CD88), this was carried out in the carp and the study contributes to extend understanding of this system in teleosts. This work was part of a contingency plan implemented for covid 19, when I was detained in China for over 1 year. Subsequent studies exploited high throughput approaches to generate a global overview of the fish embryo (yolked and unyolked) proteome and transcriptome. In chapter 4 a highly precise proteome was established for white sea bream, meagre and gilthead sea bream embryos at two developmental stages (before hatch and hatch) and it was used to characterize conserved and species specific processes and to identify potential immune related molecules in these early stages. Some isoforms of the hatching enzyme family (developmental biomarker factors) were characterized and although no evidence for an immune function was found they represent potentially interesting developmental and quality biomarkers for the future if expression levels can be linked to hatching success. The gilthead sea bream embryonic transcriptome was reported in chapter 5 for embryos at two different developmental stages (before hatch and hatch) and for embryos batches with different quality performances. Several immune genes of maternal origin and embryonic ncRNAs and their potential protein targets were reported and potential insight into a new class of immune regulators was obtained, and other potential markers of developmental quality were proposed.

6.1.1 Fish innate immune genes as quality markers of embryo development

Fish are prone to diseases as they live in a microbial rich water environment and

pathogenic microbes can impair normal development and compromise egg, larvae and juvenile fish performance. The lysozymes are a large group of immune enzymes that play a key role in host protection and defense against bacterial invasion. Despite their crucial importance in immunity relatively little information is available about lysozyme characteristics, evolution and function in teleosts and most studies available are focused on mammals (Callewaert & Michiels 2010; Irwin *et al.* 2011). The lysozymes are a group of enzymes that are classified within the humoral branch of the non-specific innate immune response. The presence of lysozymes in fish embryos means their activity was studied during embryo and larval development in several teleost fish (Table 6.1) and they are proposed to have anti-bacterial activity since they break down bacterial cell walls (Saurabh & Sahoo 2008) and are involved in the activation of the embryonic fish immune system. Several studies have reported that the enzyme activity identified in early developmental stages corresponds to lysozyme c and g in several different fish species and linked the function to fish innate immunity and pathogenic defense (see table 1). But studies characterizing the lysozyme gene family in fish are limited and gene number, distribution and activity remain largely underreported especially during early developmental stage when species specific expression and activity may exist. The lysozyme gene family was characterized in depth during my thesis since it is an important enzyme of innate immunity in early stages of fish development. The characteristics of the gene family and gene transcription and activity in early stages of sea bream and sea bass was developed to contribute to a better understanding of their role across different fish species. Extensive bioinformatic analysis was performed and members of the fish lysozyme gene family was identified and characterized. In teleosts two major lysozyme gene families were identified and within each family the number of members was found to be distinct across species suggesting that the evolution and persistence of the lysozyme genes in teleost genomes was species-specific. This suggests that different fish species have acquired a unique capacity to respond to bacteria, although the reason for the divergent evolutionary trajectories in fish was not established. In contrast to previous studies of lysozymes in fish that only reported lyz c and g, during my thesis work a novel fish lysozyme member, lalba-like, was identified for the first time in fish and has previously only been described in mammals where it is related to milk production (Layman *et al.* 2018). The existence and persistence of this gene in a restricted group of fish species raises interesting question about its function in the immune response or other processes and further studies will be needed in the future to identify its function (Chapter 2). In relation to the ambition of establishing quality markers of embryonic quality, insufficient evidence was gathered in the

work on lysozymes or on the proteome and transcriptome studies to indicate the utility of this enzyme in the assessment of embryonic quality. Nonetheless, an intriguing observation was that the expression of *lalba* and *lyg2* gene differed between embryos of good and poor quality in the gilthead sea bream. The functional significance of the expression pattern of *lalba* and *lyg2* in embryos of gilthead sea bream was not established during the thesis work but the results indicate that further work on lysozymes may validate lysozyme as a potential quality biomarker for fish embryos (Chapter 2).

Table 6.1. Studies of lysozymes during early fish development.

Species	Main observations	Authors
Plaice (*Pleuronectes platessa*)	Protein location in tissues	(Murray & Fletcher 1976)
European sea bass *Dicentrarchus labrax*	Lysozyme activity in embryos and larvae	(Cecchini *et al.* 2000)
Gilthead sea bream (*Sparus aurata*)	Enzyme activity in eggs and larvae	(Hanif *et al.* 2004)
Zebrafish (*Danio rerio*)	Enzyme activity and protein level in fertilized eggs	(Wang & Zhang 2010)
Zebrafish (*Danio rerio*)	Gene expression in larvae	(Oyarbide *et al.* 2012)
Caspian kutum (*Rutilus frisii kutum*)	Enzyme activity in fertilized eggs and larvae	(Abdollahi *et al.* 2016)
Persian Sturgeon (*Acipenser persicus*) and Sterlet (*Acipenser ruthenus*)	Enzyme activity in unfertilized and fertilized eggs and larvae	(Zamaninia *et al.* 2017)
Siberian sturgeon (*Acipenser baerii*)	Enzyme activity in unfertilized and fertilized eggs and larvae	(Valipour *et al.* 2018)
Turbot (*Scophthalmus maximus*)	Enzyme activity in unfertilized eggs	(Jiang *et al.* 2019)

Another important component of the humoral response of innate immunity is the complement (C) system. This is a group of proteins that are activated in a cascade and cause lysis after binding to the surface of microorganisms (Janeway *et al.* 2001). Complement 5 (C5) was included as part of the thesis (during retention for over a year in China due to covid 19). In this study C5 genes were identified in different fish species and were characterized and preliminary information about protein-receptor interactions was established using structural modelling. C5 is a member of the C3 family (Najafpour *et al.* 2020) and has two biologically active subunits C5a and C5b, and the former subunit is a chemoattractant (C5a) that stimulates myeloid cells (e.g. neutrophils, eosinophils) and the second subunit, C5b is crucial for the assembly of the membrane attack complex (MAC) that causes lysis. A brief consideration of

the studies that exist about C5 in early stages of fish (Table 6.2) reveals a limited understanding of this system in fish from different water environments. Since aquatic environments have a highly variable content of microbes a better understanding of the complement system in early stages may provide insight about evolutionary adaptations acquired by different species of fish to reduce the risk from the environment in early developmental stages. Hence, in chapter 3, C5 and its receptor C5R was analyzed from a structure-function perspective to extend the insight from gene transcript to protein function. Insights about systematic functions were obtained by comparing C5 and C5R in two Cyprinid fish (grass carp and zebrafish) and human (Chapter 3).

In teleost fish a single C5 gene was identified however in common carp duplicates exist as this species is allotetraploid due to a recent whole genome duplication (Xu *et al.* 2019b). Multiple C5R (CD88) were found in the genomes of some species suggesting that the receptor gene underwent species-specific evolution and that the regulation of the C5 system may be distinct across different teleosts. Based on structural similarity it seems likely that in teleosts C5 shares similar immune functions to the human homologue as revealed by PPI (protein-protein interaction) network analysis. Homology modeling and molecular docking of C5a-CD88 revealed that 290Ala residue from this downstream signaling molecule, will interact with 72Val from C5a and this represents an important site for the assembly of the protein-receptor complex. Intriguingly, integrating analysis from the yolked embryo transcriptome data (Chapter 5) revealed a significantly different gene expression pattern (p-value < 0.05) for sea bream C5 in fertilized egg batches classified as good compared to those classified as poor and between embryos before hatch (BH) and at hatch (H). Furthermore, the variation of C5 with embryo batch was not affected by the hatchery of origin, which suggests that changes in management practice or broodstock do not influence gene expression, suggesting that this gene and its protein product are likely to be important in early immune protection and possibly yet undescribed other functions in the gilthead sea bream and probably other fish species.

Table 6.2. Previously published studies about C5 during early developmental stages in teleost fish.

Species	**Main observations**	**Authors**
Rainbow trout (*Oncorhynchus mykiss*)	Gene expression in embryos and larvae	(Lvoll *et al.* 2006; Castro *et al.* 2015)
Atlantic halibut (*Hippoglossus hippoglossus*)	Gene expression in embryos	(Mommens *et al.* 2014)
Orange-spotted grouper (*Epinephelus coioides*)	Gene expression in larvae	(Wang *et al.* 2014a)
Zebrafish (*Danio rerio*)	Gene knockout and expression in larvae	(Swaminathan *et al.* 2022)

6.1.2 Proteome related to embryo development and hatching

Fish eggs are rich in proteins, but protein levels vary considerably between species (Bekhit *et al.* 2009; Kowalska-Goralska *et al.* 2020; Mol & Turan 2008; Machado *et al.* 2016; Inanli *et al.* 2019; Ozturk *et al.* 2019). In Chapter 4, the yolked embryo proteomes of 3 commercially important teleost species used for aquaculture were produced and comparative analysis revealed that white sea bream embryos are the most protein rich and that 694 proteins were common in the proteome of all three species analyzed. There were 104 differentially expressed proteins between yolked embryos collected 24 hours before hatch and at hatch and based on their gene ontology classification they are most likely to be involved in developmental, metabolic and regulatory biological processes and pathways. Taking into consideration the specific interest in characterizing endocrine and immune related processes during development the proteome of the different fish species was analyzed using GO analysis to identify enriched processes. Only two endocrine-related pathways (response to estrogen and cellular response to estrogen stimulus) and no immune-related pathways were found. This contrasts to previous reports applying proteomics to embryogenesis in other teleosts such as zebrafish (*Danio rerio*) (Shaik *et al.* 2014; Yilmaz *et al.* 2017), rainbow trout (*Oncorhynchus mykiss*) (Niksirat *et al.* 2020), common carp (*Cyprinus carpio*) and grass carp (*Ctenopharyngodon idellus*) (Wang *et al.* 2022) where immune related biological processes were found to be significantly enriched when protein abundance between different developmental stages were considered (e.g. high and low abundant) or when comparisons were made between quality, stage or species. However, relatively few studies have carried out comparative proteomics with the view of linking the proteome to quality, stage, and species. In a previous study of deyolked embryos 24 h post-fertilization in the zebrafish, 8363 proteins were identified (Shaik *et al.* 2014). Another study of the same species showed that the number of extracted proteins from non-deyolked embryos at different developmental stages was generally lower than in deyolked embryos (approx. 2700) at the same stage and this was assigned in part to the masking effect of abundant yolk proteins on other cellular proteins (Purushothaman *et al.* 2019). Comparison of previous proteomics studies with our proteomics study suggested the outcome was highly accurate and the abundant proteins identified were mapped to metabolic process, cellular component organization in common with the observation in previous studies of the zebrafish. A comparison of the five top common biological processes in the yolked embryos of the three fish species studied in our study were protein and amine related processes. These results suggest that the developmental difference between different fish is reflected in the dynamic patterns of protein expression and

activity. Moreover, common DEPs and biological processes obtained from the three species in our study is indicative of the general conservation in their developmental ontogeny and their evolutionary proximity. Furthermore, the use of a highly accurate proteomic sequencing platform, the experimental design and the care taken to ensure samples of a similar developmental status were collected contributed to guarantee the quality and comparability of the data generated for the 3 species studied.

In zebrafish embryo development the maturation of the immune system occurred after 24 hpf and a group of immune genes (*eomesb*, *ikzf1*, *rag1*, *gata3*, *lck* and *trac*) started to be expressed concomitant with the development of the immune organs (Ito *et al.* 2008). None of these proteins were differentially expressed in the previously reported proteomics studies, which may be indicative of differences in developmental progression between different fish species or may be due to differences in the experimental design, analytical approach, and analysis. RT-PCR analysis in the previous studies of zebrafish was used to assess the expression of the specific immune-related genes identified during embryonic development. In general, further studies are required in aquaculture species to build on the quantitative proteomics study presented in my thesis and could include RT-PCR, Quantitative Western Blot and functional studies of the specific immune-related proteins identified during fish embryonic development.

The fish Hatching enzymes (He) protein was used to validate the robustness of embryonic proteome produced and to provide a novel perspective on hatching in fish as this protein was differentially expressed between different embryonic stages. The He are a group of enzymes that accumulate in the egg and embryo prior to hatching and are secreted by the embryo into the perivitelline space and their function is to disrupt by proteolysis the egg envelop to facilitate the release of the embryo from the egg (Nagasawa *et al.* 2015). The proteome analysis of the gilthead sea bream identified three protein isoforms of He that are encoded by three individual genes in the genome and have a differing expression pattern in the yolked embryo proteome suggesting functional divergence during embryo development and hatch. The outcome of the proteomics study is summarized in Figure 6.2. Linking into the data from chapter 2 of the thesis it was observed that the abundance of one of the hatching enzymes, He2, was significantly decreased between yolked embryo samples before hatch (BH) and at hatch (H) in the white sea bream and gilthead sea bream, suggesting a possible role for this isoform of the protein in driving hatching in these two species. In contrast, in the meagre, which had a proteome more divergent than the breams, He2 was not significantly modified from BH to H, revealing species-specific differences. A general consideration of development as reflected by the proteome

revealed many biological processes including fin generation, muscle development, eye and visual system development and pigmentation was activated and were related to morphological alterations reported in Eupercaria during embryo development. Of note was the identification of regulatory factors associated with embryonic development and growth (Figure 6.2), and this revealed the importance of the crucial maternal cargo of proteins that support embryo development. The proteome between high quality and poor-quality embryos was not characterized as was initially foreseen in the thesis workplan and it will be important in the future to conduct this analysis as it may provide insight into the biological basis of embryonic quality and potentially yield putative quality biomarkers.

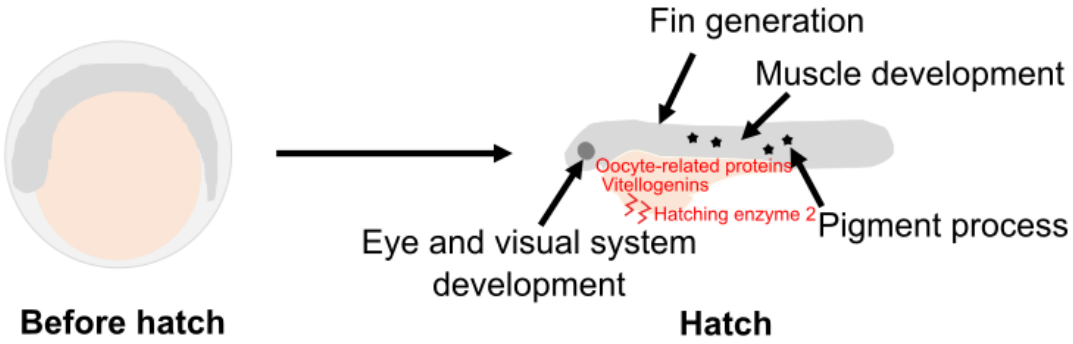


Figure 6.2. Schematic representation of the putative functional role of some of the DE proteins in fish in the present study. Terms marked in black represent significantly enriched GO biological processes and red text represents the functional DE proteins involved in the yolk; the stars indicate pigmentation deposited on the body surface.

6.1.3 Molecular dynamics of the gilthead sea bream embryo development

Transcriptomes of yolked embryos of different quality and at developmental stages were developed. This represents the first study using field samples obtained from different hatcheries and revealed highly variable transcriptomes and differentially expressed genes (DEGs) from sea bream larvae from different sampling sites (Chapter 5). Forty-two common quality marker candidates were found between good and poor-quality embryos from multiple hatcheries. The transcriptome analysis provides for the first time a comprehensive overview of the common molecular events that regulate embryo development across different production sites irrespective of broodstock and management practices. Furthermore, the common quality-linked markers are good candidates, if validated, for screening embryonic quality in sea bream. Only two immune related pathways (MAPK signaling pathway from A and Salmonella infection pathway constituted by 7 immune-related genes including *Rac family small GTPase 1a* (*rac1a*),

actin related protein 2/3 complex, subunit 1B (*arpc1b*), *caspase 3, apoptosis-related cysteine peptidase b* (*casp3b*), *cytochrome c, somatic b* (*cycsb*), *filamin C, gamma b* (*actin binding protein 280*; *flncb*), *heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 2* (*hsp90aa1.2*), *toll-like receptor 18* (*tlr18*) of total 19 enriched DEGs) were enriched in the DEGs, highlighting their importance in quality assessment. Several non-coding RNAs (ncRNA), which are likely to regulate gene expression during development were identified in the transcriptome. Prediction of the protein targets of ncRNA revealed ELAV-like protein 1, Serine/arginine-rich splicing factor 1, KH domain-containing, RNA-binding, signal transduction-associated protein 3, YTH domain-containing protein 1, and RNA-binding motif protein X chromosome as putative targets for embryo quality assessment. The differential expression of ncRNAs detected in transcriptomics, suggests they may play an essential role in the regulation of gene transcription during embryo development and that this may be closely associated with quality. Clearly further studies are required in the future if the role of ncRNA in development and embryonic quality are to be defined.

Of high interest were the 543 maternally inherited proteins identified in the yolked embryo proteome, which included some immune-related regulators. Previous studies have revealed the onset early during development of several immune molecules of maternal origin such as *irf7* in cod (*Gadus morhua*) embryos (Rise *et al.* 2012), *mhc2a* in Atlantic halibut (*Hippoglossus hippoglossus*) embryos (Mommens *et al.* 2014), and complement (*C3, C4, C5, C7*) in rainbow trout (*Oncorhynchus mykiss*) (Lvoll *et al.* 2006). It has been proposed that these genes with an early onset of expression during ontogeny may contribute to the zygote immune barrier during fish embryo development. It is unclear if the diversity of immune molecules that have been identified in different species is a consequence of technical factors linked to the techniques and methodologies used or if they represent biological differences. Our results provide support for the importance of the inherited maternal factors especially of those related to the immune system and the role of lncRNA in the regulation of embryonic development. Much more work is still required to establish the biological significance of the differentially expressed genes that were identified, particularly those related to the immune system and to validate them as putative quality markers.

6.2 Conclusion and future perspectives

6.2.1 Conclusion

This thesis provides for the first time a comprehensive description and characterization of immune and other molecular factors which govern the development of several economically important marine aquaculture fish. We found that the members of the two immune families studied using a candidate gene approach are distinct between the different teleost species compared suggesting that the importance and regulation of different immune genes and their function across fish may be species-specific, particularly in early developmental stages. We propose that lysozyme gene function and expression vary between different quality embryos and thus may be a quality marker for assessing embryonic quality, although the biological basis for this observation was not established and applicability of this observation across species needs to be assessed. Fish eggs are an important protein source for the developing embryo, but the protein content is different between embryos at the same stages across different species although many common proteins are also found. Two endocrine-related processes were found to be significantly enriched in three different fish yolked embryo proteomes indicating that the proteins involved in these two processes play an important role during fish embryo development. The discovery of hatching enzyme proteins and their functional divergence across fish fertilized eggs suggests the isoforms responsible for embryo hatching in different species and the timeframe of their activity differs. A large-scale yolked embryo transcriptome identified a core of conserved genes potentially related to embryonic quality, some of which were associated with immune-related processes. Overall, we found that the molecular dynamics of gene persistence and expression in sea bream embryos across different developmental stages and sampling sites had both conserved and divergent characteristics and that the candidate markers identified contribute to understanding the biological basis of quality in sea bream and in this way to achieving the goal of high-yield in aquaculture production. It should be noted, based on the outcome of the comparative analysis performed in the present thesis, from both a candidate gene and global perspective, the results obtained in relation to quality markers need to be applied with caution in other marine fish due to the existence of species-specific gene patterns linked to their biology and physiology and even for the same species there may be variation between production sites due to differences in broodstock, environmental conditions and the management regime.

6.1.2 Future perspectives

This thesis provides a general overview of the potential use of biomarkers to assess fish embryo quality for aquaculture hatcheries. To look for potential markers of embryonic quality transcriptome and proteome data was acquired. However, currently developmental omics data is scarce and most studies of embryonic stages use zebrafish (a non-aquaculture and non-marine teleost) as the reference. In the future it will be essential to characterize the dynamics of molecular expression patterns during the developmental ontology of aquaculture species and look for common quality markers that are effective irrespective of species, environment or management regime. This is important if sustainable and constant high quality aquaculture products are to be produced.

Through proteome and transcriptome analysis, we gained an understanding of transcription and translation processes during the embryonic development of the gilthead sea bream. The progress achieved through the work reported in the present thesis was to establish conserved molecular patterns during development across a diversity of hatcheries providing insight into dominant molecular processes that are strongly programmed and are minimally influenced by differences in the environment and drive development and immunity around the onset of embryonic development. It was not possible to map these molecules in the embryo although maternal molecules were defined and in future studies, egg inner structure and their developmental status should be integrated with multi-omics to unveil the origin of the molecular signals or cell populations during embryo development.

This thesis discovered a wide heterogeneity of genes expressed in sea bream fertilized egg batches obtained from different production sites and this implies that gene expression was influenced by different geographical conditions due to environmental factors acting on the genome and transcriptome of different egg batches. The differences in the genetic background of the egg batches from different hatcheries across the Mediterranean should not be overlooked when fertilized eggs are shared between different production sites due to the potential risk of genotype - environment interactions, which means eggs from broodstock selected under given environmental conditions may not perform well in another environment and this may influence the identification of effective quality indicators and may also limit the effectiveness of a single Mediterranean wide breeding program. In connection with this, Mediterranean fish aquaculture hatcheries should consider such factors and future research should be carried out into this issue and include studies of fish physiology particularly when it influences or determines commercially interesting traits.

Although the current thesis identified a series of putative maternally inherited molecules and embryo quality markers, accurate and comprehensive criterion for assessing the quality of embryo batches and more accessible sensitive and specific quality markers are still absent. This will require further investigation as well as extensive validation of the putative makers identified in the present thesis to obtain reliable criterion, not only for within species comparisons but also for application in species with a much greater evolutionary distant. For quality markers to be effective assessment methods need to be simple and easy to perform either by production companies with analytical laboratories or service provider, ideally, they should be applicable and reliable indicators of quality early in development to avoid occupying hatchery facilities with what may turn out to be unproductive batches of fertilized eggs. Although non-viable and extremally low-quality gametes and zygotes can be identified easily by buoyancy, appearance, or motility in some species, effective markers for accurately identifying the quality of fish gamete or zygote at a general level remain limited. Noticeably, survival at a specific embryonic stage in a stable manner can be monitored to characterize the ability of the fertilized egg to develop successfully in most fish species. From another perspective, a thorough analysis of developmental defects/failure can also be a means for unveiling the cause of poor developmental quality, but also indicate quality although such analysis tends to be time consuming (Migaud *et al.* 2013).

Many aquaculture hatcheries still rely on broodstock or seed harvested from wild populations for some farmed fish species since closure of the production cycle has not been achieved. This emphasizes that for some species there is still the need for comprehensive investigation of non-model commercial fish species to understand their biological characteristics and the molecular basis of the characteristics of each developmental stage. In addition, to classical breeding methods, a new era of genome editing has emerged with Clustered Regularly Interspaced Short Palindromic Repeats - associated protein 9 (CRISPR-Cas 9) and the availability of molecular markers (genes and proteins) may be an opportunity to accelerate selection for favourable traits linked to egg and embryo quality of commercial fish species. Gene editing is a promising approach that could be exploited to artificially create fish strains with desirable traits in the future, although there are many ethical issues linked to genome editing that need to be considered and debated within the public domain. However, before gene editing for commercial purposes it is essential to fully understand the function of the molecules and markers on phenotype variations of fish. Furthermore, an added complexity in aquaculture is the impact of environmental factors on phenotype since phenotypic plasticity

is common in cold-blooded fish species. Hence, understanding the interaction between a gene and the environment is important and stringent environmental control would be necessary if the benefits of gene editing are to be exploited for aquaculture. Single Nucleotide Polymorphism (SNP) may appropriately be used in this case for marker screening preliminarily, thereby efficiently targeting molecular markers and eventually accelerating fish re-domestication process.

Currently, the genome of European sea bass and gilthead sea bream are available, but the genome is not available for many other important aquaculture species such as the white sea bream and meagre. Information specific to each species is essential to understand their physiology since many gene families have evolved in a species-specific manner and the use of transversal molecular markers linked to quality may not be adequate. Hence, performing more genome sequencing in the future will be a way to establish if species-specific genes evolved in different ways, and will identify novel candidates for the study of embryo development.

Additionally, the role of the putative markers linked to quality in embryo development and embryo immunity remains to be explored, especially to establish if their action is direct or indirect and their mode of action in biological processes they regulate. Positional mapping of the molecular markers in the developing embryo will provide insight into the organs and tissues they regulate across a developmental time scale. Hence, future studies are required to uncover crucial and unexplored molecular functions that influence developmental ontogeny and egg and embryo quality of aquaculture species.

Bibliography

Aanes H, Winata CL, Lin CH, Chen JQP, Srinivasan KG, Lee SGP, Lim AYM, Hajan HS, Collas P, Bourque G, Gong ZY, Korzh V, Alestrom P, Mathavan S (2011) Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res,* **21**, 1328-1338.

Abdollahi R, Heidari B, Aghamaali M (2016) Evaluation of lysozyme, complement C3, and total protein in different developmental stages of Caspian kutum (*Rutilus frisii kutum* K.). *Arch Pol Fish,* **24**, 15-22.

Abrams EW, Mullins MC (2009) Early zebrafish development: It's in the maternal genes. *Curr Opin Genet Dev,* **19**, 396-403.

Adebayo IA (2018) Effect of Photoperiod on Eggs Hatchability, Growth and Survivability of Hybrid Catfish (*Heterobranchus bidorsalis X Clarias gariepinus*) Larvae. *Int J Fish Sci Res,* **2**, 1004.

Adlandsvik B, Coombs S, Sundby S, Temple G (2001) Buoyancy and vertical distribution of eggs and larvae of blue whiting (*Micromesistius poutassou*): observations and modelling. *Fish Res,* **50**, 59-72.

Adrian-Kalchhauser I, Walser JC, Schwaiger M, Burkhardt-Holm P (2018) RNA sequencing of early round goby embryos reveals that maternal experiences can shape the maternal RNA contribution in a wild vertebrate. *Bmc Evol Biol,* **18**, 1-14.

Aegerter S, Jalabert B (2004) Effects of post-ovulatory oocyte ageing and temperature on egg quality and on the occurrence of triploid fry in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture,* **231**, 59-71.

Aegerter S, Jalabert B, Bobe J (2004) Messenger RNA stockpile of cyclin B, insulin-like growth factor I, insulin-like growth factor II, insulin-like growth factor receptor Ib, and p53 in the rainbow trout oocyte in relation with developmental competence. *Mol Reprod Dev,* **67**, 127-135.

Aegerter S, Jalabert B, Bobe J (2005) Large scale real-time PCR analysis of mRNA abundance in rainbow trout eggs in relationship with egg quality and post-ovulatory ageing. *Mol Reprod Dev,* **72**, 377-385.

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res,* **46**, W537-W544.

Ahi EP, Singh P, Lecaudey LA, Gessl W, Sturmbauer C (2018) Maternal mRNA input of growth and stress-response-related genes in cichlids in relation to egg size and trophic specialization. *Evodevo,* **9**, 1-17.

Akarte SR, Mudgal MB (2017) Biochemical composition of fish eggs from local water reservoirs around Amravati city. *Biosci Biotechno Res,* **10**, 49-52.

Akdogan-Ozdilek B, Duval KL, Goll MG (2020) Chromatin dynamics at the maternal to zygotic transition: recent advances from the zebrafish model. *F1000Research,* **9**, 299.

Al-Bachry W (2018) The Effect of Using Low Fat Milk To Remove the Adhesive Substance from Common Carp Fish Eggs (*Cyprinus Carpio* L.) During Artificial Reproduction Euphrates. In: *Directorate of Agriculture Babylon Conference,* pp. 1-5.

Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB (2010) Annotating non-coding regions of the genome. *Nat Rev Genet,* **11**, 559-571.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2018) Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep,* **8**, 1-17.

Alli Shaik A, Wee S, Li RH, Li Z, Carney TJ, Mathavan S, Gunaratne J (2014) Functional mapping of the zebrafish early embryo proteome and transcriptome. *J Proteome Res,* **13**, 5536-5550.

Almeida DB, Magalhaes C, Sousa Z, Borges MT, Silva E, Blanquet I, Mucha AP (2021) Microbial community dynamics in a hatchery recirculating aquaculture system (RAS) of sole (*Solea senegalensis*). *Aquaculture,* **539**, 736592.

Aly SM, Eissa AE, ElBanna NI, Albutti A (2021a) Efficiency of monovalent and polyvalent *Vibrio alginolyticus* and *Vibrio Parahaemolyticus* vaccines on the immune response and protection in gilthead sea bream, *Sparus aurata* (L.) against vibriosis. *Fish Shellfish Immunol,* **111**, 145-151.

Aly SM, Sharaf SM, Hassanin AAI, Griesh AS (2021b) Relation of gilthead seabream (*Sparus aurata*) seasonal reproductive activity to hematology, serum biochemistry, histopathology, and Brdt gene expression. *Fish Physiol Biochem,* **47**, 961-977.

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington RA, Gerdol M, Aken B, Biscotti MA, Barucca M, Baurain D, Berlin AM, Blatch GL, Buonocore F, Burmester T, Campbell MS, Canapa A, Cannon JP, Christoffels A, De Moro G, Edkins AL, Fan L, Fausto AM, Feiner N, Forconi M, Gamieldien J, Gnerre S, Gnirke A, Goldstone JV, Haerty W, Hahn ME, Hesse U, Hoffmann S, Johnson J, Karchner SI, Kuraku S, Lara M, Levin JZ, Litman GW, Mauceli E, Miyake T, Mueller MG, Nelson DR, Nitsche A, Olmo E, Ota T, Pallavicini A, Panji S, Picone B, Ponting CP, Prohaska SJ, Przybylski D, Saha NR, Ravi V, Ribeiro

FJ, Sauka-Spengler T, Scapigliati G, Searle SMJ, Sharpe T, Simakov O, Stadler PF, Stegeman JJ, Sumiyama K, Tabbaa D, Tafer H, Turner-Maier J, van Heusden P, White S, Williams L, Yandell M, Brinkmann H, Volff J-N, Tabin CJ, Shubin N, Schartl M, Jaffe DB, Postlethwait JH, Venkatesh B, Di Palma F, Lander ES, Meyer A, Lindblad-Toh K (2013) The African coelacanth genome provides insights into tetrapod evolution. *Nature,* **496**, 311-316.

Amini F, Zamini AA, Ahmadi MR (2007) Intergeneric hybridization between kutum, *Rutilus frisii kutum*, and bream, *Abramis brama orientalis*, of the Caspian Sea. *J World Aquacult Soc,* **38**, 497-505.

Anastasiou S, Nathanailides C, Logothetis P, Kanlis G (2014) Management Challenges and Product Quality of Fish Farms in Greece. *Int J of Mech Ind En,* **8**, 3616-3620.

Anjo SI, Santa C, Manadas B (2015) Short GeLC-SWATH: a fast and reliable quantitative approach for proteomic screenings. *Proteomics,* **15**, 757-762.

Anjos L, Pinto PIS, Santos S, Estêvão MD, Santa C, Manadas B, Canário AVM, Power DM (2022) Proteome dataset of sea bass (*Dicentrarchus labrax*) skin-scales exposed to fluoxetine and estradiol. *Data Brief,* **41**, 107971.

Arezo MJ, Pereiro L, Berois N (2005) Early development in the annual fish Cynolebias viarius. *J Fish Biol,* **66**, 1357-1370.

Arreola MFO, Ortega GFA, López VG, Hernández RC, Weirich C, Morris T, Tenorio MD, Gámez CI (2015) Reproductive broodstock performance and egg quality of wild-caught and first-generation domesticated *Seriola rivoliana* reared under same culture conditions. *Lat Am J Aquat Res,* **43**, 953-962.

Arukwe A, Goksøyr A (2003) Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. *Comp Hepatol,* **2**, 1-21.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* **25**, 25-29.

Aydın İ, Okumuş İ (2017) Effects of triploid induction to the early cleavage patterns of black sea turbot (*Psetta maxima*) embryos. *Genet Aquat Org,* **1**, 15-20.

Babaheydari SB, Keyvanshokooh S, Dorafshan S, Johari SA (2017) Modifications in the proteome of rainbow trout (*Oncorhynchus mykiss*) embryo and fry as an effect of triploidy induction. *Fish Physiol Biochem,* **43**, 579-589.

Bachali S, Bailly X, Jolles J, Jolles P, Deutsch JS (2004) The lysozyme of the starfish Asterias rubens. A paradygmatic type i lysozyme. *Eur J Biochem,* **271**, 237-242.

Bachali S, Jager M, Hassanin A, Schoentgen F, Jolles P, Fiala-Medioni A, Deutsch JS (2002) Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J Mol Evol,* **54**, 652-664.

Bachvarova RF (1992) A Maternal Tail of Poly(a) - the Long and the Short of It. *Cell,* **69**, 895-897.

Badiola M, Mendiola D, Bostock J (2012) Recirculating Aquaculture Systems (RAS) analysis: Main issues on management and future challenges. *Aquacult Eng,* **51**, 26-35.

Bagarinao TU (1993) Sulfide as a toxicant in aquatic habitats. *SEAFDEC Asian Aquacult,* **15**, 2-4.

Baker J, Liu JP, Robertson EJ, Efstratiadis A (1993) Role of Insulin-Like Growth-Factors in Embryonic and Postnatal-Growth. *Cell,* **75**, 73-82.

Baki B, Ozturk DK, Tomgisi S (2021) Comparative Analysis of Egg Biochemical Composition and Egg Productivity Rainbow Trout (*Oncorhynchus mykiss* Walbaum, 1792) in Different Stations in Turkey. *Aquacult Stud,* **21**, 117-127.

Bally-Cuif L, Schatz WJ, Ho RK (1998) Characterization of the zebrafish Orb/CPEB-related RNA binding protein and localization of maternal components in the zebrafish oocyte. *Mech Dev,* **77**, 31-47.

Bandziulis RJ, Swanson MS, Dreyfuss G (1989) Rna-Binding Proteins as Developmental Regulators. *Gene Dev,* **3**, 431-437.

Baras E, Arifin OZ, Slembrouck J, Subagja J, Kristanto AH, Legendre M (2018) Oil globule size in fish eggs: A matter of biome and reproductive strategy. *Fish Fish,* **19**, 996-1002.

Barathkumar TR, Jeganathan M, Muthusamy T (2014) Phylogenetic analysis and codon usage bias estimation of the cDNA of catfish, Plotosus lineatus encoding toxin I natterin-like protein. *Global J Tradit Med Syst,* **3**, 1-7.

Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C (2014) jvenn: an interactive Venn diagram viewer. *Bmc Bioinf,* **15**, 1-7.

Barnes ME, Cordes RJ, Sayler WA, Hanten RP (2003) Soft-egg disease in landlocked fall chinook salmon eggs: Possible causes and therapeutic treatments. *N Am J Aquacult,* **65**, 126-133.

Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, Fu WL, Hamilton JK, Etkin LD, Lipshitz HD (1999) Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. *Embo J,* **18**, 2610-2620.

Basu S, Hadzhiev Y, Petrosino G, Nepal C, Gehrig J, Armant O, Ferg M, Strahle U, Sanges R, Muller F (2016)

The Tetraodon nigroviridis reference transcriptome: developmental transition, length retention and microsynteny of long non-coding RNAs in a compact vertebrate genome. *Sci Rep-Uk,* **6**, 1-14.

Batut B, Hiltemann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, Bretaudeau A, Brillet-Gueguen L, Cech M, Chilton J, Clements D, Doppelt-Azeroual O, Erxleben A, Freeberg MA, Gladman S, Hoogstrate Y, Hotz HR, Houwaart T, Jagtap P, Lariviere D, Le Corguille G, Manke T, Mareuil F, Ramirez F, Ryan D, Sigloch FC, Soranzo N, Wolff J, Videm P, Wolfien M, Wubuli A, Yusuf D, Taylor J, Backofen R, Nekrutenko A, Gruning B, Network GT (2018) Community-Driven Data Analysis Training for Biology. *Cell Syst,* **6**, 752-+.

Bazzini AA, del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin YD, Yao J, Khokha MK, Giraldez AJ (2016) Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *Embo J,* **35**, 2087-2103.

Bazzoli N (1992) Oogenesis in Neotropical freshwater teleosts. Federal University of Minas Gerais.

Bazzoli N, Rizzo E (1990) A comparative cytological and cytochemical study of the oogenesis in ten Brazilian teleost fish species. *Eur Arch Biol,* **101**, 399-410.

Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (2011) The quantitative proteome of a human cell line. *Mol Syst Biol,* **7**, 549.

Bekhit AEDA, Morton JD, Dawson CO, Zhao JH, Lee HYY (2009) Impact of maturity on the physicochemical and biochemical properties of chinook salmon roe. *Food Chem,* **117**, 318-325.

Bell JG, Sargent JR (2003) Arachidonic acid in aquaculture feeds: current status and future opportunities. *Aquaculture,* **218**, 491-499.

Bell JL, Wachter K, Muleck B, Pazaitis N, Kohn M, Lederer M, Huttelmaier S (2013) Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cell Mol Life Sci,* **70**, 2657-2675.

Bessonart M, Izquierdo MS, Salhi M, Hernandez-Cruz CM, Gonzalez MM, Fernandez-Palacios H (1999) Effect of dietary arachidonic acid levels on growth and survival of gilthead sea bream (*Sparus aurata* L.) larvae. *Aquaculture,* **179**, 265-275.

Bizuayehu TT, Mommens M, Sundaram AYM, Dhanasiri AKS, Babiak I (2019) Postovulatory maternal transcriptome in Atlantic salmon and its relation to developmental potential of embryos. *Bmc Genomics,* **20**, 1-13.

Bjarnmark NA, Yarra T, Churcher AM, Felix RC, Clark MS, Power DM (2016) Transcriptomics provides insight into Mytilus galloprovincialis (Mollusca: Bivalvia) mantle function and its role in biomineralisation. *Mar Genom,* **27**, 37-45.

Bobe J (2015a) Egg quality in fish: Present and future challenges. *Animal Frontiers,* **5**, 66-72.

Bobe J (2015b) Egg quality in fish: Present and future challenges. *Anim Front,* **5**, 66-72.

Bobe J, Labbe C (2010) Egg and sperm quality in fish. *Gen Comp Endocr,* **165**, 535-548.

Bonnet E, Fostier A, Bobe J (2007a) Characterization of rainbow trout egg quality: A case study using four different breeding protocols, with emphasis on the incidence of embryonic malformations. *Theriogenology,* **67**, 786-794.

Bonnet E, Fostier A, Bobe J (2007b) Microarray-based analysis of fish egg quality after natural or controlled ovulation. *BMC Genomics,* **8**, 1-17.

Boulekbache H (1981) Energy-Metabolism in Fish Development. *Am Zool,* **21**, 377-389.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, Berlin AM, Campbell MS, Barrell D, Martin KJ, Mulley JF, Ravi V, Lee AP, Nakamura T, Chalopin D, Fan S, Wcisel D, Canestro C, Sydes J, Beaudry FEG, Sun Y, Hertel J, Beam MJ, Fasold M, Ishiyama M, Johnson J, Kehr S, Lara M, Letaw JH, Litman GW, Litman RT, Mikami M, Ota T, Saha NR, Williams L, Stadler PF, Wang H, Taylor JS, Fontenot Q, Ferrara A, Searle SMJ, Aken B, Yandell M, Schneider I, Yoder JA, Volff J-N, Meyer A, Amemiya CT, Venkatesh B, Holland PWH, Guiguen Y, Bobe J, Shubin NH, Di Palma F, Alfoldi J, Lindblad-Toh K, Postlethwait JH (2016) The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet,* **48**, 427-437.

Brew K, Vanaman TC, Hill RL (1967) Comparison of the amino acid sequence of bovine α-lactalbumin and hens egg white lysozyme. *J Biol Chem,* **242**, 3747-3748.

Bringer A, Cachot JM, Dubillot E, Lalot B, Thomas H (2021) Evidence of deleterious effects of microplastics from aquaculture materials on pediveliger larva settlement and oyster spat growth of Pacific oyster, *Crassostrea gigas*. *Sci Total Environ,* **794**, 148708.

Brooks S, Tyler CR, Sumpter JP (1997) Quality in fish: what makes a good egg? *Rev Fish Biol Fisher,* **7**, 387-416.

Brown LL, Cox WT, Levine RP (1997) Evidence that the causal agent of bacterial cold-water disease *Flavobacterium psychrophilum* is transmitted within salmonid eggs. *Dis Aquat Org,* **29**, 213-218.

Bruning A, Holker F, Wolter C (2011) Artificial light at night: implications for early life stages development in four temperate freshwater fish species. *Aquat Sci,* **73**, 143-152.

Buonocore F, Randelli E, Trisolino P, Facchiano A, de Pascale D, Scapigliati G (2014) Molecular characterization,

gene structure and antibacterial activity of a g-type lysozyme from the European sea bass (*Dicentrarchus labrax* L.). *Mol Immunol,* **62**, 10-18.
Buric M, Bláhovec J, Kouril J (2014) A simple and effective recirculating hatchery for salmonids. *J Aquacult Res Dev,* **5**, 1-5.
Cabada MO, Darnbrough C, Ford PJ, Turner PC (1977) Differential accumulation of two size classes of poly (A) associated with messenger RNA during oogenesis in *Xenopus laevis*. *Dev Biol,* **57**, 427-439.
Cabrera-Quio LE, Schleiffer A, Mechtler K, Pauli A (2021) Zebrafish Ski7 tunes RNA levels during the oocyte-to-embryo transition. *PLoS Genet,* **17**, e1009390.
Cai JN, Galli GL (2021) Top 10 species groups in global, regional and national aquaculture 2019. FAO*,* pp. 1-325.
Cain SA, Coughlan T, Monk PN (2001) Mapping the ligand-binding site on the C5a receptor: arginine74 of C5a contacts aspartate282 of the C5a receptor. *Biochemistry,* **40**, 14047-14052.
Cain SA, Higginbottom A, Monk PN (2003) Characterisation of C5a receptor agonists from phage display libraries. *Biochem Pharmacol,* **66**, 1833-1840.
Callewaert L, Michiels CW (2010) Lysozymes in the animal kingdom. *J Biosci,* **35**, 127-160.
Campbell PM, Pottinger TG, Sumpter JP (1992) Stress reduces the quality of gametes produced by rainbow trout. *Biol Reprod,* **47**, 1140-1150.
Carballo C, Chronopoulou EG, Letsiou S, Spanidi E, Gardikis K, Labrou NE, Manchado M (2019) Genomic and phylogenetic analysis of choriolysins, and biological activity of hatching liquid in the flatfish Senegalese sole. *Plos One,* **14**.
Carnevali O, Carletta R, Cambi A, Vita A, Bromage N (1999) Yolk formation and degradation during oocyte maturation in seabream *Sparus aurata*: Involvement of two lysosomal proteinases. *Biol Reprod,* **60**, 140-146.
Carnevali O, Cionna C, Tosti L, Lubzens E, Maradonna F (2006) Role of cathepsins in ovarian follicle growth and maturation. *Gen Comp Endocr,* **146**, 195-203.
Carnevali O, Mosconi G, Cambi A, Ridolfi S, Zanuy S, Polzonetti-Magni AM (2001a) Changes of lysosomal enzyme activities in sea bass (*Dicentrarchus labrax*) eggs and developing embryos. *Aquaculture,* **202**, 249-256.
Carnevali O, Mosconi G, Cardinali M, Meiri I, Polzonetti-Magni A (2001b) Molecular components related to egg viability in the gilthead sea bream, *Sparus aurata*. *Mol Reprod Dev,* **58**, 330-335.
Carrillo M, Bromage N, Zanuy S, Serrano R, Prat F (1989) The Effect of Modifications in Photoperiod on Spawning Time, Ovarian Development and Egg Quality in the Sea Bass (*Dicentrarchus-Labrax* L). *Aquaculture,* **81**, 351-365.
Casimiro ACR, Vizintim Marques AC, Claro-Garcia A, Garcia DAZ, de Almeida FS, Orsi ML (2022) Hatchery fish stocking: case study, current Brazilian state, and suggestions for improvement. *Aquacult Int*, 1-18.
Castanho S, Califano G, Soares F, Costa R, Mata L, Pousao-Ferreira P, Ribeiro L (2017) The effect of live feeds bathed with the red seaweed Asparagopsis armata on the survival, growth and physiology status of Sparus aurata larvae. *Fish Physiol Biochem,* **43**, 1043-1054.
Castro R, Jouneau L, Tacchi L, Macqueen DJ, Alzaid A, Secombes CJ, Martin SAM, Boudinot P (2015) Disparate developmental patterns of immune responses to bacterial and viral infections in fish. *Sci Rep-Uk,* **5**, 15458.
Cavallo M, Frangoudes K, Agundez JP, Raux P (2020) Exploring Troubles, Attitudes, and Strategies Related to Integrated Aquaculture. A Case of the Andalusia Region (South of Spain). *J Mar Sci Eng,* **8**, 684.
Cecchini S, Terova G, Caricato G, Saroglia M (2000) Lysozyme activity in embryos and larvae of sea bass (Dicentrarchus labrax L.), spawned by broodstocks fed with vitamin C enriched diets. *B Eur Assoc Fish Pat,* **20**, 120-124.
Cejas JR, Almansa E, Jerez S, Bolanos A, Felipe B, Lorenzo A (2004) Changes in lipid class and fatty acid composition during development in white seabream (*Diplodus sargus*) eggs and larvae. *Comp Biochem Phys B,* **139**, 209-216.
Cerda J (2009) Molecular pathways during marine fish egg hydration: the role of aquaporins. *J Fish Biol,* **75**, 2175-2196.
Cerdà J, Maes T, Bobe J, Babin PJ, Admon A, Lubzens E (2007) Aplicación de la genómica funcional y proteómica en el estudio de la reproducción y calidad de los gametos de teleósteos. *Genética y Genómica en Acuicultura,* **13**, 573-619.
Cerda J, Zapater C, Chauvigne F, Finn RN (2013) Water homeostasis in the fish oocyte: new insights into the role and molecular regulation of a teleost-specific aquaporin. *Fish Physiol Biochem,* **39**, 19-27.
Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Komen H, Nielsen EE, Bargelloni L (2016) A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquacult Int,* **24**, 1287-1307.
Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant,* **13**, 1194-1202.
Chen F, Smith C, Wang YK, He J, Xia WL, Xue G, Chen J, Xie P (2021) The Evolution of Alternative Buoyancy

Mechanisms in Freshwater Fish Eggs. *Front Ecol Evol,* **9**, 736718.
Cheung C, Nguyen T, Le Cam A, Patinote A, Journot L, Reynes C, Bobe J (2018a) Lost in translation: egg transcriptome reveals molecular signature to predict developmental success and novel maternal-effect genes. *bioRxiv*.
Cheung C, Nguyen TTV, Le Cam A, Patinote A, Zarski D, Journot L, Reynes C, Bobe J (2018b) Molecular portrait of egg quality and novel maternal-effect genes. In: *11. International Symposium on Reproductive Physiology of Fish,* pp. 251 p.
Cheung CT, Nguyen TV, Le Cam A, Patinote A, Journot L, Reynes C, Bobe J (2019) What makes a bad egg? Egg transcriptome reveals dysregulation of translational machinery and novel fertility genes important for fertilization. *BMC Genomics,* **20**, 584.
Cheung CT, Patinote A, Guiguen Y, Bobe J (2018c) *foxr1* is a novel maternal-effect gene in fish that is required for early embryonic success. *Peerj,* **6**, e5534.
Chiu WWC, Erikson EKL, Sole CA, Shelling AN, Chamley LW (2004) SPRASA, a novel sperm protein involved in immune-mediated infertility. *Hum Reprod,* **19**, 243-249.
Choi EK, Sohn JH, Han SS, Kim DH (2019) Comparative ultrastructure of the fertilized egg envelope in *Corydoras adolfoi* and *Corydoras sterbai*, Callichthyidae, Teleostei. *Microsc Res Techniq,* **82**, 33-38.
Çiftci Y, Üstündağ C, Erteken A, Özongun M, Ceylan B, Haşimoğlu A, Güneş E, Yoseda K, Sakamoto F, Nezaki G (2002) Manual for the seed production of turbot, *Psetta maxima* in the Black Sea. *Special Publication*, 1-80.
Ciuhandu CS, Stevens ED, Wright PA (2005) The effect of oxygen on the growth of *Oncorhynchus mykiss* embryos with and without a chorion. *J Fish Biol,* **67**, 1544-1551.
Clough S, Mamo J, Hoevenaars K, Bardocz T, Petersen P, Rosendorf P, Atiye T, Gukelberger E, Guya E, Hoinkis J (2020) Innovative Technologies to Promote Sustainable Recirculating Aquaculture in Eastern Africa-A Case Study of a Nile Tilapia (*Oreochromis niloticus*) Hatchery in Kisumu, Kenya. *Integr Environ Asses,* **16**, 934-941.
Collins BC, Gillet LC, Rosenberger G, Rost HL, Vichalkovski A, Gstaiger M, Aebersold R (2013) Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods,* **10**, 1246-1253.
Collins LA, Nelson SG (1993) Effects of Temperature on Oxygen-Consumption, Growth, and Development of Embryos and Yolk-Sac Larvae of *Siganus-Randalli* (Pisces, Siganidae). *Mar Biol,* **117**, 195-204.
Contreras-Sanchez WM, Schreck CB, Fitzpatrick MS, Pereira CB (1998) Effects of stress on the reproductive performance of rainbow trout (*Oncorhynchus mykiss*). *Biol Reprod,* **58**, 439-447.
Corripio-Miyar Y, Bird S, Treasurer JW, Secombes CJ (2007) RAG-1 and IgM genes, markers for early development of the immune system in the gadoid haddock, *Melanogrammus aeglefinus*, L. *Fish Shellfish Immunol,* **23**, 71-85.
Council AA (2021) The provision of ecosystem services by European aquaculture.
Craik JCA, Harvey SM (1987) The Causes of Buoyancy in Eggs of Marine Teleosts. *J Mar Biol Assoc Uk,* **67**, 169-182.
Crespel A, Rime H, Fraboulet E, Bobe J, Fauvel C (2008) Egg quality in domesticated and wild seabass (*Dicentrarchus labrax*): A proteomic analysis. *Cybium,* **32**, 205-205.
Cucchi P, Sucre E, Santos R, Leclere J, Charmantier G, Castille R (2012) Embryonic development of the sea bass *Dicentrarchus labrax*. *Helgoland Mar Res,* **66**, 199-209.
Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SMJ, Clamp M (2004) The Ensembl automatic gene annotation system. *Genome Res,* **14**, 942-950.
Daffre S, Kylsten P, Samakovlis C, Hultmark D (1994) The lysozyme locus in *Drosophila melanogaster*: an expanded gene family adapted for expression in the digestive tract. *Mol Gen Genet,* **242**, 152-162.
Dang Y, Xu X, Shen Y, Hu M, Zhang M, Li L, Lv L, Li J (2016) Transcriptome Analysis of the Innate Immunity-Related Complement System in Spleen Tissue of *Ctenopharyngodon idella* Infected with *Aeromonas hydrophila*. *Plos One,* **11**, e0157413.
Dautigny A, Prager EM, Phamdinh D, Jolles J, Pakdel F, Grinde B, Jolles P (1991) Cdna and Amino-Acid-Sequences of Rainbow-Trout (*Oncorhynchus-Mykiss*) Lysozymes and Their Implications for the Evolution of Lysozyme and Lactalbumin. *J Mol Evol,* **32**, 187-198.
Davies B, Bromage N (2002) The effects of fluctuating seasonal and constant water temperatures on the photoperiodic advancement of reproduction in female rainbow trout *Oncorhynchus mykiss*. *Aquaculture,* **205**, 183-200.
De Souza AG, MacCormack TJ, Wang N, Li L, Goss GG (2009) Large-scale proteome profile of the zebrafish (*Danio rerio*) gill for physiological and biomarker discovery studies. *Zebrafish,* **6**, 229-238.
Deguara S, Jauncey K, Agius C (2003) Enzyme activities and pH variations in the digestive tract of gilthead sea bream. *J Fish Biol,* **62**, 1033-1043.
DelGiacco L, Vanoni C, Bonsignorio D, Duga S, Mosconi G, Santucci A, Cotelli F (1998) Identification and spatial

distribution of the mRNA encoding the gp49 component of the gilthead sea bream, *Sparus aurata*, egg envelope. *Mol Reprod Dev,* **49**, 58-69.
Deung YK, Reu DS, Kim DH (1997) Comparative ultrastructures of the fertilized egg envelopes in golden severum, convic cichlid and discus, Cichlidae, teleost. *Appl Microsc,* **27**, 417-432.
Di Santo V, Tran AH, Svendsen JC (2016) Progressive hypoxia decouples activity and aerobic performance of skate embryos. *Conserv Physiol,* **4**.
Ding JF, Li JA, Bao YB, Li L, Wu FC, Zhang GF (2011) Molecular characterization of a mollusk chicken-type lysozyme gene from *Haliotis discus hannai* Ino, and the antimicrobial activity of its recombinant protein. *Fish Shellfish Immunol,* **30**, 163-172.
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* **32**, 1792-1797.
Edvardsen RB, Leininger S, Kleppe L, Skaftnesmo KO, Wargelius A (2014) Targeted Mutagenesis in Atlantic Salmon (*Salmo salar* L.) Using the CRISPR/Cas9 System Induces Complete Knockout Individuals in the F0 Generation. *Plos One,* **9**, e108622.
Elkatatny NM, El Nahas AF, Helal MA, Fahmy HA, Tanekhy M (2020) The impacts of seasonal variation on the immune status of Nile tilapia larvae and their response to different immunostimulants feed additives. *Fish Shellfish Immunol,* **96**, 270-278.
Ellis AE (1988) Ontogeny of the immune system in teleost fish. *Fish vaccination*, 20-31.
Ellis AE (1990) Lysozyme assays. *Tech Fish Immunol,* **1**, 101-103.
Espinal CA, Matulić D (2019) Recirculating aquaculture technologies. *Aquaponics food production systems*, 35-76.
Evsikov AV, Graber JH, Brockman JM, Hampl A, Holbrook AE, Singh P, Eppig JJ, Solter D, Knowles BB (2006) Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Gene Dev,* **20**, 2713-2727.
Fabra M, Raldua D, Power DM, Deen PMT, Cerda J (2005) Marine fish egg hydration is aquaporin-mediated. *Science,* **307**, 545-545.
Fan SG, Zhang QY, Luo C (2009) Sequence cloning and expression analysis of Rag genes in goldfish. *Acta Hydrobiol Sin,* **33**, 603-612.
FAO (2020) The state of world fisheries and aquaculture 2020. In: *Sustainability in Action*, Rome.
FAO (2022) The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. Rome, Italy.
Fast MD, Sims DE, Burka JF, Mustafa A, Ross NW (2002) Skin morphology and humoral non-specific defence parameters of mucus and plasma in rainbow trout, coho and Atlantic salmon. *Comp Biochem Phys A,* **132**, 645-657.
Feist G, Schreck CB, Fitzpatrick MS, Redding JM (1990) Sex steroid profiles of coho salmon (*Oncorhynchus kisutch*) during early development and sexual differentiation. *Gen Comp Endocrinol,* **80**, 299-313.
Felix RC, Trindade M, Pires IRP, Fonseca VG, Martins RS, Silveira H, Power DM, Cardoso JCR (2015) Unravelling the Evolution of the Allatostatin-Type A, KISS and Galanin Peptide-Receptor Gene Families in Bilaterians: Insights from Anopheles Mosquitoes. *Plos One,* **10**, e0130347.
Ferosekhan S, Turkmen S, Perez-Garcia C, Xu HL, Gomez A, Shamna N, Afonso JM, Rosenlund G, Fontanillas R, Gracia A, Izquierdo M, Kaushik S (2021) Influence of Genetic Selection for Growth and Broodstock Diet n-3 LC-PUFA Levels on Reproductive Performance of Gilthead Seabream, *Sparus aurata*. *Animals,* **11**, 519.
Finn RN, Kristoffersen BA (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**, e169.
Finn RN, Wamboldt M, Fyhn HJ (2002) Differential processing of yolk proteins during oocyte hydration in marine fishes (Labridae) that spawn benthic and pelagic eggs. *Mar Ecol Prog Ser,* **237**, 217-226.
Firat K, Saka S, Kamaci HO (2005) Embryonic and larval development of striped sea bream (*Lithognathus mormyrus* L 1758). *Isr J Aquacult-Bamid,* **57**, 131-140.
Fishelson Z (1991) The third component of complement: Chemistry and biology : Edited by John D. Lambris . Current topics in microbiology and immunology, vol. 153. Springer, Heidelberg, 1990. **28**, 562.
Foley DA, Cheng TC (1977) Degranulation and other changes of molluscan granulocytes associated with phagocytosis. *J invertebr pathol,* **29**, 321-325.
Forniés MA, Mañanós E, Carrillo M, Rocha A, Laureau S, Mylonas CC, Zohar Y, Zanuy S (2001) Spawning induction of individual European sea bass females (*Dicentrarchus labrax*) using different GnRHa-delivery systems. *Aquaculture,* **202**, 221-234.
Franchini S, Zarkadis IK, Sfyroera G, Sahu A, Moore WT, Mastellos D, LaPatra SE, Lambris JD (2001) Cloning and purification of the rainbow trout fifth component of complement (C5). *Dev Comp Immunol,* **25**, 419-430.
Fraser AJ, Gamble JC, Sargent JR (1988) Changes in Lipid-Content, Lipid Class Composition and Fatty-Acid Composition of Developing Eggs and Unfed Larvae of Cod (*Gadus Morhua*). *Mar Biol,* **99**, 307-313.

Fridman S, Bron J, Rana K (2012) Influence of salinity on embryogenesis, survival, growth and oxygen consumption in embryos and yolk-sac larvae of the Nile tilapia. *Aquaculture,* **334-337**, 0-190.

Frost M, Baxter JM, Buckley PJ, Cox M, Dye SR, Harvey NW (2012) Impacts of climate change on fish, fisheries and aquaculture. *Aquat Conserv,* **22**, 331-336.

Fu GH, Bai ZY, Xia JH, Liu F, Liu P, Yue GH (2013) Analysis of Two Lysozyme Genes and Antimicrobial Functions of Their Recombinant Proteins in Asian Seabass. *Plos One,* **8**.

Fu W, Zhang JP, Xu WT, He S, Long MD, Liao QQ, Liu JH, Peng LY, Liu WB, Xiao YM (2021) Characteristics of hatching enzymes and egg envelope in cross progenies from crucian carp (*Carassius auratus* var.) and zebrafish (*Barchydanio rerio* var.). *Reprod Breed,* **1**, 81-88.

Fuhrmann M, Hausherr A, Ferbitz L, Schödl T, Heitzer M, Hegemann P (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol Biol,* **55**, 869-881.

Fujimoto T, Kataoka T, Otani S, Saito T, Aita T, Yamaha E, Arai K (2004) Embryonic stages from cleavage to gastrula in the loach *Misgurnus anguillicaudatus*. *Zoolog Sci,* **21**, 747-755.

Furuita H, Tanaka H, Yamamoto T, Shiraishi M, Takeuchi T (2000) Effects of n-3 HUFA levels in broodstock diet on the reproductive performance and egg and larval quality of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture,* **187**, 387-398.

Garcia-Gasca A, Galaviz MA, Gutierrez JN, Garcia-Ortega A (2006) Development of the digestive tract, trypsin activity and gene expression in eggs and larvae of the bullseye puffer fish *Sphoeroides annulatus*. *Aquaculture,* **251**, 366-376.

Gavilanes JG, Menendez-Arias L, Rodriguez R (1984) Comparative study on the secondary structure of lysozymes from different sources. *Comp Biochem Physiol B,* **77**, 83-88.

Gehring MR, Shiels BR, Northemann W, de Bruijn MH, Kan CC, Chain AC, Noonan DJ, Fey GH (1987) Sequence of rat liver alpha 2-macroglobulin and acute phase control of its messenger RNA. *J Biol Chem,* **262**, 446-454.

Gerdol M, Venier P (2015) An updated molecular basis for mussel immunity. *Fish Shellfish Immunol,* **46**, 17-38.

Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol Cell Proteomics,* **11**, O111.016717.

Gimenez G, Estevez A, Lahnsteiner F, Zecevic B, Bell JG, Henderson RJ, Pinera JA, Sanchez-Prado JA (2006) Egg quality criteria in common dentex (*Dentex dentex*). *Aquaculture,* **260**, 232-243.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science,* **308**, 833-838.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* **312**, 75-79.

Gohin M, Fournier E, Dufort I, Sirard MA (2014) Discovery, identification and sequence analysis of RNAs selected for very short or long poly A tail in immature bovine oocytes. *Mol Hum Reprod,* **20**, 127-138.

Goldstein RA (2008) The structure of protein evolution and the evolution of protein structure. *Curr Opin Struct Biol,* **18**, 170-177.

Graham M, Shin DH, Smith SL (2009) Molecular and expression analysis of complement component C5 in the nurse shark (*Ginglymostoma cirratum*) and its predicted functional role. *Fish Shellfish Immunol,* **27**, 40-49.

Greeley MSJ, Hols H, Wallace RA (1991) Changes in size, hydration and low molecular weight osmotic effectors during meiotic maturation of Fundulus oocytes in vivo. *Comp Biochem Physiol Part A: Physiol,* **100**, 639-647.

Grigorakis K, Alexis MN, Taylor KDA, Hole M (2002) Comparison of wild and cultured gilthead sea bream (*Sparus aurata*); composition, appearance and seasonal variations. *Int J Food Sci Tech,* **37**, 477-484.

Groh KJ, Nesatyy VJ, Segner H, Eggen RI, Suter MJ (2011) Global proteomics analysis of testis and ovary in adult zebrafish (*Danio rerio*). *Fish Physiol Biochem,* **37**, 619-647.

Grove BD, Wourms JP (1991) The Follicular Placenta of the Viviparous Fish, Heterandria-Formosa .1. Ultrastructure and Development of the Embryonic Absorptive Surface. *J Morphol,* **209**, 265-284.

Grutter MG, Weaver LH, Matthews BW (1983) Goose Lysozyme Structure - an Evolutionary Link between Hen and Bacteriophage Lysozymes. *Nature,* **303**, 828-831.

Guardiola FA, Saraiva-Fraga M, Cuesta A, Esteban MA (2018) Changes in natural haemolytic complement activity induced by stress in gilthead seabream (*Sparus aurata* L.). *Fish Shellfish Immunol,* **78**, 317-321.

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

Gulzari B, Komen H, Nammula VR, Bastiaansen JWM (2022) Genetic parameters and genotype by environment interaction for production traits and organ weights of gilthead seabream (*Sparus aurata*) reared in sea

cages. *Aquaculture,* **548**, 737555.
Guo RF, Huber-Lang M, Wang X, Sarma V, Padgaonkar VA, Craig RA, Riedemann NC, McClintock SD, Hlaing T, Shi MM (2000) Protective effects of anti-C5a in sepsis-induced thymocyte apoptosis. *J Clin Invest,* **106**, 1271-1280.
Haga S, Uji S, Suzuki T (2008) Evaluation of the effects of retinoids and carotenoids on egg quality using a microinjection system. *Aquaculture,* **282**, 111-116.
Hall TE, Smith P, Johnston IA (2004) Stages of embryonic development in the Atlantic cod *Gadus morhua*. *J Morphol,* **259**, 255-270.
Hallerman E, Esteban MA, Baldisserotto B (2022) Current Advances and Challenges in Fisheries and Aquaculture Science. *Fishes-Basel,* **7**, 87.
Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins,* **47**, 409-443.
Hamre K (2006) Nutrition in cod (*Gadus morhua*) larvae and juveniles. *Ices J Mar Sci,* **63**, 267-274.
Hang CY, Moriya S, Ogawa S, Parhar IS (2016) Deep Brain Photoreceptor (*val-opsin*) Gene Knockout Using CRISPR/Cas Affects Chorion Formation and Embryonic Hatching in the Zebrafish. *Plos One,* **11**, e0165535.
Hanif A, Bakopoulos V, Dimitriadis GJ (2004) Maternal transfer of humoral specific and non-specific immune parameters to sea bream (*Sparus aurata*) larvae. *Fish Shellfish Immunol,* **17**, 411-435.
Hara A, Hiramatsu N, Fujita T (2016) Vitellogenesis and choriogenesis in fishes. *Fisheries Sci,* **82**, 187-202.
Harikrishnan R, Kim JS, Kim MC, Balasundaram C, Heo MS (2011) Molecular characterization, phylogeny, and expression pattern of c-type lysozyme in kelp grouper, *Epinephelus bruneus*. *Fish Shellfish Immunol,* **31**, 588-594.
Hassan A, Okomoda VT, Nurhayati MN (2018) Embryonic development o diploid and triploid eggs of *Clarias gariepinus* (Burchell, 1822). *Caryologia,* **71**, 372-379.
Helvik JV, Hamre K, Hordvik I, Van der Meeren T, Ressem H, Tveiten H, Øie G, Schartl M (2009) The fish larva: a transitional life form, the foundation for aquaculture and fisheries. *The Research Council of Norway*, 5-35.
Henrotte E, Overton JL, Kestemont P (2008) Effects of dietary n-3 and n-6 fatty acids levels on egg and larval quality of Eurasian perch. *Cybium,* **32**, 271-272.
Herbomel P, Thisse B, Thisse C (1999) Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development,* **126**, 3735-3745.
Hiramatsu N, Todo T, Sullivan CV, Schilling J, Reading BJ, Matsubara T, Ryu YW, Mizuta H, Luo WS, Nishimiya O, Wu MQ, Mushirobira Y, Yilmaz O, Hara A (2015) Ovarian yolk formation in fishes: Molecular mechanisms underlying formation of lipid droplets and vitellogenin-derived yolk proteins. *Gen Comp Endocr,* **221**, 9-15.
Hochberg Y, Benjamini Y (1990) More Powerful Procedures for Multiple Significance Testing. *Stat Med,* **9**, 811-818.
Holland MC, Lambris JD (2002) The complement system in teleosts. *Fish Shellfish Immunol,* **12**, 399-420.
Honji RM, Tolussi CE, Mello PH, Caneppele D, Moreira RG (2012) Embryonic development and larval stages of *Steindachneridion parahybae* (Siluriformes: Pimelodidae) - implications for the conservation and rearing of this endangered Neotropical species. *Neotrop Ichthyol,* **10**, 313-327.
Hosokawa K (1985) Electron-Microscopic Observation of Chorion Formation in the Teleost, *Navodon-Modestus*. *Zool Sci,* **2**, 513-522.
Howley C, Ho RK (2000) mRNA localization patterns in zebrafish oocytes. *Mech Dev,* **92**, 305-309.
Huang C, Zhu B, Leng DL, Ge W, Zhang XHD (2021a) Long noncoding RNAs implicated in embryonic development in Ybx1 knockout zebrafish. *Febs Open Bio,* **11**, 1259-1276.
Huang JS, Amenyogbe E, Chen G, Wang WZ (2021b) Biochemical composition and activities of digestive and antioxidant enzymes during the egg and yolk-sac larval development of the cobia (*Rachycentron canadum*). *Aquacult Res,* **52**, 1643-1656.
Huang LY, Li BX, Luo C, Xie JY, Chen P, Liang SP (2004) Proteome comparative analysis of gynogenetic haploid and diploid embryos of goldfish (*Carassius auratus*). *Proteomics,* **4**, 235-243.
Huang SQ, Yang LJ, Zhang L, Sun B, Gao J, Chen ZJ, Zhong L, Cao XJ (2022) Endogenic upregulations of HIF/VEGF signaling pathway genes promote air breathing organ angiogenesis in bimodal respiration fish. *Funct Integr Genomic,* **22**, 65-76.
Huang SY (2015) Exploring the potential of global protein-protein docking: an overview and critical assessment of current programs for automatic ab initio docking. *Drug Discov Today,* **20**, 969-977.
Hughes AL (1994) Phylogeny of the C3/C4/C5 complement-component gene family indicates that C5 diverged first. *Mol Biol Evol,* **11**, 417-425.
Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar

RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature,* **545**, 505-509.
Hwang PP, Wu SM, Lin JH, Wu LS (1992) Cortisol content of eggs and larvae of teleosts. *Gen Comp Endocrinol,* **86**, 189-196.
Ibarz A, Martin-Perez M, Blasco J, Bellido D, de Oliveira E, Fernandez-Borras J (2010) Gilthead sea bream liver proteome altered at low temperatures by oxidative stress. *Proteomics,* **10**, 963-975.
Ienaga N, Higuchi K, Takashi T, Gen K, Tsuda K, Terayama K (2021) Vision-based egg quality prediction in Pacific bluefin tuna (*Thunnus orientalis*) by deep neural network. *Sci Rep-Uk,* **11**.
Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*, 13-34.
Imsland AK, Foss A, Koedijk R, Folkvord A, Stefansson SO, Jonassen TM (2006) Short- and long-term differences in growth, feed conversion efficiency and deformities in juvenile Atlantic cod (*Gadus morhua*) startfed on rotifers or zooplankton. *Aquacult Res,* **37**, 1015-1027.
Inanli AG, Coban OE, Yilmaz O, Ozpolat E, Kuzgun NK (2019) Assessment of vitamin compositions and cholesterol levels of carp (*Cyprinus carpio carpio*) and rainbow trout (*Oncorhynchus mykiss*) caviars. *Su Urunleri Dergisi,* **36**, 293-299.
Inohaya K, Yasumasu S, Ishimaru M, Ohyama A, Iuchi I, Yamagami K (1995) Temporal and Spatial Patterns of Gene-Expression for the Hatching Enzyme in the Teleost Embryo, *Oryzias-Latipes*. *Dev Biol,* **171**, 374-385.
Inohaya K, Yasumasu S, Yasumasu I, Iuchi I, Yamagami K (1999) Analysis of the origin and development of hatching gland cells by transplantation of the embryonic shield in the fish, *Oryzias latipes*. *Dev Growth Differ,* **41**, 557-566.
Intarasirisawat R, Benjakul S, Visessanguan W (2011) Chemical compositions of the roes from skipjack, tongol and bonito. *Food Chem,* **124**, 1328-1334.
Interdonato F (2012) Recirculating aquaculture system (RAS) biofilters: focusing on bacterial communities complexity and activity. Università degli studi di Messina.
Irie N, Kuratani S (2011) Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun,* **2**, 248.
Irwin DM (2014) Evolution of the vertebrate goose-type lysozyme gene family. *Bmc Evol Biol,* **14**, 188.
Irwin DM, Biegel JM, Stewart CB (2011) Evolution of the mammalian lysozyme gene family. *Bmc Evol Biol,* **11**, 166.
Irwin DM, Gong Z (2003) Molecular evolution of vertebrate goose-type lysozyme genes. *J Mol Evol,* **56**, 234-242.
Irwin DM, Wilson AC (1990) Concerted Evolution of Ruminant Stomach Lysozymes - Characterization of Lysozyme Cdna Clones from Sheep and Deer. *J Biol Chem,* **265**, 4944-4952.
Ishigaki M, Taketani A, Sato H (2014) Discrimination of fish egg quality and viability by Raman spectroscopy. *Anal Methods,* **6**, 9206-9211.
Issa FO, Aderinoye-AbdulWahab S, Kagbu JH (2022) Assessment of aquaculture development programmes in Nigeria. *J Agricult Ext,* **26**, 10-17.
Ito K, Takizawa F, Yoshiura Y, Ototake M, Nakanishi T (2008) Expression profile of cytokine and transcription factor genes during embryonic development of zebrafish *Danio rerio*. *Fisheries Sci,* **74**, 391-396.
Ito Y, Yoshikawa A, Hotani T, Fukuda S, Sugimura K, Imoto T (1999) Amino acid sequences of lysozymes newly purified from invertebrates imply wide distribution of a novel class in the lysozyme family. *Eur J Biochem,* **259**, 456-461.
Itoh N, Takahashi KG (2009) A novel peptidoglycan recognition protein containing a goose-type lysozyme domain from the Pacific oyster, *Crassostrea gigas*. *Mol Immunol,* **46**, 1768-1774.
Iwamatsu T (2004) Stages of normal development in the medaka *Oryzias latipes*. *Mech Dev,* **121**, 605-618.
Izumi H, Gen K, Lokman PM, Hagihara S, Horiuchi M, Tanaka T, Ijiri S, Adachi S (2019) Maternal transcripts in good and poor quality eggs from Japanese eel, *Anguilla japonica*-their identification by large-scale quantitative analysis. *Mol Reprod Dev,* **86**, 1846-1864.
Janeway JCA, Travers P, Walport M, Shlomchik MJ (2001) The complement system and innate immunity. In: *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science.
Janssen K, Chavanne H, Berentsen P, Komen H (2017) Impact of selective breeding on European aquaculture. *Aquaculture,* **472**, 8-16.
Jia YD, Meng Z, Liu XF, Lei JL (2014) Biochemical composition and quality of turbot (*Scophthalmus maximus*) eggs throughout the reproductive season. *Fish Physiol Biochem,* **40**, 1093-1104.
Jiang CY, Wang P, Liu SS, Wang YF, Li J, Ji GD (2019) Administration of Yeast Glucan on Immunity of Offspring in Turbot *Scophthalmus maximus*: A Trans-Generational Immune-Enhancing Effect. *J Ocean U China,* **18**, 997-1003.
Jiang H, Deng R, Yang X, Shang J, Lu S, Zhao Y, Song K, Liu X, Zhang Q, Chen Y, Chinn YE, Wu G, Li J, Chen G, Yu J, Zhang J (2017) Peptidomimetic inhibitors of APC-Asef interaction block colorectal cancer

migration. *Nat Chem Biol,* **13**, 994-1001.
Jimenez-Cantizano RM, Infante C, Martin-Antonio B, Ponce M, Hachero I, Navas JI, Manchado M (2008) Molecular characterization, phylogeny, and expression of c-type and g-type lysozymes in brill (*Scophthalmus rhombus*). *Fish Shellfish Immunol,* **25**, 57-65.
Johansen W, Grove S, Anonsen JH, Moen A, Agusti-Ridaura C, Azar AS, Straetkvern KO (2019) Complement factor C5 in Atlantic salmon (*Salmo salar*): Characterization of cDNA, protein and glycosylation. *Dev Comp Immunol,* **100**, 103424.
Johnsson P, Lipovich L, Grandér D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta Gen Subj,* **1840**, 1063-1071.
Jollès P (1996) *Lysozymes--model enzymes in biochemistry and biology,* Birkhäuser Verlag, Basel ; Boston, viii, 449 p. p.
Jonsson B, Jonsson N (2019) Phenotypic plasticity and epigenetics of fish: embryo temperature affects later-developing life-history traits. *Aquat Biol,* **28**, 21-32.
Joo KB, Kim DH (2013) Comparative ultrastructures of the fertilized egg envelopes in *Danio rerio* and *Danio rerio* var. frankei, Cyprinidae, Teleostei. *Appl Microsc,* **43**, 14-20.
Joris M, Schloesser M, Baurain D, Hanikenne M, Muller M, Motte P (2017) Number of inadvertent RNA targets for morpholino knockdown in *Danio rerio* is largely underestimated: evidence from the study of Ser/Arg-rich splicing factors. *Nucleic Acids Res,* **45**, 9547-9557.
Jukam D, Shariati SAM, Skotheim JM (2017) Zygotic Genome Activation in Vertebrates. *Dev Cell,* **42**, 316-332.
Jung KM, Folkvord A, Kjesbu OS, Agnalt AL, Thorsen A, Sundby S (2012) Egg buoyancy variability in local populations of Atlantic cod (*Gadus morhua*). *Mar Biol,* **159**, 1969-1980.
Jung KM, Folkvord A, Kjesbu OS, Sundby S (2014) Experimental Parameterisation of Principal Physics in Buoyancy Variations of Marine Teleost Eggs. *Plos One,* **9**, e104089.
Kaizu A, Fagutao FF, Kondo H, Aoki T, Hirono I (2011) Functional analysis of C-type lysozyme in penaeid shrimp. *J Biol Chem,* **286**, 44344-44349.
Kaleem O, Sabi AFBS (2021) Overview of aquaculture systems in Egypt and Nigeria, prospects, potentials, and constraints. *Aquacult Fish,* **6**, 535-547.
Kamacı HO, Saka Ş, Fırat K (2005) The Cleavage and Embryonic Phase of Gilthead Sea Bream (*Sparus aurata* Linnaeus, 1758) Eggs. *Su Ürünleri Dergisi,* **22**, 205-209.
Kaminski R, Kamler E, Korwin-Kossakowski M, Myszkowski L, Wolnicki J (2006) Effects of different incubation temperatures on the yolk-feeding stage of *Eupallasella percnurus* (Pallas). *J Fish Biol,* **68**, 1077-1090.
Kanamori A, Naruse K, Mitani H, Shima A, Hori H (2003) Genomic organization of ZP domain containing egg envelope genes in medaka (*Oryzias latipes*). *Gene,* **305**, 35-45.
Kane DA, Kimmel CB (1993) The Zebrafish Midblastula Transition. *Development,* **119**, 447-456.
Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature,* **447**, 714-719.
Katselis G, Tsolakos K, Theodorou JA (2022) Mapping of Greek Marine Finfish Farms and Their Potential Impact on the Marine Environment. *J Mar Sci Eng,* **10**, 286.
Kaushik K, Leonard VE, Shamsudheen KV, Lalwani MK, Jalali S, Patowary A, Joshi A, Scaria V, Sivasubbu S (2013) Dynamic Expression of Long Non-Coding RNAs (lncRNAs) in Adult Zebrafish. *Plos One,* **8**, e83616.
Kazemi R, Yarmohammadi M, Hallajian A, Jalilpour J, Esmaeili F (2020) Influence of the photoperiod and light intensity on growth performance, Melatonin and Insulin-like growth factors gene expression on *Acipenser persicus* during the embryonic stage. *Iran J Fish Sci,* **19**, 1175-1192.
Keller M, Molenaar P, de Leeuw J, Mooij W, Rijnsdorp A, van de Wolfshaar K (2020) Temperature effects on egg and larval development rate in European smelt, *Osmerus eperlanus*, experiments and a 50 year hindcast. *J Fish Biol,* **96**, 1422-1433.
Kim DH (2020) Ultrastructure of the fertilized egg envelopes in *Ancistrus cirrhosus*, Loricariidae, Teleostei. *Appl Microsc,* **50**, 1-7.
Kim DH, Chang BS, Teng YC, Kim S, Joo KB, Lee KJ (2009) Ultrastructure of the fertilized egg envelope in *Cichlasoma managuensis*, Cichlidae, Teleost. *Appl Microsc,* **39**, 9-15.
Kim DH, Deung YK, Kim WJ, Reu DS, Kang SJ (1999) Comparative Ultranstructures of the Fertilized Egg Envelopes from Three-spot gourami, Pearl gourami and Marble gourami, Belontiidae, Teleost. *Appl Microsc,* **29**, 343-351.
Kim K-W, Nam TJ, Choi YH (2018) Protein variation and involvement of insulin-like growth factor during embryonic development in the olive flounder *Paralichthys olivaceus*. *Fish Aquat Sci,* **21**, 4.
Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of Embryonic-Development of the Zebrafish. *Dev Dynam,* **203**, 253-310.
Kjørsvik E (1994) Egg quality in wild and broodstock cod *Gadus morhua* L. *J World Aquacult Soc,* **25**, 22-29.
Kjorsvik E, Hoehne-Reitan K, Reitan KI (2003) Egg and larval quality criteria as predictive measures for juvenile

production in turbot (*Scophthalmus maximus* L.). *Aquaculture,* **227**, 9-20.
Kjorsvik E, Mangorjensen A, Holmefjord I (1990) Egg Quality in Fishes. *Adv Mar Biol,* **26**, 71-113.
Kleene KC (2016) Position-dependent interactions of Y-box protein 2 (YBX2) with mRNA enable mRNA storage in round spermatids by repressing mRNA translation and blocking translation-dependent mRNA decay. *Mol Reprod Dev,* **83**, 190-207.
Kleppe L, Edvardsen RB, Kuhl H, Malde K, Furmanek T, Drivenes O, Reinhardt R, Taranger GL, Wargelius A (2012) Maternal 3 ' UTRs: from egg to onset of zygotic transcription in Atlantic cod. *Bmc Genomics,* **13**, 1-14.
Kohn YY, Symonds JE, Kleffmann T, Nakagawa S, Lagisz M, Lokman PM (2015) Proteomic analysis of early-stage embryos: implications for egg quality in hapuku (*Polyprion oxygeneios*). *Fish Physiol Biochem,* **41**, 1403-1417.
Komar C, Turnbull JF, Roque A, Fajer E, Duncan NJ (2004) Effect of water treatment and aeration on the percentage hatch of demersal, adhesive eggs of the bullseye puffer (*Sphoeroides annulatus*). *Aquaculture,* **229**, 147-158.
Kong R, Wang F, Zhang J, Wang F, Chang S (2019) CoDockPP: A Multistage Approach for Global and Site-Specific Protein-Protein Docking. *J Chem Inf Model,* **59**, 3556-3564.
Korwin-Kossakowski M (2012) Fish hatching strategies: a review. *Rev Fish Biol Fisher,* **22**, 225-240.
Koumansvandiepen JCE, Tavernethiele JJ, Vanrens BTTM, Rombout JHWM (1994) Immunocytochemical and Flow Cytometric Analysis of B-Cells and Plasma-Cells in Carp (*Cyprinus-Carpio* L) - an Ontogenic Study. *Fish Shellfish Immunol,* **4**, 19-28.
Kowalska-Goralska M, Formicki K, Dobrzanski Z, Wondolowska-Grabowska A, Skrzynska E, Korzelecka-Orkisz A, Nedzarek A, Tanski A (2020) Nutritional Composition of Salmonidae and Acipenseridae Fish Eggs. *Ann Anim Sci,* **20**, 629-645.
Krogdahl A, Hemre GI, Mommsen TP (2005) Carbohydrates in fish nutrition: digestion and absorption in postlarval stages. *Aquacult Nutr,* **11**, 103-122.
Kronja I, Whitfield ZJ, Yuan BB, Dzeyk K, Kirkpatrick J, Krijgsveld J, Orr-Weaver TL (2014) Quantitative proteomics reveals the dynamics of protein changes during Drosophila oocyte maturation and the oocyte-to-embryo transition. *P Natl Acad Sci USA,* **111**, 16023-16028.
Kudo S (1992) Enzymatic Basis for Protection of Fish Embryos by the Fertilization Envelope. *Experientia,* **48**, 277-281.
Kudo S, Teshima C (1991) Enzyme-Activities and Antifungal Action of Fertilization Envelope Extract from Fish Eggs. *J Exp Zool,* **259**, 392-398.
Kwon JK, Jung HS, Kim DH (2015) Comparative ultrastructures of the fertilized egg envelopes in *Nothobranchius guentheri* and *Nothobranchius patrizii*, Nothobranchiidae, Teleostei. *Appl Microsc,* **45**, 144-149.
Kwon O, Sohn JH, Chung DY, Kim EJ, Kim DH (2017) Comparative ultrastructures of the fertilized egg envelopes in *Nothobranchius foerschi* and *Nothobranchius rachovii*, Nothobranchiidae, teleostei. *Appl Microsc,* **47**, 70-74.
L. Tort JCB, S. Mackenzie (2003) Fish immune system. A crossroads between innate and adaptive responses. *Inmunología,* **22**, 277-286.
Laale HW (1980) The Perivitelline Space and Egg Envelopes of Bony Fishes - a Review. *Copeia*, 210-226.
Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227**, 680-685.
Lahnsteiner F, Gimenez G, Estevez A (2008) Egg quality determination based on the shape of the lipid vesicle in common dentex, *Dentex dentex*. *Aquacult Res,* **39**, 144-149.
Lahnsteiner F, Patarnello P (2004a) Biochemical egg quality determination in the gilthead seabream, *Sparus aurata*: reproducibility of the method and its application for sharpsnout seabream, *Puntazzo puntazzo*. *Aquaculture,* **237**, 433-442.
Lahnsteiner F, Patarnello P (2004b) Egg quality determination in the gilthead seabream, *Sparus aurata*, with biochemical parameters. *Aquaculture,* **237**, 443-459.
Lahnsteiner F, Patarnello P (2005) The shape of the lipid vesicle is a potential marker for egg quality determination in the gilthead seabream, *Sparus aurata*, and in the sharpsnout seabream, *Diplodus puntazzo*. *Aquaculture,* **246**, 423-435.
Lambert JP, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin ZY, Zhong Q, Lindquist S, Vidal M, Aebersold R, Pawson T, Bonner R, Tate S, Gingras AC (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat Methods,* **10**, 1239-1245.
Lambris JD, Reid KB, Volanakis JE (1999) The evolution, structure, biology and pathophysiology of complement. *Immunol Today,* **20**, 207-211.
Lanes CFC, Bizuayehu TT, Fernandes JMD, Kiron V, Babiak I (2013) Transcriptome of Atlantic Cod (*Gadus morhua* L.) Early Embryos from Farmed and Wild Broodstocks. *Mar Biotechnol,* **15**, 677-694.
Langfelder P, Castellani LW, Zhou Z, Paul E, Davis R, Schadt EE, Lusis AJ, Horvath S, Mehrabian M (2012) A

systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochim Biophys Acta,* **1821**, 435-447.
Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? *PLoS Comput Biol,* **7**.
Langlois RE, Lu H (2008) Chapter 3 Machine Learning for Protein Structure and Function Prediction. *Annu Rep Comput Chem,* **4**, 41-66.
Larsson J (2020) eulerr: Area-Proportional Euler and Venn Diagrams with ellipses. *R package version 6.1.0*.
Lasram FBR, Guilhaumon F, Mouillot D (2009) Fish diversity patterns in the Mediterranean Sea: deviations from a mid-domain model. *Mar Ecol Prog Ser,* **376**, 253-267.
Laue K, Rajshekar S, Courtney AJ, Lewis ZA, Goll MG (2019) The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nat Commun,* **10**, 1-10.
Lavens P, Lebegue E, Jaunet H, Brunel A, Dhert P, Sorgeloos P (1999) Effect of dietary essential fatty acids and vitamins on egg quality in turbot broodstocks. *Aquacult Int,* **7**, 225-240.
Lavner Y, Kotlar D (2005) Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene,* **345**, 0-138.
Layman DK, Lönnerdal B, Fernstrom JD (2018) Applications for α-lactalbumin in human nutrition. *Nutr Rev,* **76**, 444-460.
Lee J-W, Lee J-H, Noh JK, Kim HC, Park C-J, Park J-W, Kim K-K (2014a) Transcriptional Onset of Lysozyme Genes during Early Development in Olive Flounder (*Paralichthys olivaceus*). *Dev Reprod,* **18**, 267-274.
Lee JW, Kim JE, Goo IB, Hwang JA, Im JH, Choi HS, Lee JH (2015) Expression of Immune-Related Genes during Loach (*Misgurnus anguillicaudatus*) Embryonic and Early Larval Development. *Dev Reprod,* **19**, 181-187.
Lee MT, Bonneau AR, Giraldez AJ (2014b) Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annu Rev Cell Dev Bi,* **30**, 581-613.
Lefort V, Longueville JE, Gascuel O (2017) SMS: Smart Model Selection in PhyML. *Mol Biol Evol,* **34**, 2422-2424.
Leifson RMQ (2003) Phospholipids in formulated starter feed for turbot (*Scophthalmus maximus* L.) and cod (*Gadus morhua* L.) larvae–effects on mitochondrial membranes in turbot larvae enterocytes. PhD Thesis. University of Tromsø, Tromsø, Norway.
Li LS, Cardoso JCR, Felix RC, Mateus AP, Canario AVM, Power DM (2021) Fish lysozyme gene family evolution and divergent function in early development. *Dev Comp Immunol,* **114**, 103772.
Li M, Leatherland JF (2012) The interaction between maternal stress and the ontogeny of the innate immune system during teleost embryogenesis: implications for aquaculture practice. *J Fish Biol,* **81**, 1793-1814.
Li X, Lu YC, Dai KZ, Torregroza I, Hla T, Evans T (2014) Elavl1a regulates zebrafish erythropoiesis via posttranscriptional control of gata1. *Blood,* **123**, 1384-1392.
Li XH, Chavali PL, Babu MM (2018) Capturing dynamic protein interactions. *Science,* **359**, 1105-1106.
Li XP, Sun L (2017) A teleost complement factor Ba possesses antimicrobial activity and inhibits bacterial infection in fish. *Dev Comp Immunol,* **71**, 49-58.
Liao Z, Wan Q, Yuan G, Su J (2019) The systematic identification and mRNA expression profiles post viral or bacterial challenge of complement system in grass carp Ctenopharyngodon idella. *Fish Shellfish Immunol,* **86**, 107-115.
Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, Di Genova A, Samy JKA, Vik JO, Vigeland MD, Caler L, Grimholt U, Jentoft S, Vage DI, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang XT, Fan DD, Liberles DA, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature,* **533**, 200-205.
Lin Y, Chen Y, Yang X, Xu D, Liang S (2009) Proteome analysis of a single zebrafish embryo using three different digestion strategies coupled with liquid chromatography-tandem mass spectrometry. *Anal Biochem,* **394**, 177-185.
Liu M, Zhang S, Liu Z, Li H, Xu A (2006) Characterization, organization and expression of AmphiLysC, an acidic c-type lysozyme gene in amphioxus Branchiostoma belcheri tsingtauense. *Gene,* **367**, 110-117.
Liu Y, Zheng ZJ, Yu LH, Wu S, Sun L, Wu SL, Xu Q, Cai SF, Qin N, Bao WB (2019) Examination of the temporal and spatial dynamics of the gut microbiome in newborn piglets reveals distinct microbial communities in six intestinal segments. *Sci Rep-Uk,* **9**, 3453.
Llorente I, Fernandez-Polanco J, Baraibar-Diez E, Odriozola MD, Bjorndal T, Asche F, Guillen J, Avdelas L, Nielsen R, Cozzolino M, Luna M, Fernandez-Sanchez JL, Luna L, Aguilera C, Basurco B (2020) Assessment of the economic performance of the seabream and seabass aquaculture industry in the European Union. *Mar Policy,* **117**, 103876.
Lonning S, Kjorsvik E, Falkpetersen IB (1988) A Comparative-Study of Pelagic and Demersal Eggs from Common

Marine Fishes in Northern Norway. *Sarsia,* **73**, 49-60.
Lopez JL (2005) Role of proteomics in taxonomy: the Mytilus complex as a model of study. *J Chromatogr B,* **815**, 261-274.
Lotan T, Chalifa-Caspi V, Ziv T, Brekhman V, Gordon MM, Admon A, Lubzens E (2014) Evolutionary conservation of the mature oocyte proteome. *EuPA Open Proteomics,* **3**, 27-36.
Lovoll M, Johnsen H, Boshra H, Bogwald J, Sunyer JO, Dalmo RA (2007) The ontogeny and extrahepatic expression of complement factor C3 in Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol,* **23**, 542-552.
Lubzens E, Bobe J, Young G, Sullivan CV (2017) Maternal investment in fish oocytes and eggs: The molecular cargo and its contributions to fertility and early development. *Aquaculture,* **472**, 107-143.
Lubzens E, Young G, Bobe J, Cerda J (2010) Oogenesis in teleosts: How fish eggs are formed. *Gen Comp Endocr,* **165**, 367-389.
Lucitt MB, Price TS, Pizarro A, Wu W, Yocum AK, Seiler C, Pack MA, Blair IA, FitzGerald GA, Grosser T (2008) Analysis of the zebrafish proteome during embryonic development. *Mol Cell Proteomics,* **7**, 981-994.
Luo MK, Wang LM, Yin HR, Zhu WB, Fu JJ, Dong ZJ (2019) Integrated analysis of long non-coding RNA and mRNA expression in different colored skin of koi carp. *Bmc Genomics,* **20**, 1-14.
Lv H, Zhou T, Dong C, Kong S, Chen L, Pu F, Li X, Xu P (2020) Genome-wide identification, evolution, and mRNA expression of complement genes in common carp (*Cyprinus carpio*). *Fish Shellfish Immunol,* **96**, 190-200.
Lvoll M, Kilvik T, Boshra H, Bogwald J, Sunyer JO, Dalmo RA (2006) Maternal transfer of complement components C3-1, C3-3, C3-4, C4, C5, C7, Bf, and Df to offspring in rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics,* **58**, 168-179.
Lyman G, J., Pelegri F (2007) Maternal factors in fish oogenesis and embryonic development. In: *The fish oocyte*. Springer, pp. 141-174.
Lyons CE, Payette KL, Price JL, Huang RC (1993) Expression and structural analysis of a teleost homolog of a mammalian zona pellucida gene. *J Biol Chem,* **268**, 21351-21358.
Ma H, Martin K, Dixon D, 2nd, Hernandez AG, Weber GM (2019a) Transcriptome analysis of egg viability in rainbow trout, *Oncorhynchus mykiss*. *BMC Genomics,* **20**, 319.
Ma H, Martin K, Dixon D, Hernandez AG, Weber GM (2019b) Transcriptome analysis of egg viability in rainbow trout, *Oncorhynchus mykiss*. *Bmc Genomics,* **20**, 1-15.
Machado TM, Tabata YA, Takahashi NS, Casarini LM, Neiva CRP, Henriques MB (2016) Caviar substitute produced from roes of rainbow trout (*Oncorhynchus mykiss*). *Acta Sci-Technol,* **38**, 233-240.
Maeland A, Rønnestad I, Waagbø R (2003) Folate in eggs and developing larvae of Atlantic halibut, *Hippoglossus hippoglossus* L. *Aquacult Nutr,* **9**, 185-188.
Magnadottir B (2010) Immunological control of fish diseases. *Mar Biotechnol (NY),* **12**, 361-379.
Magnadottir B, Lange S, Gudmundsdottir S, Bogwald J, Dalmo RA (2005) Ontogeny of humoral immune parameters in fish. *Fish Shellfish Immunol,* **19**, 429-439.
Magnadottir B, Lange S, Steinarsson A, Gudmundsdottir S (2004) The ontogenic development of innate immune parameters of cod (*Gadus morhua* L.). *Comp Biochem Phys B,* **139**, 217-224.
Mandal A, Klotz KL, Shetty J, Jayes FL, Wolkowicz MJ, Bolling LC, Coonrod SA, Black MB, Diekman AB, Haystead TA, Flickinger CJ, Herr JC (2003) SLLP1, a unique, intra-acrosomal, non-bacteriolytic, c lysozyme-like protein of human spermatozoa. *Biol Reprod,* **68**, 1525-1537.
Manthey HD, Woodruff TM, Taylor SM, Monk PN (2009) Complement component 5a (C5a). *Int J Biochem Cell Biol,* **41**, 2114-2117.
Marimuthu K (2019) A short review on induced spawning and seed production of African Catfish *Clarias gariepinus* in Malaysia. *Iop C Ser Earth Env,* **348**, 012134.
Martins CIM, Pistrin MG, Ende SSW, Eding EH, Verreth JAJ (2009) The accumulation of substances in Recirculating Aquaculture Systems (RAS) affects embryonic and larval development in common carp *Cyprinus carpio*. *Aquaculture,* **291**, 65-73.
Mathavan S, Lee SGP, Mak A, Miller LD, Murthy KRK, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan YJ, Korzh V, Gong ZY, Liu ET, Lufkin T (2005) Transcriptome analysis of zebrafish embryogenesis using microarrays. *Plos Genet,* **1**, 260-276.
Matsubara T, Koya Y (1997) Course of proteolytic cleavage in three classes of yolk proteins during oocyte maturation in barfin flounder *Verasper moseri*, a marine teleost spawning pelagic eggs. *J Exp Zool,* **278**, 189-200.
Matsubara T, Ohkubo N, Andoh T, Sullivan CV, 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. *Dev Biol,* **213**, 18-32.
McDade JE, Tripp M (1967) Lysozyme in the hemolymph of the oyster, *Crassostrea virginica*. *J Invertebr Pathol,* **9**, 531-535.

McHenery JG, Birkbeck TH, Allen JA (1979) The occurrence of lysozyme in marine bivalves. *Comp Biochem Physiol Part B: Comp Biochem,* **63**, 25-28.
McKenzie HA, White FH, Jr. (1991) Lysozyme and alpha-lactalbumin: structure, function, and interrelationships. *Adv Protein Chem,* **41**, 173-315.
Medzhitov R, Janeway CA, Jr. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell,* **91**, 295-298.
Mei J, Yue HM, Li Z, Chen B, Zhong JX, Dan C, Zhou L, Gui JF (2014) C1q-like Factor, a Target of miR-430, Regulates Primordial Germ Cell Development in Early Embryos of *Carassius auratus*. *Int J Biol Sci,* **10**, 15-24.
Meng FX, Sun YN, Liu XZ, Wang JX, Xu TJ, Wang RX, Marc RR (2012) Analysis of C3 Suggests Three Periods of Positive Selection Events and Different Evolutionary Patterns between Fish and Mammals. *Plos One,* **7**, e37489.
Merle NS, Noe R, Halbwachs-Mecarelli L, Fremeaux-Bacchi V, Roumenina LT (2015) Complement System Part II: Role in Immunity. *Front Immunol,* **6**, 257.
Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res,* **43**, W566-570.
Meynard CN, Mouillot D, Mouquet N, Douzery EJ (2012) A phylogenetic perspective on the evolution of Mediterranean teleost fishes. *Plos One,* **7**, e36443.
Migaud H, Bell G, Cabrita E, McAndrew B, Davie A, Bobe J, Herraez MP, Carrillo M (2013) Gamete quality and broodstock management in temperate fish. *Rev Aquacult,* **5**, S194-S223.
Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *grid computing environments,* pp. 1-8.
Minagawa S, Hikima J, Hirono I, Aoki T, Mori H (2001) Expression of Japanese flounder c-type lysozyme cDNA in insect cells. *Dev Comp Immunol,* **25**, 439-445.
Mishima Y, Giraldez AJ, Takeda Y, Fujiwara T, Sakamoto H, Schier AF, Inoue K (2006) Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr Biol,* **16**, 2135-2142.
Mishima Y, Tomari Y (2016) Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol Cell,* **61**, 874-885.
Modig C, Modesto T, Canario A, Cerda J, von Hofsten J, Olsson PE (2006) Molecular characterization and expression pattern of zona pellucida proteins in gilthead seabream (*Sparus aurata*). *Biol Reprod,* **75**, 717-725.
Mol S, Turan S (2008) Comparison of proximate, fatty acid and amino acid compositions of various types of fish roes. *Int J Food Prop,* **11**, 669-677.
Mommens M, Fernandes JM, Tollefsen KE, Johnston IA, Babiak I (2014) Profiling of the embryonic Atlantic halibut (*Hippoglossus hippoglossus* L.) transcriptome reveals maternal transcripts as potential markers of embryo quality. *BMC Genomics,* **15**, 829.
Mommsen TP, Vijayan MM, Moon TW (1999) Cortisol in teleosts: dynamics, mechanisms of action, and metabolic regulation. *Rev Fish Biol Fisher,* **9**, 211-268.
Monk PN, Scola AM, Madala P, Fairlie DP (2007) Function, structure and therapeutic potential of complement C5a receptors. *Br J Pharmacol,* **152**, 429-448.
Moravec CE, Voit GC, Otterlee J, Pelegri F (2021) Identification of maternal-effect genes in zebrafish using maternal crispants. *Development,* **148**, dev199536.
Morretti A, Pedini FCM, Cittolin G, Guidastri R (1999) *Manual on hatchery production of seabass and gilthead seabream,* Food Agricult Org.
Mosconi G, Carnevali O, Habibi HR, Sanyal R, Polzonetti-Magni AM (2002) Hormonal mechanisms regulating hepatic vitellogenin synthesis in the gilthead sea bream, *Sparus aurata*. *Am J Physiol-Cell Physiol,* **283**, C673-C678.
Mugwanya M, Dawood MAO, Kimera F, Sewilam H (2022) Anthropogenic temperature fluctuations and their effect on aquaculture: a comprehensive review. *Aquacult Fish,* **7**, 223-243.
Mulero I, García-Ayala A, Meseguer J, Mulero V (2007) Maternal transfer of immunity and ontogeny of autologous immunocompetence of fish: a minireview. *Aquaculture,* **268**, 244-250.
Muller EHJ (1988) Molecular organization and function of the complement system. *Annu Rev Biochem,* **57**, 321-347.
Muniesa A, Furones D, Rodgers C, Basurco B (2022) An assessment of health management and biosecurity procedures in marine fish farming in Spain. *Aquacult Rep,* **25**, 101199.
Murakami Y, Horibe T, Kinoshita M (2019) Development of an efficient bioreactor system for delivering foreign proteins secreted from liver into eggs with a vitellogenin signal in medaka *Oryzias latipes*. *Fisheries Sci,* **85**, 677-685.
Murata K, Sugiyama H, Yasumasu S, Iuchi I, Yasumasu I, Yamagami K (1997) Cloning of cDNA and estrogen-induced hepatic gene expression for choriogenin H, a precursor protein of the fish egg envelope (chorion).

P Natl Acad Sci USA, **94**, 2050-2055.
Murray CK, Fletcher TC (1976) The immunohistochemical localization of lysozyme in plaice (*Pleuronectes platessa* L.) tissues. *J Fish Biol,* **9**, 329-334.
Murray F, Bostock J, Fletcher D (2014) Review of recirculation aquaculture system technologies and their commercial application.
Mustika R (2022) Profitability In Fish Breeding Business: A Case Study Of As-Syifa Pond Fisheries Business. *Enrich J Manage,* **12**, 1686-1695.
Nagasawa T, Kawaguchi M, Sano K, Yasumasu S (2015) Sturgeon Hatching Enzyme and the Mechanism of Egg Envelope Digestion: Insight Into Changes in the Mechanism of Egg Envelope Digestion During the Evolution of Ray-Finned Fish. *J Exp Zool Part B,* **324**, 720-732.
Nagler JJ (2000) In vivo treatment with cycloheximide or actinomycin D inhibits early embryonic development in rainbow trout (*Oncorhynchus mykiss*). *Fish Physiol Biochem,* **22**, 61-66.
Najafpour B, Cardoso JCR, Canario AVM, Power DM (2020) Specific Evolution and Gene Family Expansion of Complement 3 and Regulatory Factor H in Fish. *Front Immunol,* **11**, 568631.
Najafpour B, Pinto PIS, Moutou KA, Canario AVM, Power DM (2021) Factors Driving Bacterial Microbiota of Eggs from Commercial Hatcheries of European Seabass and Gilthead Seabream. *Microorganisms,* **9**, 2275.
Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res,* **17**, 1254-1265.
Namgung J, Mizuta H, Yamaguchi Y, Nagata J, Todo T, Yilmaz O, Hiramatsu N (2021) Knock out of a major vitellogenin receptor gene with eight ligand binding repeats in medaka (*Oryzias latipes*) using the CRISPR/Cas9 system. *Comp Biochem Physiol Part A-Mol Integr Physiol,* **257**, 110967.
Naraine R, Iegorova V, Abaffy P, Franek R, Soukup V, Psenicka M, Sindelka R (2022) Evolutionary conservation of maternal RNA localization in fishes and amphibians revealed by TOMO-Seq. *Dev Biol,* **489**, 146-160.
Nayak SP, Mohanty BR, Mishra J, Rauta PR, Das A, Eknath AE, Sahoo PK (2011) Ontogeny and tissue-specific expression of innate immune related genes in rohu, *Labeo rohita* (Hamilton). *Fish Shellfish Immunol,* **30**, 1197-1201.
Nelson HM, Coffing GC, Chilson S, Hester K, Carillo C, Ostreicher S, Tomamichel W, Hanlon S, Burns AR, Lafontant PJ (2019) Structure, Development, and Functional Morphology of the Cement Gland of the Giant Danio, *Devario malabaricus*. *Dev Dyn*.
Ni SS, Zhou Y, Song LL, Chen Y, Wang X, Du XY, Zhang SC (2021) ELAVL1a is an immunocompetent protein that protects zebrafish embryos from bacterial infection. *Commun Biol,* **4**, 1-13.
Nieto D, Norambuena R, González E, González L, Brett D (2010) Smolt production systems in Chile. Analysis of alternatives from the environmental, sanitary, and economic perspectives. *World Wildlife Fund, Valdivia*.
Niksirat H, Andersson L, Golpour A, Chupani L, James P (2017) Quantification of egg proteome changes during fertilization in sterlet *Acipenser ruthenus*. *Biochem Bioph Res Co,* **490**, 189-193.
Niksirat H, Levander F, Kouba A, James P (2020) Proteomic changes after fertilization and before first cleavage in rainbow trout *Oncorhynchus mykiss*. *Aquaculture,* **520**, 734951.
Nipkow M, Wirthgen E, Luft P, Rebl A, Hoeflich A, Goldammer T (2018) Characterization of igf1 and igf2 genes during maraena whitefish (*Coregonus maraena*) ontogeny and the effect of temperature on embryogenesis and igf expression. *Growth Horm IGF Res,* **40**, 32-43.
Nissling A (2004) Effects of temperature on egg and larval survival of cod (*Gadus morhua*) and sprat (*Sprattus sprattus*) in the Baltic Sea - implications for stock development. *Hydrobiologia,* **514**, 115-123.
Nitta K, Sugai S (1989) The Evolution of Lysozyme and Alpha-Lactalbumin. *Eur J Biochem,* **182**, 111-118.
Nitta K, Tsuge H, Shimazaki K, Sugai S (1988) Calcium-binding lysozymes. *Biol Chem Hoppe Seyler,* **369**, 671-675.
Okada K, Takase R, Hamaoka Y, Honda A, Ikeda A, Hokazono Y, Maeda Y, Hayasaka O, Kotani T, Komatsu M, Shiozaki K (2020) Establishment and characterization of Neu1-knockout zebrafish and its abnormal clinical phenotypes. *Biochem J,* **477**, 2841-2857.
Oko AO (2019) Comparative Analyses of the Nutritional Composition of Fish, Fish Eggs and Fowl Eggs. *American-Eurasian J. Agric. & Environ. Sci,* **19**, 242-248.
Okomoda VT, Koh ICC, Hassan A, Amornsakun T, Shahreza MS (2017) Embryonic and larvae development of reciprocal crosses between *Pangasianodon hypophthalmus* (Sauvage, 1878) and *Clarias gariepinus* (Burchell, 1822). *Egypt J Aquat Res,* **43**, 321-327.
Oldham MC, Horvath S, Geschwind DH (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A,* **103**, 17973-17978.
Oliveros JC (2015) Venny. An interactive tool for comparing lists with Venn's diagrams.
Ostbye TK, Wetten OF, Tooming-Klunderud A, Jakobsen KS, Yafe A, Etzioni S, Moen T, Andersen O (2007) Myostatin (MSTN) gene duplications in Atlantic salmon (*Salmo salar*): evidence for different selective pressure on teleost MSTN-1 and -2. *Gene,* **403**, 159-169.

Oyarbide U, Rainieri S, Pardo MA (2012) Zebrafish (*Danio rerio*) larvae as a system to test the efficacy of polysaccharides as immunostimulants. *Zebrafish,* **9**, 74-84.

Ozturk DK, Baki B, Ozturk R, Karayucel S, Goren GU (2019) Determination of growth performance, meat quality and colour attributes of large rainbow trout (*Oncorhynchus mykiss*) in the southern Black Sea coasts of Turkey. *Aquacult Res,* **50**, 3763-3775.

Pachoensuk T, Fukuyo T, Rezanujjaman M, Wanlada K, Yamamoto C, Maeno A, Rahaman MM, Ali MH, Tokumoto T (2021) Zebrafish *stm* is involved in the development of otoliths and of the fertilization envelope. *Reprod Fertil,* **2**, 7-16.

Palomino J, Gomez C, Otarola MT, Dettleff P, Patino-Garcia D, Orellana R, Moreno RD (2021) Embryo Buoyancy and Cell Death Gene Expression During Embryogenesis of Yellow-Tail Kingfish *Seriola lalandi*. *Front Cell Dev Biol,* **9**.

Panase P, Saenphet S, Saenphet K (2017) Visceral and serum lysozyme activities in some freshwater fish (three catfish and two carps). *Comp Clin Pathol,* **26**, 169-173.

Papakostas S, Vasemagi A, Vaha JP, Himberg M, Peil L, Primmer CR (2012) A proteomics approach reveals divergent molecular responses to salinity in populations of European whitefish (*Coregonus lavaretus*). *Mol Ecol,* **21**, 3516-3530.

Parry RM, Jr., Chandan RC, Shahani KM (1965) A Rapid and Sensitive Assay of Muramidase. *Proc Soc Exp Biol Med,* **119**, 384-386.

Pauletto M, Manousaki T, Ferraresso S, Babbucci M, Tsakogiannis A, Louro B, Vitulo N, Quoc VH, Carraro R, Bertotto D, Franch R, Maroso F, Aslam ML, Sonesson AK, Simionati B, Malacrida G, Cestaro A, Caberlotto S, Sarropoulou E, Mylonas CC, Power DM, Patarnello T, Canario AVM, Tsigenopoulos C, Bargelloni L (2018) Genomic analysis of *Sparus aurata* reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. *Commun Biol,* **1**, 1-13.

Pavlidis MA, Mylonas CC (2011) *Sparidae: Biology and aquaculture of gilthead sea bream and other species,* John Wiley & Sons.

Pavlov DA, Moksness E (1995) Development of wolffish eggs at different temperature regimes. *Aquacult Int,* **3**, 315-335.

Pelegri F (2003) Maternal factors in zebrafish development. *Dev Dyn,* **228**, 535-554.

Perera E, Rosell-Moll E, Naya-Catala F, Simo-Mirabet P, Calduch-Giner J, Perez-Sanchez J (2021) Effects of genetics and early-life mild hypoxia on size variation in farmed gilthead sea bream (*Sparus aurata*). *Fish Physiol Biochem,* **47**, 121-133.

Perez-Riverol Y, Bai JW, Bandla C, Garcia-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang SB, Brazma A, Vizcaino JA (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res,* **50**, D543-D552.

Pettersson-Kastberg J, Aits S, Gustafsson L, Mossberg A, Storm P, Trulsson M, Persson F, Mok KH, Svanborg C (2009) Can misfolded proteins be beneficial? The HAMLET case. *Ann Med,* **41**, 162-176.

Pickering AD (1993) Growth and Stress in Fish Production. *Aquaculture,* **111**, 51-63.

Pierce AL, Breves JP, Moriyama S, Hirano T, Grau EG (2011) Differential regulation of Igf1 and Igf2 mRNA levels in tilapia hepatocytes: effects of insulin and cortisol on GH sensitivity. *J Endocrinol,* **211**, 201-210.

Pinto PI, Anjos L, Estevao MD, Santos S, Santa C, Manadas B, Monsinjon T, Canario AVM, Power DM (2022) Proteomics of sea bass skin-scales exposed to the emerging pollutant fluoxetine compared to estradiol. *Sci Total Environ,* **814**, 152671.

Pinto PIS, Guerreiro CC, Costa RA, Martinez-Blanch JF, Carballo C, Codoner FM, Manchado M, Power DM (2019) Understanding pseudo-albinism in sole (*Solea senegalensis*): a transcriptomics and metagenomics approach. *Sci Rep-Uk,* **9**.

Pinto PIS, Thorne MAS, Power DM (2017) European sea bass (*Dicentrarchus labrax*) skin and scale transcriptomes. *Mar Genom,* **35**, 35-37.

Pipe RK (1990) Hydrolytic enzymes associated with the granular haemocytes of the marine mussel *Mytilus edulis*. *Histochem J,* **22**, 595-603.

Planas M, Labarta U, Fernandez-Reiriz MJ, Ferreiro MJ, Munilla R, Garrido JL (1993) Chemical changes during development in turbot (*Scophthalmus maximus*) eggs and larvae. In: *Physiological and biochemical aspects of fish development* (ed. by Walter BT, Fyhn HJ), University of Bergen, Bergen, Norway, pp. 269-278.

Pottinger TG (2008) *The stress response in fish-mechanisms, effects and measurement*, 32-48 p.

Poupard G, Andre M, Durliat M, Ballagny C, Boeuf G, Babin PJ (2000) Apolipoprotein E gene expression correlates with endogenous lipid nutrition and yolk syncytial layer lipoprotein synthesis during fish development. *Cell Tissue Res,* **300**, 251-261.

Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature,* **465**, 922-926.

Prager EM, Jolles P (1996) Animal lysozymes c and g: an overview. *Exs,* **75**, 9-31.
Preena PG, Kumar VJR, Singh ISB (2021) Nitrification and denitrification in recirculating aquaculture systems: the processes and players. *Rev Aquacult,* **13**, 2053-2075.
Pulido-Rodriguez LF, Cardinaletti G, Secci G, Randazzo B, Bruni L, Cerri R, Olivotto I, Tibaldi E, Parisi G (2021) Appetite Regulation, Growth Performances and Fish Quality Are Modulated by Alternative Dietary Protein Ingredients in Gilthead Sea Bream (*Sparus aurata*) Culture. *Animals,* **11**, 1919.
Purushothaman K, Das PP, Presslauer C, Lim TK, Johansen SD, Lin QS, Babiak I (2019) Proteomics Analysis of Early Developmental Stages of Zebrafish Embryos. *Int J Mol Sci,* **20**, 1-14.
Qasba PK, Kumar S (1997) Molecular divergence of lysozymes and alpha-lactalbumin. *Crit Rev Biochem Mol Biol,* **32**, 255-306.
Ramachandra RK, Salem M, Gahr S, Rexroad CE, Yao J (2008) Cloning and characterization of microRNAs from rainbow trout (*Oncorhynchus mykiss*): Their expression during early embryonic development. *Bmc Dev Biol,* **8**, 1-11.
Ramakrishnaiah Y, Kuhlmann L, Tyagi S (2020) Towards a comprehensive pipeline to identify and functionally annotate long noncoding RNA (lncRNA). *Comput Biol Med,* **127**, 104028.
Ramakrishnaiah Y, Kuhlmann L, Tyagi S (2021) linc2function: A deep learning model to identify and assign function to long noncoding RNA (lncRNA). *bioRxiv*.
Rana S, Hasan MN, Bari AA, Shimul SA, Ahmed SI, Nahid SAA (2022) Problems and prospects of fish farming in the Chattogram Hill Tracts of Bangladesh: Community-based aquaculture might be a right choice. *Aquacult Fish,* **In press**.
Ravi V, Venkatesh B (2008) Rapidly evolving fish genomes and teleost diversity. *Curr Opin Genet Dev,* **18**, 544-550.
Rawat VS, Rani KV, Phartyal R, Sehgal N (2013) Vitellogenin genes in fish: differential expression on exposure to estradiol. *Fish Physiol Biochem,* **39**, 39-46.
Raya A, Koth CM, Buscher D, Kawakami Y, Itoh T, Raya RM, Sternik G, Tsai HJ, Rodriguez-Esteban C, Izpisua-Belmonte JC (2003) Activation of Notch signaling pathway precedes heart regeneration in zebrafish. *P Natl Acad Sci USA,* **100**, 11889-11895.
Read TD, Petit RA, 3rd, Joseph SJ, Alam MT, Weil MR, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb DH, Tan M, Dove ADM (2017) Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. *BMC genomics,* **18**, 532.
Reading BJ, Andersen LK, Ryu YW, Mushirobira Y, Todo T, Hiramatsu N (2018) Oogenesis and Egg Quality in Finfish: Yolk Formation and Other Factors Influencing Female Fertility. *Fishes-Basel,* **3**, 45.
Ren F, Lin QH, Gong GR, Du X, Dan H, Qin WY, Miao R, Xiong Y, Xiao R, Li XH, Gui JF, Mei J (2020) Igf2bp3 maintains maternal RNA stability and ensures early embryo development in zebrafish. *Commun Biol,* **3**, 1-10.
Rhodes MA, Phelps RP (2008) Evaluation of the ciliated protozoa, *Fabrea salina* as a first food for larval red snapper, *Lutjanus campechanus* in a large scale rearing experiment. *J Appl Aquacult,* **20**, 120-133.
Richter JD (1999) Cytoplasmic polyadenylation in development and beyond. *Microbiol Mol Biol R,* **63**, 446-456.
Ricklin D, Hajishengallis G, Yang K, Lambris JD (2010) Complement: a key system for immune surveillance and homeostasis. *Nat Immunol,* **11**, 785-797.
Riehl R, Patzner RA (1998) Minireview: The modes of egg attachment in teleost fishes. *Ital J Zool,* **65**, 415-420.
Rise ML, Hall JR, Alcock BP, Hori TS (2012) Dynamic expression profiles of virus-responsive and putative antimicrobial "peptide-encoding transcripts during Atlantic cod (*Gadus morhua*) embryonic and. early larval development. *Gene,* **509**, 232-246.
Rise ML, Nash GW, Hall JR, Boornan M, Honi TS, Trippel EA, Gamperl AK (2014) Variation in embryonic mortality and maternal transcript expression among Atlantic cod (*Gadus morhua*) broodstock: A functional genomics study. *Mar Genom,* **18**, 3-20.
Rizzo E, Bazzoli N (1991) The zona pellucida of the Brazilian white piranha, *Serrasalmus brandtii* Reinhardt 1874 (Pisces, Characidae): a cytological and cytochemical study. *Funct Dev Morphol,* **1**, 21-24.
Rizzo E, Sato Y, Barreto BP, Godinho HP (2002) Adhesiveness and surface patterns of eggs in neotropical freshwater teleosts. *J Fish Biol,* **61**, 615-632.
Robertson N, Rappas M, Dore AS, Brown J, Bottegoni G, Koglin M, Cansfield J, Jazayeri A, Cooke RM, Marshall FH (2018) Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature,* **553**, 111-114.
Robledo D, Palaiokostas C, Bargelloni L, Martinez P, Houston R (2018) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev Aquacult,* **10**, 670-682.
Romano M, Rosanova P, Anteo C, Limatola E (2004) Vertebrate yolk proteins: A review. *Mol Reprod Dev,* **69**, 109-116.
Romero H, Zavala A, Musto H, Bernardi G (2003) The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene,* **317**, 141-147.

Romney AL, Podrabsky JE (2017) Transcriptomic analysis of maternally provisioned cues for phenotypic plasticity in the annual killifish, *Austrofundulus limnaeus*. *Evodevo,* **8**, 1-20.
Rønnestad I, Fyhn HJ (1993) Metabolic aspects of free amino acids in developing marine fish eggs and larvae. *Rev Fish Sci,* **1**, 239-259.
Rønnestad I, Lie Ø, Waagbø R (1997) Vitamin B6 in Atlantic halibut, *Hippoglossus hippoglossus*—endogenous utilization and retention in larvae fed natural zooplankton. *Aquaculture,* **157**, 337-345.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol,* **61**, 539-542.
Rurangwa E, Poelman M (2011) Hatchery manual for broodstock management and larval production of tubrot (*Psetta maxima*). IMARES*,* pp. 1-52.
Saborido-Rey F, Kjesbu OS, Thorsen A (2003) Buoyancy of Atlantic cod larvae in relation to developmental stage and maternal influences. *J Plankton Res,* **25**, 291-307.
Sahin T, Gunes E, Aydin I, Polat H (2008) Reproductive characteristics and egg development in flounder (*Pleuronectes flesus luscus*) in the southern Black Sea. *Isr J Aquacult-Bamid,* **60**, 20-26.
Sahoo S, Banu H, Prakash A, Tripathi G (2021) Immune System of Fish: An Evolutionary Perspective. In: *Antimicrobial Immune Response*. IntechOpen.
Saillant EA, Leclercq E, Bardon-Albaret A, Sarkisian B, Apeitos A, Brown-Peterson NJ, Lemus JT, Buentello A, Pohlenz C, Gatlin DM (2013) Development of aquaculture of the red snapper *Lutjanus campechanus*: research on larval nutrition. 352-356.
Saleh M, Kumar G, Abdel-Baki AAS, Dkhil MA, El-Matbouli M, Al-Quraishy S (2019) Quantitative proteomic profiling of immune responses to *Ichthyophthirius multifiliis* in common carp skin mucus. *Fish Shellfish Immunol,* **84**, 834-842.
Sano K, Kawaguchi M, Watanabe S, Yasumasu S (2014) Neofunctionalization of a duplicate hatching enzyme gene during the evolution of teleost fishes. *Bmc Evol Biol,* **14**, 1-15.
Santa C, Anjo SI, Manadas B (2016) Protein precipitation of diluted samples in SDS-containing buffer with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone approach. *Proteomics,* **16**, 1847-1851.
Santos JAD, Soares CM, Bialetzki A (2020) Effects of pH on the incubation and early development of fish species with different reproductive strategies. *Aquat Toxicol,* **219**, 105382.
Sargent JR (1995) Origins and function of eggs lipids: Nutritional implication. In: *Broodstock management and egg and larval quality* (ed. by Bromage N, Roberts RJ), Blackwell Science, Cambridge, UK*,* pp. 353-372.
Sargent JR, Henderson RJ, Tocher DR (1989) The lipids. In: *Fish Nutrition* (ed by Halver JE), Academic Press, New York*,* pp. 153–218.
Saurabh S, Sahoo PK (2008) Lysozyme: an important defence molecule of fish innate immune system. *Aquacult Res,* **39**, 223-239.
Sawaguchi S, Kagawa H, Ohkubo N, Hiramatsu N, Sullivan CV, 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. *Mol Reprod Dev,* **73**, 719-736.
Scapigliati G, Meloni S, Mazzini M (1999) A monoclonal antibody against chorion proteins of the sea bass *Dicentrarchus labrax* (Linnaeus, 1758): Studies of chorion precursors and applicability in immunoassays. *Biol Reprod,* **60**, 783-789.
Schisa JA (2012) New Insights into the Regulation of Rnp Granule Assembly in Oocytes. *Int Rev Cel Mol Bio,* **295**, 233-289.
Schreck CB, Contreras-Sanchez W, Fitzpatrick MS (2001) Effects of stress on fish reproduction, gamete quality, and progeny. *Aquaculture,* **197**, 3-24.
Sennels L, Bukowski-Wills JC, Rappsilber J (2009) Improved results in proteomics by use of local and peptide-class specific false discovery rates. *Bmc Bioinf,* **10**, 1-11.
Seoka M, Yamada S, Iwata Y, Yanagisawa T, Nakagawa T, Kumai H (2003) Differences in the biochemical content of buoyant and non-buoyant eggs of the Japanese eel, *Anguilla japonica*. *Aquaculture,* **216**, 355-362.
Seppola M, Bakkemo KR, Mikkelsen H, Myrnes B, Helland R, Irwin DM, Nilsen IW (2016) Multiple specialised goose-type lysozymes potentially compensate for an exceptional lack of chicken-type lysozymes in Atlantic cod. *Sci Rep,* **6**, 28318.
Seppola M, Johnsen H, Mennen S, Myrnes B, Tveiten H (2009) Maternal transfer and transcriptional onset of immune genes during ontogenesis in Atlantic cod. *Dev Comp Immunol,* **33**, 1205-1211.
Serrano V, Bernay B, Moreno-Legorreta M, Ortega-Rubio A, Maldonado-Garcia M (2020) Egg proteomic characterization of *Seriola rivoliana* in captivity. *Lat Am J Aquat Res,* **48**, 877-885.
Servili A, Canario AVM, Mouchel O, Munoz-Cueto JA (2020) Climate change impacts on fish reproduction are mediated at multiple levels of the brain-pituitary-gonad axis. *Gen Comp Endocr,* **291**, 113439.
Shabanipour N, Heidari B (2017) A histological study of the zona radiata during late oocyte developmental stages

in the Caspian Sea mugilid, *Liza aurata* (Risso 1810). *J Morphol Sci,* **21**, 191-195.
Shaik AA, Wee S, Li RHX, Li Z, Carney TJ, Mathavan S, Gunaratne J (2014) Functional Mapping of the Zebrafish Early Embryo Proteome and Transcriptome. *J Proteome Res,* **13**, 5536-5550.
Shao C, Bao B, Xie Z, Chen X, Li B, Jia X, Yao Q, Orti G, Li W, Li X, Hamre K, Xu J, Wang L, Chen F, Tian Y, Schreiber AM, Wang N, Wei F, Zhang J, Dong Z, Gao L, Gai J, Sakamoto T, Mo S, Chen W, Shi Q, Li H, Xiu Y, Li Y, Xu W, Shi Z, Zhang G, Power DM, Wang Q, Schartl M, Chen S (2017) The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet,* **49**, 119-124.
Sharp PM, Li WH (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res,* **15**, 1281-1295.
Shen MH, Di GL, Li M, Fu JQ, Dai Q, Miao XL, Huang MQ, You WW, Ke CH (2018) Proteomics Studies on the three Larval Stages of Development and Metamorphosis of *Babylonia areolata* (vol 8, 6269, 2018). *Sci Rep-Uk,* **8**.
Shi SL, Stanley P (2006) Evolutionary origins of Notch signaling in early development. *Cell Cycle,* **5**, 274-278.
Shields RJ, Brown NP, Bromage NR (1997) Blastomere morphology as a predictive measure of fish egg viability. *Aquaculture,* **155**, 1-12.
Shields RJ, Kotani T, Molnar A, Marion K, Kobashigawa J, Tang L (2005) Intensive cultivation of a subtropical paracalanid copepod, Parvocalanus sp., as prey for small marine fish larvae. *Copepods Aquacult*, 209-223.
Shih DF, Hsiao CD, Min MY, Lai WS, Yang CW, Lee WT, Lee SJ (2013) Aromatic L-Amino Acid Decarboxylase (AADC) Is Crucial for Brain Development and Motor Functions. *Plos One,* **8**, e71741.
Shin SC, Ahn DH, Kim SJ, Pyo CW, Lee H, Kim M-K, Lee J, Lee JE, Detrich HW, Postlethwait JH, Edwards D, Lee SG, Lee JH, Park H (2014) The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol,* **15**, 468.
Shirota M, Ishida T, Kinoshita K (2009) Development of a new meta-score for protein structure prediction from seven all-atom distance dependent potentials using support vector regression. In: *Genome Informatics 2009: Genome Informatics Series Vol. 23*. World Scientific, pp. 149-158.
Shixi C, Wanshu H, Qiyong Z, Renxie W, Qiong W (2006) Rates of oxygen consumption and tolerance of hypoxia and desiccation in Chinese black sleeper (*Bostrichthys sinensis*) and mudskipper (*Boleophthalmus pectinirostris*) embryos. *Acta Oceanol Sin-English Edition,* **25**, 91-98.
Sieger D, Stein C, Neifer D, van der Sar AM, Leptin M (2009) The role of gamma interferon in innate immunity in the zebrafish embryo. *Dis Model Mech,* **2**, 571-581.
Siitonen L, Gall GAE (1989) Response to Selection for Early Spawn Date in Rainbow Trout, *Salm Gairdneri*. *Aquaculture,* **78**, 153-161.
Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts A, Otillar RP, Terry AY, Boore JL, Grigoriev IV, Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS (2013) Insights into bilaterian evolution from three spiralian genomes. *Nature,* **493**, 526-531.
Singh SK, Sundaram CS, Shanbhag S, Idris MM (2010) Proteomic profile of zebrafish brain based on two-dimensional gel electrophoresis matrix-assisted laser desorption/ionization MS/MS analysis. *Zebrafish,* **7**, 169-177.
Skjermo J, Vadstein O (1999) Techniques for microbial control in the intensive rearing of marine larvae. *Aquaculture,* **177**, 333-343.
Smith GR, Sternberg MJ (2002) Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol,* **12**, 28-35.
Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SMJ, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson Y-W, Nanlohy KG, Libants SV, Yeh C-Y, McCauley DW, Langeland JA, Pancer Z, Fritzsch B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W (2013) Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet,* **45**, 415-421, 421e411-412.
Sohn JH, Kim DH (2021) Ultrastructure of the fertilized egg envelope from *Melanotaenia praecox*, Melanotaeniidae, Teleostei. *Appl Microsc,* **51**, 3.
Speers-Roesch B, Richards JG, Brauner CJ, Farrell AP, Hickey AJR, Wang YS, Renshaw GMC (2012) Hypoxia tolerance in elasmobranchs. I. Critical oxygen tension as a measure of blood oxygen transport during hypoxia exposure. *J Exp Biol,* **215**, 93-102.
Stehr CM, Hawkes JW (1979) Comparative Ultrastructure of the Egg Membrane and Associated Pore Structures in the Starry Flounder, *Platichthys-Stellatus* (Pallas), and Pink Salmon, *Oncorhynchus-Gorbuscha*

(Walbaum). *Cell Tissue Res,* **202**, 347-356.
Subbotkina T, Subbotkin M, Ha VT (2018) Comparative lysozyme analysis in carp fishes (Cyprinidae, Cypriniformes). *Труды Института биологии внутренних вод РАН*.
Sullivan CV, Yilmaz O (2018) Vitellogenesis and Yolk Proteins, Fish. *Encyclopedia of Reproduction (Second Edition),* **6**, 266-277.
Sun JW, Yan L, Shen WM, Meng AM (2018) Maternal Ybx1 safeguards zebrafish oocyte maturation and maternal-to-zygotic transition by repressing global translation. *Development,* **145**, dev166587.
Sun LL, Bertke MM, Champion MM, Zhu GJ, Huber PW, Dovichi NJ (2014) Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. *Sci Rep-Uk,* **4**.
Sundby S (1991) Factors Affecting the Vertical-Distribution of Eggs. *Ices Mar Sc,* **192**, 33-38.
Sundby S, Kristiansen T (2015) The Principles of Buoyancy in Marine Fish Eggs and Their Vertical Distributions across the World Oceans. *Plos One,* **10**.
Supungul P, Rimphanitchayakit V, Aoki T, Hirono I, Tassanakajon A (2010) Molecular characterization and expression analysis of a c-type and two novel muramidase-deficient i-type lysozymes from Penaeus monodon. *Fish Shellfish Immunol,* **28**, 490-498.
Sussarellu R, Suquet M, Thomas Y, Lambert C, Fabioux C, Pernet MEJ, Le Goic N, Quillien V, Mingant C, Epelboin Y, Corporeau C, Guyomarch J, Robbens J, Paul-Pont I, Soudant P, Huvet A (2016) Oyster reproduction is affected by exposure to polystyrene microplastics. *P Natl Acad Sci USA,* **113**, 2430-2435.
Sutin S, Tina FW (2020) Embryonic development of a mountainous fish species *Garra cambodgiensis* (Tirant, 1883) in southern Thailand. *J Anim Behav Biomete,* **8**, 88-94.
Sveinsdottir H, Vilhelmsson O, Gudmundsdottir A (2008) Proteome analysis of abundant proteins in two age groups of early Atlantic cod (*Gadus morhua*) larvae. *Comp Biochem Physiol Part D Genomics Proteomics,* **3**, 243-250.
Swain P, Nayak SK (2009) Role of maternally derived immunity in fish. *Fish Shellfish Immunol,* **27**, 89-99.
Swain P, Nayak SK, Sahu A, Mohapatra BC, Meher PK (2002) Bath immunisation of spawn, fry and fingerlings of Indian major carps using a particulate bacterial antigen. *Fish Shellfish Immunol,* **13**, 133-140.
Swaminathan A, Gliksberg M, Anbalagan S, Wigoda N, Levkowitz G (2022) Stress resilience is established during development and is regulated by complement factors. *bioRxiv*.
Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res,* **47**, D607-D613.
Tadros W, Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts. *Development,* **136**, 3033-3042.
Tao WJ, Yang L, Mayden RL, He SP (2019) Phylogenetic relationships of Cypriniformes and plasticity of pharyngeal teeth in the adaptive radiation of cyprinids. *Sci China Life Sci,* **62**, 553-565.
Taranger GL, Haux C, Stefansson SO, Bjornsson BT, Walther BT, Hansen T (1998) Abrupt changes in photoperiod affect age at maturity, timing of ovulation and plasma testosterone and oestradiol-17 beta profiles in Atlantic salmon, Salmo salar. *Aquaculture,* **162**, 85-98.
Taslima K, Al-Emran M, Rahman MS, Hasan J, Ferdous Z, Rohani MF, Shahjahan M (2022) Impacts of heavy metals on early development, growth and reproduction of fish–a review. *Toxicol Rep,* **9**, 858-868.
Tata JR (1986) Coordinated assembly of the developing egg. *Bioessays,* **4**, 197-201.
Tay TL, Lin QS, Seow TK, Tan KH, Hew CL, Gong ZY (2006) Proteomic analysis of protein profiles during early development of the zebrafish, *Danio rerio*. *Proteomics,* **6**, 3176-3188.
Thorsen A, Fyhn HJ (1991) Osmotic effectors during preovulatory swelling of marine fish eggs. In: *Proceedings of the fourth international symposium on the reproductive physiology of fish, Sheffield* (eds Scott AP, Sumpter JP, Kime DE, Rolfe MS)*,* pp. 312-314.
Tieri P, Farina L, Petti M, Astolfi L, Castiglione F (2018) *Network Inference and Reconstruction in Bioinformatics*.
Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierne N, Volckaert FAM, Bargelloni L, Power DM, Bonhomme F, Canario AVM, Reinhardt R (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun,* **5**, 5770.
Tingaud-Sequeira A, Cerda J (2007) Phylogenetic relationships and gene expression pattern of three different cathepsin L (Ctsl) isoforms in zebrafish: Ctsla is the putative yolk processing enzyme. *Gene,* **386**, 98-106.
Torrans L, Steeby J (2008) Effects of dissolved oxygen concentration on oxygen consumption and development of channel catfish eggs and fry: Implications for hatchery management. *N Am J Aquacult,* **70**, 286-295.
Tsai KL, Chen LW, Yang LJ, Shin H, Chen HW (2022) IoT based Smart Aquaculture System with Automatic Aerating and Water Quality Monitoring. *J Internet Technol,* **23**, 177-184.
Tveiten H, Jobling M, Andreassen I (2004) Influence of egg lipids and fatty acids on egg viability, and their utilization during embryonic development of spotted wolf-fish, *Anarhichas minor* Olafsen. *Aquacult Res,*

35, 152-161.
Ulitsky I, Bartel DP (2013) lincRNAs: Genomics, Evolution, and Mechanisms. *Cell,* **154**, 26-46.
Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP (2011) Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell,* **147**, 1537-1550.
Urrutia AO, Hurst LD (2003) The signature of selection mediated by expression on human genes. *Genome Res,* **13**, 2260-2264.
Valenti WC, Barros HP, Moraes-Valenti P, Bueno GW, Cavalli RO (2021) Aquaculture in Brazil: past, present and future. *Aquacult Rep,* **19**, 100611.
Valipour A, Heidari B, Hadavi M, Yousefi A (2018) Changes in immune parameters (lysozyme, IgM, C3) in early life stages and broodstock of Siberian sturgeon, *Acipenser baerii*. *Arch Pol Fish,* **26**, 21-30.
Vallee M, Aiba K, Piao Y, Palin MF, Ko MSH, Sirard MA (2008) Comparative analysis of oocyte transcript profiles reveals a high degree of conservation among species. *Reproduction,* **135**, 439-448.
van der Vaart M, Spaink HP, Meijer AH (2012) Pathogen recognition and activation of the innate immune response in zebrafish. *Adv Hematol,* **2012**, 159807.
Van Eeden FJM, Holley SA, Haffter P, Nusslein-Volhard C (1998) Zebrafish segmentation and pair-rule patterning. *Dev Genet,* **23**, 65-76.
Van Herreweghe JM, Michiels CW (2012) Invertebrate lysozymes: diversity and distribution, molecular mechanism and in vivo function. *J Biosci,* **37**, 327-348.
Vazquez DM, de Astarloa JMD, Gabbanelli V, Mabragana E (2020) Comparative embryonic development patterns in three deep-water skates from the southwest Atlantic. *Deep-Sea Res Pt I,* **161**, 103301.
Vázquez R, González S, Rodriguez A, Mourente G (1994) Biochemical composition and fatty acid content of fertilized eggs, yolk sac stage larvae and first-feeding larvae of the Senegal sole (*Solea senegalensis* Kaup). *Aquaculture,* **119**, 273-286.
Velsen F (1980) Embryonic development in eggs of sockeye salmon, *Oncorhynchus nerka*. *Can Spec Publ Fish Aquat Sci,* **49**, 1-19.
Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay B-H, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature,* **505**, 174-179.
Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J (2011) The zebrafish transcriptome during early development. *Bmc Dev Biol,* **11**, 1-18.
Vetemaa M, Saat T (1996) Effects of salinity on the development of fresh-water and brackish-water ruffe *Gymnocephalus cernuus* (L) embryos. *Ann Zool Fenn,* **33**, 687-691.
Villalba A, Coll O, Gebauer F (2011) Cytoplasmic polyadenylation and translational control. *Curr Opin Genet Dev,* **21**, 452-457.
Vogel AM, Gerster T (1997) Expression of a zebrafish Cathepsin L gene in anterior mesendoderm and hatching gland. *Dev Genes Evol,* **206**, 477-479.
Volanakis JE (1998) *The human complement system in health and disease,* Crc Press, New York: Marcel Dekker, 23 p.
Vong YH, Sivashanmugam L, Leech R, Zaucker A, Jones A, Sampath K (2021) The RNA-binding protein Igf2bp3 is critical for embryonic and germline development in zebrafish. *Plos Genet,* **17**, e1009667.
Vonwesternhagen H, Dethlefsen V, Cameron P, Berg J, Furstenberg G (1988) Developmental Defects in Pelagic Fish Embryos from the Western Baltic. *Helgolander Meeresun,* **42**, 13-36.
Wallace RA, Selman K (1981) Cellular and Dynamic Aspects of Oocyte Growth in Teleosts. *Am Zool,* **21**, 325-343.
Wang J, Fu LY, Koganti PP, Wang L, Hand JM, Ma H, Yao JB (2016a) Identification and Functional Prediction of Large Intergenic Noncoding RNAs (lincRNAs) in Rainbow Trout (*Oncorhynchus mykiss*). *Mar Biotechnol,* **18**, 271-282.
Wang L, Han G, Wang R, Chen G, Xu R, Xiao H, Li X, Geng S, Li Y, Li X, Wang J, Feng J, Riedemann NC, Guo R, Shen B, Li Y (2010) Regulation of IL-8 production by complement-activated product, C5a, in vitro and in vivo during sepsis. *Clin Immunol,* **137**, 157-165.
Wang M, Jiang S, Wu W, Yu F, Chang WG, Li PF, Wang K (2018) Non-coding RNAs Function as Immune Regulators in Teleost Fish. *Front Immunol,* **9**, 2801.
Wang Q, Wang C, Mu C, Wu H, Zhang L, Zhao J (2013a) A novel C-type lysozyme from *Mytilus galloprovincialis*: insight into innate immunity and molecular evolution of invertebrate C-type lysozymes. *Plos One,* **8**, e67469.
Wang RJ, Feng JB, Li C, Liu SK, Zhang Y, Liu ZJ (2013b) Four lysozymes (one c-type and three g-type) in catfish are drastically but differentially induced after bacterial infection. *Fish Shellfish Immunol,* **35**, 136-145.
Wang SC, Wang RX, Xu TJ (2013c) The evolutionary analysis on complement genes reveals that fishes C3 and

C9 experience different evolutionary patterns. *Fish Shellfish Immunol,* **35**, 2040-2045.
Wang SH, Wang Y, Ma J, Ding YC, Zhang SC (2011) Phosvitin Plays a Critical Role in the Immunity of Zebrafish Embryos via Acting as a Pattern Recognition Receptor and an Antimicrobial Effector. *J Biol Chem,* **286**, 22653-22664.
Wang ST, Meng XZ, Li LS, Dang YF, Fang Y, Shen Y, Xu XY, Wang RQ, Li JL (2017) Biological parameters, immune enzymes, and histological alterations in the livers of grass carp infected with *Aeromonas hydrophila*. *Fish Shellfish Immunol,* **70**, 121-128.
Wang X, Du XY, Li HY, Zhang SC (2016b) Identification of the Zinc Finger Protein ZRANB2 as a Novel Maternal Lipopolysaccharide-binding Protein That Protects Embryos of Zebrafish against Gram-negative Bacterial Infections. *J Biol Chem,* **291**, 4019-4034.
Wang Y, Lu Y, Zhang Y, Ning Z, Li Y, Zhao Q, Lu H, Huang R, Xia X, Feng Q, Liang X, Liu K, Zhang L, Lu T, Huang T, Fan D, Weng Q, Zhu C, Lu Y, Li W, Wen Z, Zhou C, Tian Q, Kang X, Shi M, Zhang W, Jang S, Du F, He S, Liao L, Li Y, Gui B, He H, Ning Z, Yang C, He L, Luo L, Yang R, Luo Q, Liu X, Li S, Huang W, Xiao L, Lin H, Han B, Zhu Z (2015) The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet,* **47**, 625-631.
Wang YD, Huang SJ, Chou HN, Liao WL, Gong HY, Chen JY (2014a) Transcriptome analysis of the effect of *Vibrio alginolyticus* infection on the innate immunity-related complement pathway in *Epinephelus coioides*. *Bmc Genomics,* **15**, 1-15.
Wang YK, Chen F, He J, Chen J, Xue G, Zhao Y, Peng YH, Xie P (2022) Comparative ultrastructure and proteomics of two economic species (common carp and grass carp) egg envelope. *Aquaculture,* **546**, 737276.
Wang YS, Bu LZ, Yang LL, Li HY, Zhang SC (2016c) Identification and functional characterization of fish-egg lectin in zebrafish. *Fish Shellfish Immunol,* **52**, 23-30.
Wang YXR, Waterman MS, Huang HY (2014b) Gene coexpression measures in large heterogeneous samples using count statistics. *P Natl Acad Sci USA,* **111**, 16371-16376.
Wang ZJ, Lu ZH, Shi P (2012) Study on toxic effects of ammonia on embryonic and yolk-sac stage larvae of rare minnow. *Huan Jing Ke Xue,* **33**, 1323-1330.
Wang ZP, Zhang SC (2010) The role of lysozyme and complement in the antibacterial activity of zebrafish (*Danio rerio*) egg cytosol. *Fish Shellfish Immunol,* **29**, 773-777.
Wang ZP, Zhang SC, Wang GF, An Y (2008) Complement Activity in the Egg Cytosol of Zebrafish *Danio rerio*: Evidence for the Defense Role of Maternal Complement Components. *Plos One,* **3**, e1463.
Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res,* **38**, W214-220.
Watanabe WO, Benetti DD, Feeley MW (2005) Status of artificial propagation of mutton, yellowtail, and red snapper (family Lutjanidae). In: *American Fisheries Society Symposium,* pp. 517-540.
Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res,* **46**, W296-W303.
Weaver LH, Grutter MG, Remington SJ, Gray TM, Isaacs NW, Matthews BW (1984) Comparison of goose-type, chicken-type, and phage-type lysozymes illustrates the changes that occur in both amino acid sequence and three-dimensional structure during evolution. *J Mol Evol,* **21**, 97-111.
Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL (2018) Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol,* **35**, 773-777.
Wei S, Huang Y, Huang X, Cai J, Wei J, Li P, Ouyang Z, Qin Q (2014) Molecular cloning and characterization of a new G-type lysozyme gene (Ec-lysG) in orange-spotted grouper, *Epinephelus coioides*. *Dev Comp Immunol,* **46**, 401-412.
Wei SN, Huang YH, Cai J, Huang XH, Fu J, Qin QW (2012) Molecular cloning and characterization of c-type lysozyme gene in orange-spotted grouper, *Epinephelus coioides*. *Fish Shellfish Immunol,* **33**, 186-196.
Witeska M, Bilska K, Sarnowski P (2010) Effects of Copper and Cadmium on Growth and Yolk Utilization in Barbel (*Barbus Barbus* L.) Larvae. *Pol J Environ Stud,* **19**, 227-230.
Wittkopp PJ (2007) Variable gene expression in eukaryotes: a network perspective. *J Exp Biol,* **210**, 1567-1575.
Wosnick N, Takatsuka V, Mello AE, Dias J, Lubitz N, de Azevedo VG (2019) Embryonic malformations in an offspring of the shortnose guitarfish. *Braz J Oceanogr,* **67**.
Woynarovich E, Horváth L (1980) Natural Propagation of Finfish. In: *The artificial propagation of warm-water finfishes - a manual for extension*. FAO Fish. Tech. Pap., pp. 183p.
Wu N, LaPatra SE, Li J, Sunyer JO, Zhang YA (2014) Complement C5a acts as molecular adjuvant in fish by enhancing antibody response to soluble antigen. *Fish Shellfish Immunol,* **40**, 616-623.
Xu HG, Cao L, Sun B, Wei YL, Liang MQ (2019a) Transcriptomic Analysis of Potential "lncRNA-mRNA" Interactions in Liver of the Marine Teleost *Cynoglossus semilaevis* Fed Diets With Different DHA/EPA

Ratios. *Front Physiol,* **10**, 331.
Xu P, McIntyre LM, Scardina J, Wheeler PA, Thorgaard GH, Nichols KM (2011) Transcriptome profiling of embryonic development rate in rainbow trout advanced backcross introgression lines. *Mar Biotechnol,* **13**, 215-231.
Xu P, Xu J, Liu GJ, Chen L, Zhou ZX, Peng WZ, Jiang YL, Zhao ZX, Jia ZY, Sun YH, Wu YD, Chen BH, Pu F, Feng JX, Luo J, Chai J, Zhang HY, Wang H, Dong CJ, Jiang WK, Sun XW (2019b) The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat Commun,* **10**, 1-11.
Yamagami K, Hamazaki TS, Yasumasu S, Masuda K, Iuchi I (1992) Molecular and Cellular Basis of Formation, Hardening, and Breakdown of the Egg Envelope in Fish. *Int Rev Cytol,* **136**, 51-92.
Yang Y, Wang L, Han X, Yang WL, Zhang MM, Ma HL, Sun BF, Li A, Xia J, Chen J, Heng J, Wu BX, Chen YS, Xu JW, Yang X, Yao H, Sun JW, Lyu C, Wang HL, Huang Y, Sun YP, Zhao YL, Meng AM, Ma JB, Liu F, Yang YG (2019) RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. *Mol Cell,* **75**, 1188-1202.
Yassien SA, Abd El-Rahim SA, Osman MF, Hamouda RE, Soliman MAM, Nageib RM (2022) Factors affecting aquaculture farms' profitability and constraints facing fish farmers in Egypt. *Egypt J Aquat Biol Fish,* **26**, 519-527.
Ye X, Gao F-Y, Zheng Q-M, Bai J-J, Wang H, Lao H-H, Jian Q (2009) Cloning and characterization of the tiger shrimp lysozyme. *Mol Biol Rep,* **36**, 1239-1246.
Yi S, Li Y, Wang W (2018) Selection shapes the patterns of codon usage in three closely related species of genus Misgurnus. *Genomics,* **110**, 134-142.
Yilmaz O, Patinote A, Com E, Pineau C, Bobe J (2021) Knock out of specific maternal vitellogenins in zebrafish (*Danio rerio*) evokes vital changes in egg proteomic profiles that resemble the phenotype of poor quality eggs. *Bmc Genomics,* **22**, 1-22.
Yilmaz O, Patinote A, Nguyen T, Bobe J (2018) Multiple vitellogenins in zebrafish (*Danio rerio*): quantitative inventory of genes, transcripts and proteins, and relation to egg quality. *Fish Physiol Biochem,* **44**, 1509-1525.
Yilmaz O, Patinote A, Nguyen TV, Com E, Lavigne R, Pineau C, Sullivan CV, Bobe J (2017) Scrambled eggs: Proteomic portraits and novel biomarkers of egg quality in zebrafish (*Danio rerio*). *Plos One,* **12**, e0188084.
Yin ZX, He JG, Deng WX, Chan SM (2003) Molecular cloning, expression of orange-spotted grouper goose-type lysozyme cDNA, and lytic activity of its recombinant protein. *Dis Aquat Organ,* **55**, 117-123.
Yoon IS, Lee GW, Kang SI, Park SY, Lee JS, Kim JS, Heu MS (2018) Chemical composition and functional properties of roe concentrates from skipjack tuna (*Katsuwonus pelamis*) by cook-dried process. *Food Sci Nutr,* **6**, 1276-1286.
Yousif AN, Albright LJ, Evelyn TPT (1991) Occurrence of Lysozyme in the Eggs of Coho Salmon *Oncorhynchus-Kisutch*. *Dis Aquat Org,* **10**, 45-49.
Yousif AN, Albright LJ, Evelyn TPT (1994) In-Vitro Evidence for the Antibacterial Role of Lysozyme in Salmonid Eggs. *Dis Aquat Org,* **19**, 15-19.
Yu Y, Santat LA, Choi S (2006) Bioinformatics packages for sequence analysis. *Appl Mycol Biotechnol,* **6**, 143-160.
Yuan XY, Liu WB, Liang C, Sun CX, Xue YF, Wan ZD, Jiang GZ (2017) Effects of partial replacement of fish meal by yeast hydrolysate on complement system and stress resistance in juvenile Jian carp (*Cyprinus carpio* var. Jian). *Fish Shellfish Immunol,* **67**, 312-321.
Zamaninia Z, Azizi A, Heidari B, Hadavi M (2017) Variations of Immune Parameters in the Early Life Stages and the Female Broodstock of the Cultured Persian Sturgeon, *Acipenser persicus* (Borodin 1897) and the Sterlet, *Acipenser ruthenus* (Linnaeus, 1758). *Turk J Fish Aquat Sc,* **17**, 1017-1024.
Zapata A, Diez B, Cejalvo T, Frias CG, Cortes A (2006) Ontogeny of the immune system of fish. *Fish Shellfish Immunol,* **20**, 126-136.
Zar JH (1999) *Biostatistical analysis,* Pearson Education India, 4th edn.
Zarski D, Le Cam A, Frohlich T, Kosters M, Klopp C, Nynca J, Ciesielski S, Sarosiek B, Dryl K, Montfort J, Krol J, Fontaine P, Ciereszko A, Bobe J (2021) Neurodevelopment vs. the immune system: Complementary contributions of maternally-inherited gene transcripts and proteins to successful embryonic development in fish. *Genomics,* **113**, 3811-3826.
Zarski D, Nguyen T, Le Cam A, Montfort J, Dutto G, Vidal MO, Fauvel C, Bobe J (2017) Transcriptomic Profiling of Egg Quality in Sea Bass (*Dicentrarchus labrax*) Sheds Light on Genes Involved in Ubiquitination and Translation. *Mar Biotechnol,* **19**, 102-115.
Zhang JA, Lanham KA, Peterson RE, Heideman W, Li LJ (2010) Characterization of the adult zebrafish cardiac proteome using online pH gradient strong cation exchange-RP 2D LC coupled with ESI MS/MS. *J Sep Sci,* **33**, 1462-1471.
Zhang QY, Fan SG, Luo C (2009) Sequence cloning and expression analysis of recombination activating gene 1

and 2 (rag1 and rag2) in grass carp, *Ctenopharyngodon idellus*. *Acta Hydrobiol Sin,* **33**, 795-803.
Zhao BS, Wang X, Beadell AV, Lu Z, Shi H, Kuuspalu A, Ho RK, He C (2017) m(6)A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature,* **542**, 475-478.
Zhao J, Song L, Li C, Zou H, Ni D, Wang W, Xu W (2007) Molecular cloning of an invertebrate goose-type lysozyme gene from *Chlamys farreri*, and lytic activity of the recombinant protein. *Mol Immunol,* **44**, 1198-1208.
Zheng WB, Tian C, Chen XH (2007) Molecular characterization of goose-type lysozyme homologue of large yellow croaker and its involvement in immune response induced by trivalent bacterial vaccine as an acute-phase protein. *Immunol Lett,* **113**, 107-116.
Zheng WW, Chu Q, Xu TJ (2021) The long noncoding RNA NARL regulates immune responses via microRNA-mediated NOD1 downregulation in teleost fish. *J Biol Chem,* **296**.
Zhou XY, Wang S, Zheng MM, Kuver A, Wan XY, Dai KZ, Li X (2019a) Phosphorylation of ELAVL1 (Ser219/Ser316) mediated by PKC is required for erythropoiesis. *Bba-Mol Cell Res,* **1866**, 214-224.
Zhou YY, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019b) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun,* **10**, 1-10.
Zilli L, Beirao J, Schiavone R, Herraez MP, Gnoni A, Vilella S (2014) Comparative Proteome Analysis of Cryopreserved Flagella and Head Plasma Membrane Proteins from Sea Bream Spermatozoa: Effect of Antifreeze Proteins. *Plos One,* **9**.
Ziv T, Gattegno T, Chapovetsky V, Wolf H, Bamea E, Lubzens E, Admon A (2008) Comparative proteomics of the developing fish (zebrafish and gilthead seabream) oocytes. *Comp Biochem Phys D,* **3**, 12-35.
Zorrilla I, Chabrillon M, Arijo S, Diaz-Rosales P, Martinez-Manzanares E, Balebona MC, Morinigo MA (2003) Bacteria recovered from diseased cultured gilthead sea bream (*Sparus aurata* L.) in southwestern Spain. *Aquaculture,* **218**, 11-20.