BABAK NAJAFPOUR

Microbiome and transcriptome ontogeny with a focus on the fish immune system: insights into developmental biology and host-bacteria interactions

UNIVERSIDADE DO ALGARVE
Faculdade de Ciências e Tecnologia

Faro, 2023

BABAK NAJAFPOUR

Microbiome and transcriptome ontogeny with a focus on the fish immune system: insights into developmental biology and host-bacteria interactions

Doutoramento em Ciências Biológicas
Especialidade em Bioquímica e Biotecnologia

Trabalho efetuado sob a orientação de:
Prof. Doutora Deborah M. Power
Doutora Patricia I Pinto

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia Faro, 2023

BABAK NAJAFPOUR

Microbiome and transcriptome ontogeny with a focus on the fish immune system: insights into developmental biology and host-bacteria interactions

Doctoral Programme in Biological Sciences
Speciality in Biochemistry and Biotechnology

Supervisors:
Prof. Deborah M. Power
Dr Patricia I Pinto

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia Faro, 2023

Título | Thesis title

Microbiome and transcriptome ontogeny with a focus on the fish immune system: insights into developmental biology and host-bacteria interactions

Declaração de autoria de trabalho:

Declaro ser a autora 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.

Babak Najafpour

Support

The European Union Horizon2020 Programme (PerformFISH, grant n° 727610) and the Portuguese Foundation for Science and Technology (FCT) project to CCMAR (UIDB/04326/2020) financed this study as the Ph.D. grant.

Apoio

O Programa Horizon2020 da União Europeia (PerformFISH, bolsa n° 727610) e o projeto da Fundação para a Ciência e Tecnologia (FCT) ao CCMAR (UIDB/04326/2020) financiaram este estudo como doutoramento conceder.


CCMAR
Centro de Ciências do Mar


PerformFISH

Dedicated to the cherished souls of my beloved family members who have passed away, especially my dear mother, who raised me with all her heart, and my father-in-law.

Acknowledgements

When I think about this chapter of my life book, I see how hard it would have been to go through the challenges without the help and support of my friends and the honourable people I encountered with them. People whose names and memories are imprinted in the book of my life forever. I must express my gratitude to my supervisors, Prof. Deborah Power and Adelino Canário, who support me in carrying out this research and my personal life, although I know that words cannot explain their help. Their activity, responsibility, ethics, and intelligence always encouraged me to go forward in this period.

I am grateful to my friends Dr. Patricia Pinto and Dr. João Cardoso, who contributed to this research and always helped me.

I am grateful to all my friends in the Comparative Endocrinology and Integrative Biology group, especially Ms. Soraia Santos.

I thank the financial sources of this research, especially PerformFISH (Horizon 2020).

I thank all the fish breeders for their constructive interaction and Dr. Katerina Moutou, manager of the PerformFish project.

I would thank all the researchers, technicians, and students at the CCMAR Center and the University of Algarve, who always generously helped me in this research.

Finally, I am grateful to my family, especially my wife, who has always been by my side.

I wish the best for all the people who participated in this research.

Abstract

Aquaculture hatcheries face persistent challenges concerning poor larvae quality and high mortality rates. The early developmental stages and the transition from larvae to juveniles are critical periods characterized by heightened vulnerability to environmental stressors. Researchers have turned to next-generation sequencing techniques, which have provided invaluable insights into organism biology and symbiotic bacteria. In the present research, RNA-seq and 16S rRNA gene sequencing have been employed to illuminate the microbiome and transcriptome during fish early development, focusing on two commercially farmed species, gilthead sea bream, and European sea bass. This investigation further delves into the host's vital biological responses to pathogens, such as the complement system, using a comprehensive phylogenetic analysis, and reveals that species-specific gene duplication and functional diversity of complement C3 may enhance the capacity of fish to activate complement through direct interaction of C3 isoforms with pathogenic agents. Designing bacterial genus-specific primers and complementary techniques like quantitative PCR enables targeted tracking of the colonization of larvae by abundant bacteria during their ontogeny. The analyses of the microbiome and transcriptome revealed dynamic profiles of bacteria and host gene expression during larval development, underscoring their potential interaction and stage-specific adaptability to rearing practices. This uncovers new management opportunities as manipulating microbiome profiles may be a potentially effective method to modulate larvae quality during the hatchery phase. Comparing larvae from different sites, stages, and species identified the different changes in the core gene expression or microbiome profiles during metamorphosis at the hatchery phase. This investigation also hints at the intriguing interaction between the microbiota and specific pathways, such as muscle development, immune response, and energy homeostasis. The study demonstrates how rearing water, live feed, season, and age are coupled with host selection of beneficial bacteria and are the main drivers of the microbiome in fish larvae. This comprehensive investigation contributes new knowledge to improve management strategies and fish health in the hatchery phase. The importance of this research extends beyond aquaculture, as transcriptome and microbiome changes associated with age provide crucial insights into the basic biology of the host and the changing holobiont throughout development.

Keywords: development, holobiont, immune system, microbiome, pathogens, transcriptome

Resumo

Os incubatórios de aquicultura enfrentam desafios persistentes relativos à má qualidade das larvas e às altas taxas de mortalidade. Os estágios iniciais de desenvolvimento e a transição de larvas para juvenis são períodos críticos caracterizados por maior vulnerabilidade a estressores ambientais. Os pesquisadores recorreram a técnicas de sequenciamento de última geração, que forneceram informações valiosas sobre a biologia dos organismos e das bactérias simbióticas. Na presente pesquisa, o sequenciamento do gene RNA-seq e 16S rRNA foi empregado para iluminar o microbioma e o transcriptoma durante o desenvolvimento inicial dos peixes, com foco em duas espécies cultivadas comercialmente, a dourada e o robalo europeu. A análise transcriptômica comparou estágios larvais críticos durante a metamorfose usando dourada e larvas de linguado senegalês coletadas de vários incubatórios comerciais, principalmente nos estágios de flexão e metamorfose média. O agrupamento dos perfis de expressão gênica obtidos foi determinado pela idade e peso em uma análise de PCA das 25 bibliotecas de dourada. Genes diferencialmente expressos foram identificados comparando larvas em diferentes estágios/grupos de desenvolvimento, onde a comparação I (flexão vs. metamorfose média) identificou 2.243 e a comparação II (metamorfose precoce vs. metamorfose média) resultou em 2.299 transcritos expressos diferencialmente. Este estudo investiga ainda mais as respostas biológicas vitais do hospedeiro a patógenos, como o sistema complemento, usando uma análise filogenética abrangente, e revela que a duplicação genética específica da espécie e a diversidade funcional do complemento C3 podem aumentar a capacidade dos peixes de ativar o complemento através interação direta de isoformas C3 com agentes patogênicos. A dourada (Sparus aurata) teve nove transcritos do gene c3 altamente expressos no fígado, embora, como relatado em outros peixes, também ocorra expressão extra-hepática. O sequenciamento do gene do RNA ribossômico 16S foi usado para traçar o perfil das comunidades bacterianas dos ovos e da água dos reprodutores de três incubatórios comerciais. As proteobactérias foram os filos mais comuns e dominantes nas amostras (49,7% em média). *Vibrio* sp. foi o gênero mais representado (7,1%), seguido por *Glaciecola* (4,8%), *Pseudoalteromonas* (4,4%) e *Colwellia* (4,2%), em ovos e água nos locais. Os desinfetantes à base de iodo usados rotineiramente reduziram ligeiramente a carga bacteriana dos ovos, mas não alteraram significativamente a sua composição. O projeto de primers específicos do gênero bacteriano e técnicas complementares como PCR quantitativa permite o rastreamento direcionado da colonização de larvas por bactérias abundantes durante sua ontogenia. Iniciadores eficientes específicos de gênero foram projetados para onze gêneros bacterianos, incluindo *Alkalimarinus*, *Colwellia*, *Enterovibrio*,

Marinomonas, *Massilia*, *Oleispira*, *Phaeobacter*, *Photobacterium*, *Polaribacter*, *Pseudomonas* e *Psychrobium*. Os microbiomas da água, da alimentação viva e das larvas de peixes diferiam, mas partilhavam as mesmas bactérias centrais, que incluíam *Vibrio* e *Pseudoalteromonas*. A criação de água e alimentos vivos teve uma contribuição diferente para a microbiota larval: a transferência de *Vibrio* foi mais provável com Artemia e rotíferos do que com algas e água. A estação (janeiro e maio) teve um efeito menor na composição microbiana nos incubatórios. Larvas mais jovens (até 23 dias pós-eclosão, dph) e mais velhas (42-77 dph) tinham diferentes comunidades microbianas, e a carga bacteriana aumentou com a idade. A seleção da microbiota do hospedeiro foi evidente no início da alimentação em comparação com o meio da metamorfose, e as larvas iniciais foram colonizadas por algumas bactérias, e. *Bifidobacterium*, *Lactobacillus*, *Bacteroides* e *Blautia*, que podem ser benéficos para o desenvolvimento. A análise funcional previu interações significativas entre o hospedeiro e a microbiota durante o desenvolvimento larval ligadas ao crescimento e resistência das bactérias ao hospedeiro ou à resposta a mudanças na fisiologia e morfologia do hospedeiro nos estágios intermediários da metamorfose (larvas mais velhas). As análises do microbioma e do transcriptoma revelaram perfis dinâmicos de bactérias e expressão gênica do hospedeiro durante o desenvolvimento larval, ressaltando sua interação potencial e adaptabilidade específica do estágio às práticas de criação. Isto revela novas oportunidades de gestão, uma vez que a manipulação de perfis de microbioma pode ser um método potencialmente eficaz para modular a qualidade das larvas durante a fase de incubação. A comparação de larvas de diferentes locais, estágios e espécies identificou as diferentes mudanças na expressão genética central ou nos perfis do microbioma durante a metamorfose na fase de incubação. Esta investigação também sugere a intrigante interação entre a microbiota e vias específicas, como o desenvolvimento muscular, a resposta imune e a homeostase energética. O estudo demonstra como a água de criação, a alimentação viva, a estação e a idade estão associadas à seleção do hospedeiro de bactérias benéficas e são os principais impulsionadores do microbioma nas larvas de peixes. Esta investigação abrangente contribui com novos conhecimentos para melhorar as estratégias de manejo e a saúde dos peixes na fase de incubação. A importância desta pesquisa vai além da aquicultura, uma vez que as alterações no transcriptoma e no microbioma associadas à idade fornecem informações cruciais sobre a biologia básica do hospedeiro e a mudança do holobiont ao longo do desenvolvimento.

Palavras-chave: desenvolvimento, holobiont, microbioma, patógenos, sistema imunológico, transcriptoma

Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter2

Chapter 3

Chapter 4

CHAPTER 1

Introduction: microbiome establishment and transcriptomic modifications during fish early development

Exploring host-bacteria interactions in early life of fish: insights from transcriptome and microbiome in commercial and model teleost – a review

Manuscript is in preparation as a review paper.

Exploring host-bacteria interactions in early life of fish: insights from transcriptome and microbiome in commercial and model teleost

Babak Najafpour[1], Adelino V. M. Canario[1,2] and Deborah M. Power[1,2#]

[1]CCMAR - Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, Edifício 7, 8005-139 Faro, Portugal

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

CReDit statement:

Babak Najafpour: Writing - Original Draft, Review & Editing; Adelino VM Canario: Writing - Review & Editing; Deborah M Power: Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

1.1. Abstract

Poor larvae quality and high mortality remain problematic bottlenecks in aquaculture hatcheries. Early development and the transition from larvae to juvenile are critical stages in the fish lifecycle due to their high vulnerability to environmental stressors. Next-generation sequencing approaches have altered and broadened knowledge about organism biology and their symbiotic bacteria. The importance of the microbiome and transcriptome in aquaculture, their dynamic profiles with age, and microbiota driven events during larval development (fish-microbe interactions, feeding, bacteria of water, and farm location) are accentuated points in this review. The best alternatives to modulate the microbiota during the hatchery phase are highlighted, such as removing the sources of pathogens in live feeds or manipulating microbiome profiles using selected mixtures of bacteria/prebiotics in rearing water/feeds. Colonization of the gut with unstable and dynamic microbiota at early life stages suggests that microbiota modulation should start early to establish a stable microbiota population. Distinct transcriptomes and microbiomes of larvae at different developmental stages suggest there is scope for further adjustments of rearing practices to optimize larval development. The significantly modified genes during larval ontogeny may represent candidate gene markers that drive phenotypic plasticity and by promoting or repressing their regulation could be used to enhance larval quality. The parallel changes in the transcriptome and microbiome during larval development hint that there is at an interaction between microbiota and specific pathways, including those of muscle development, thyroid and GH-IGF1 axes, energy homeostasis, gastrointestinal maturation, immune response, and tissue homeostasis. New knowledge from transcriptome and microbiome studies can contribute new knowledge, concepts and strategies to increase hatchery performance and harmonize future research to decipher host-bacteria interactions and their influence and importance for larval ontogeny and microbiome establishment.

Key words: aquaculture, bacteria establishment, early ontogeny, hatchery, host transcriptomes

1.1. Introduction

The aquaculture industry is growing fast and is expected to make an increasingly important impact on the worldwide food sector. The potential benefits of shifting human diets from meat to other protein sources, including seafood, are well described (Froehlich et al., 2018). Aquaculture has already had a positive impact on food production (aquatic foods providing about 17 % of animal protein) and boosted economic growth in coastal and rural areas (supporting the livelihoods of 10 - 12% of the global population, employing > 58 million people, FAO, 2022). According to the World Bank, billions of people worldwide, especially the world's poorest, rely on healthy oceans as a source of income and food. This is under threat from the environmental crises that are modifying the climate and changing the oceans (the largest global carbon sink). This makes climate actions and innovative solutions a priority to proof oceanic food production and safeguard the livelihoods of coastal communities (World Bank, 2023). The importance of aquaculture and its potential to deliver food security is clear since fish is a valuable source of protein for some 950 million people worldwide (Pradeepkiran, 2019). Overall, global fish production has increased in all the continents, and Asia has had the most significant gain, particularly China (FAO, 2020). Finfish production has the largest share of aquatic animal farming with an inland harvest of 47 million tonnes and marine and coastal aquaculture contributing 7.3 million tonnes (FAO, 2020). The increased production from aquaculture contrasts with the decline in fisheries resources due to their overexploitation and factors such as water pollution and technological developments (Perissi et al., 2017). The potential of aquaculture to contribute to food security depends in part on improved management through optimized production cycles in hatcheries that can guarantee high-quality larvae for the grow-out stage while also potentially contributing to fishery resources. A good understanding of the impact of biotic and abiotic factors in intensive larviculture is essential since changes in morphology, organ function and transitional stages (e.g., endogenous-exogenous feeding or larvae-juvenile) may demand specific cultivation measures that are critical for larval survival and performance.

The European sea bass (*Dicentrarchus labrax*) and the gilthead sea bream (*Sparus aurata*) are examples of commercially farmed fish species in the Mediterranean region, in which the intensive production model includes grow-on of juveniles in sea cages (FAO, 2018). Aquaculture of sea bream and sea bass is the second most valuable aquaculture industry after salmonids in the European Union (Llorente et al., 2020). In general, the economic performance of the sector has been poor in the last ten years, and considerable effort has been directed at

improving efficiency and overcoming profitability issues (Llorente et al., 2020). Manipulation of biotic and abiotic factors to improve efficiency and productivity have been the target of many scientific studies with the aim of identifying the biological response, and interaction of fish with their environment (e.g., microbiota). However, a limiting factor of many experimental studies of aquaculture species is that they do not cover the scale or diversity of the production systems in different hatcheries and countries (Llorente et al., 2020). The aim of this review is to give insight into the status of microbiome and transcriptome studies in larviculture, with particular attention given to gilthead sea bream and European sea bass and in this way identify challenges, gaps in knowledge and the research still needed to define optimal rearing practices.

Embryonic and larval development and the transition from larvae to juveniles (metamorphosis) are critical stages for determining production traits due to their vulnerability to environmental stressors (Martínez et al., 2021; Najafpour et al., 2019; Sifa and Mathias, 1987). Specific pathogens of fish eggs and larvae (Hansen and Olafsen, 1999; Merrifield and Rodiles, 2015; Muniesa et al., 2020; Olafsen, 2001) and low adaptation due to inadequate changes in morpho-physiology (e.g., digestive tract and enzymes activity) and behavior of larvae during developmental transitions (Najafpour et al., 2021a; Searcy and Sponaugle, 2001) are proposed as potential causes of larval vulnerability at metamorphosis. The yolk is the main energy source during the autotrophic stage (from fertilization to the onset of exogenous feeding) for most fishes (**Figure 1.1**) and a clear difference between the early development of most fish and other oviparous animals (e.g., invertebrates, reptiles, and birds) is that the yolk is not fully resorbed before hatching (Kamler, 1992). Hatching of poorly developed larvae and the major morphological and functional changes that need to occur up to the juvenile stage (**Figure 1.1**) explains the vulnerability of fish larvae to environmental factors. The nutritional requirements, physiology, differing stages of development and biological plasticity contribute to the differing vulnerability of different developmental stages of fish larvae (**Figure 1.1**), and may also explain the nonuniform transcriptional responses detected in experiments and between studies. The susceptibility of early life stages and the limited availability of studies on larvae compared to adults, and the fundamental importance of this stage from a commercial and biodiversity perspective justifies the review about fish larvae from a microbiome and transcriptome perspective.

Recent advances in the understanding of the microbiome and its influence on fish biology during early development arise from the availability of cost effective and rapid approaches that yield in-depth analysis of the holobionte. The biological processes or pathways

associated with larval development and the interaction of larvae with microbiota are potential core factors that determine larvae survival and performance. For example, a fully developed digestive tract is essential after the onset of the exogenous feeding (Mazurais et al., 2011) and interactions between the microbiome and fish mucosal tissues are essential for health (Merrifield and Rodiles, 2015). In the present review, the outcome of transcriptome and microbiome studies directed at larval ontogeny are considered to highlight microbiota establishment, the predominant larval transcriptomic responses and the potential interactions between the host and microbiota at different developmental stages. Knowledge gaps will be identified and the need for future studies highlighted.

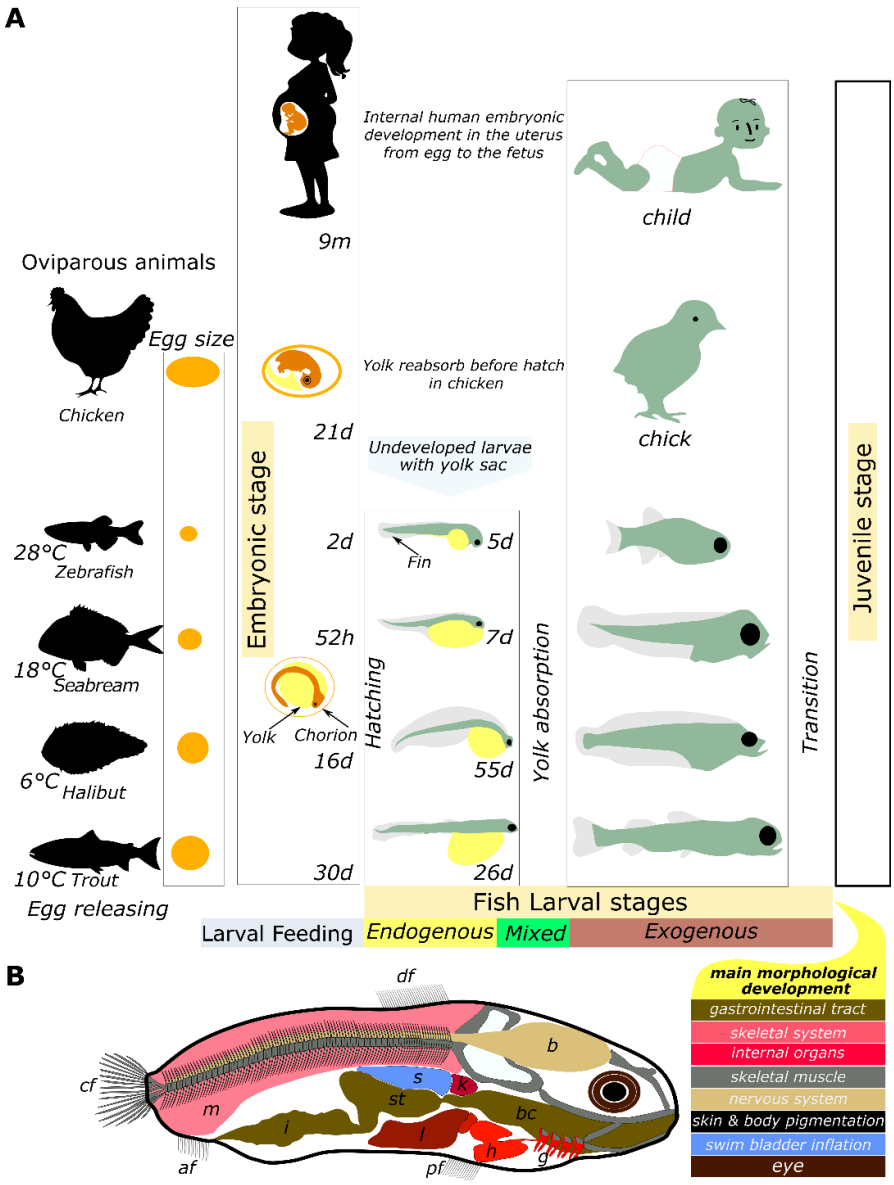


Figure 1.1. Early development of fish larvae. A) a comparison between fish and other

organisms shows a poorly developed fish larvae hatches from the egg, and also the duration and key stages of embryonic and larval development in several different fish species. B) The core morphological changes that occur during fish larval development are highlighted, b = brain and nerve cord, bc = buccopharyngeal cavity, fins (cf = caudal fin; af = anal, df = dorsal, pc = pectoral), g = gills, h = heart (and cardiovascular system), i = intestine, k = kidney, m = muscle, st = stomach, s = swim bladder.

1.1.1. PhD objectives

The current Ph.D. thesis aims to profile the microbiome and transcriptome associated with larval performance and development under hatchery conditions on a large scale. The ultimate objective is to enhance larval resistance against pathogens. To achieve this goal, the research is structured around the following specific objectives:

1- **Transcriptomics**: this objective seeks to provide a comprehensive understanding of the biological capabilities of the larvae at the molecular level.

2- **Innate immune system evaluation**: this objective focuses on the assessment of the fish's innate immune system, with a particular emphasis on the complement system in European sea bass and gilthead sea bream.

3- **Egg microbiome characterization**: this aims to determine the composition of the egg microbiome and explore the impact of various factors, including disinfection methods.

4- **Microbiome profiling of larvae**: this objective involves characterizing the microbiome of first-feeding and mid-metamorphic larvae sourced from hatcheries across Mediterranean coastal water.

Since the research findings have been or are in the process of being published, the document is structured in paper-based as follows

Chapter 1: this chapter presents a comprehensive review of microbiome and transcriptome studies during larval development. It provides insight into potential interactions between host and bacteria based on previous research and the outcomes of the current thesis.

Chapter 2: this chapter includes the modifications in the transcriptome during the larvae's metamorphosis.

Chapter 3: here, an in-depth analysis of the complement system, a vital component of innate immunity, is presented.

Chapter 4: this chapter profiles the microbiota during early development and emphasizes the key factors influencing the establishment of microbiota in both eggs and larvae.

1.1.2. New challenges and opportunities from the holobiont in finfish aquaculture

There are numerous studies and several excellent reviews since the 1980's that have characterized the substantial morpho-physiological changes that occur during larval development (Dabrowski, 1986; Falk-Petersen, 2005; Izquierdo et al., 2000; Leclercq et al., 2010; Pittman et al., 2013; Portella et al., 2014; Rønnestad et al., 2013, 2007; Varsamos et al., 2005; Zambonino Infante and Cahu, 2001). There are far fewer studies that have analysed development from a comparative, molecular perspective and exploited it to establish a robust understanding and identification of molecular markers that give insight into larval condition during hatchery production. A range of factors may explain the lack of robust and common molecular markers for characterization of larvae and include: **1**- different developmental rates across fish species, **2**- the practice of mixing together in a single production batch of eggs and larvae (differing by hours to days), which is associated with variable gene expression patterns in transcriptome studies, **3** – the multitude of different fish species cultured for aquaculture (> 300 globally, FAO, 2022) and the limited conservation of gene sequences between species, which limits the introduction of common gene markers for all aquaculture species, **4** – the fact that the up- and down-regulation of gene markers may be stage-and species specific so patterns of gene expression can change across development and so require that the ontogenic expression pattern of genes be established in different stages and species, **5** - most molecular studies of larval development have focussed on candidate genes rather than whole transcriptome analysis, which limits the scope and likelyhood of detecting quality predictive gene markers, **6** - aging, diet, and environmental factors can affect the epigenome and in this way the phenotype of juveniles and adults (e.g., change in DNA methylation patterns, Nafee et al., 2008). The increasing availability of transcriptome studies in fish larvae, makes the review timely since the aim is to highlight common biological processes with significant changes at a transcriptional level during larval development in different fish species. By identifying species-specific and species-independent genes, their expression patterns and their correlation with larval quality the general aim is to identify interesting candidate genes that can be targeted in future work (e.g., genetic selection, quality markers etc).

1.1.3. The link between bacteria and host and the potential impact of microbiota symbiosis during host development

Beyond fish biology, the complex and emerging relationship between fish, their microbiota and the environmental microbiota during larval ontogeny is still unclear. Dysbiosis and pathogen outbreaks can contribute to high mortality during larvae production. The identification of pathogenic bacteria in core microbiota of eggs and lavae and the overall low incidence of bacteria with beneficial characteristics in hatcheries emphasizes the ever present risk of disease outbreaks and the need for more knowledge about the main factors that drive the microbiome (Najafpour et al., 2023). Therefore, other specific objectives of this review are to highlight, a) the importance of the microbiota in aquaculture, b) factors that determine and influence establishment of the microbiome, and c) the potential biological pathways influenced by microbiota during larval ontogeny. Among the studies that exist, a correlation has been identified between altered gut microbiota and host-specific gene expression (Yoon et al., 2022). Significant positive and negative correlations have been found in Gilthead sea bream (*S. aurata*) juveniles between the gut microbial community abundance and the expression level of some candidate genes related with the host immune response (e.g., *igm*, *il8*, *tnfα*, *il10*, *lgals1*, *lgals8*, *atlr2*, *tlr5*, *tlr9*, *muc13)* and *scd1*- and *elovl6*-mediated lipogenesis in the liver (Naya-Català et al., 2021). A recent review highlighted the importance of antimicrobial defence mechanisms in fish including some innate (e.g., TLR) and adaptive immune genes (e.g., IgM) and the contribution of the microbiota to larval nutrition and physiology (Vadstein et al., 2013).

Nonetheless, an appropriate tool to integrate microbiology and immunity is still missing. Furthermore, the knowledge that has recently been produced about molecular processes underpinning fish larvae biological responses and the microbiome using next generation sequencing (NGS) methods, need to be integrated. The current review aims to identify and decipher significant changes in the most common molecular pathways, identify common pathways during development in different species and decipher potential host-bacteria interactions under hatchery conditions by comparison and integration of reported transcriptome and microbiome studies. For example, transcriptome analysis of grouper during metamorphosis identified several significantly modified pathways and molecules (e.g., leukocyte transendothelial migration and phagosome, cdh5) linked to immunity (Xiao Joe et al., 2019). The application of RNA-seq (called dual RNA-seq) and third-generation sequencing (e.g., nanopore sequencing) technologies enables simultaneous analysis of gene expression in the host and pathogen. The application of these approaches to studies of immunity is making a

significant impact on the state of the art and shifting research from hypothesis-based to unbiased discovery-based approaches (Westermann et al., 2012), with the potential to generate new insight into the holobiont and larval development.

1.1.4. *Effective management of diseases requires a better understanding of the host biology and its relationship with the environment and symbiotic bacteria*

In general, a range of management strategies and technologies have been developed and applied in aquaculture production to overcome bottlenecks including low larvae quality and high mortality, species diversification, live and formulated feeds, selective breeding, parentage genetics, marker-assisted selection, and disease and water management (Yue and Shen, 2022). Low genetic variability, low quality gametes, inadequate nutrition, and pathogen outbreaks all negatively impact hatchery performance (Sbordoni et al., 1986; Vadstein et al., 1993, 2013). A range of management strategies have been used to prevent disease outbreaks in aquaculture, including vaccination and therapeutic approaches such as pre- and probiotics (Gudding and Van Muiswinkel, 2013; Pérez-Sánchez et al., 2018). Nonetheless, the cost of disease outbreaks and their impact on the aquaculture industry, larviculture, and health management programs are problems that challenge the success and sustainability of intensive rearing in commercial hatcheries (Bondad-Reantaso et al., 2005; Vadstein et al., 2013). The host species, fish age, and the disease status (acute, chronic, subclinical carrier) are determinant factors for pathogen detection and disease (Toranzo et al., 2005). Examples of common diseases during early development include infectious pancreatic necrosis (IPN) in young salmonid fry that causes up to 100% mortality (Crane and Hyatt, 2011), *Flexibacter ovolyticus* in eggs and yolk-sac larvae of the Atlantic halibut (Bergh et al., 1992), and vibriosis, caused by *Vibrio* species such as *Vibrio anguillarum*, which causes severe economic losses in marine aquaculture (Novriadi, 2016; Silva et al., 2014; Touraki et al., 2012). Better characterization of microbiome establishment, composition, succession, and function and the holobionte may offer new solutions for management. Improved understanding of immune capacity in larval stages, is by analogy with terrestrial farming, essential for the design of effective management approaches for pathogens, such as vaccination (Liu et al., 2022; Mulero et al., 2008).

Studies suggest that fish larvae gain immunocompetence sometime after the formation of the lymphomyeloid organs (head–kidney, thymus, and spleen, Lam et al., 2004; Mulero et al., 2008). The assessment of candidate immune gene expression during development led to the proposal that immunocompetence in larval fish such as the gilthead sea bream is

primarily dependent on innate immunity (Mulero et al., 2008). The identification of the age at which a protective immune response, based on memory develops, is essential knowledge for the formulation of successful vaccination protocols (Liu et al., 2022). In parallel, the microbiome, especially of the gut, may play a significant role in modulating an effective response to vaccines administered by immersion, and dysbiosis due to a range of factors (e.g., antibiotic intake, stress, diet, environmental factors, and concomitant diseases) can reduce the effectiveness of vaccines with age (Bosco and Noti, 2021; Ferreira et al., 2010; Valdez et al., 2014). A range of challenges still exist for the elaboration of effective fish vaccines since most are currently delivered by injection, which makes mass immunization costly and causes problems of safety, compliance, and morbidity (Adams, 2019; Miquel-Clopés et al., 2019).

Oral vaccine delivery is widely used for fish immunization at a young age since it permits application on a population-scale, and reduces stress, although the efficacy of the method compared to intraperitoneal injection is still controversial (Adams, 2019; Dang et al., 2021; Miquel-Clopés et al., 2019). To ensure effective mucosal vaccine delivery appropriate adjuvants are essential and improved basic knowledge about fish larvae biology, immunology and their microbiome are needed (Adams, 2019). The composition and function of the gut microbiota are proposed to be crucial factors modulating the immune response to vaccination since they can act as natural adjuvants and microbial metabolites and microbiota-encoded epitopes may be cross-reactive with vaccine antigens (Lynn et al., 2021). The interest in modulating the gut microbiota using diets that include prebiotics and probiotics as promoters of overall health and as potential enhancers of vaccine efficacy has increased dramatically (Wargo, 2020). The exploitation of microbiota for the modulation of fish health requires knowledge about the intrinsic biological capacity of larvae (e.g., immune response) and their interaction with microbiota. The advent of high-throughput sequencing approaches has enabled detailed studies of larval transcriptomes and their microbiota during their ontogeny. Integrating the outcome of such research in aquaculture may improve understanding of larval development and plasticity and the influence of the microbiota and may originate new approaches for monitoring and managing their performance.

1.1.5. NGS approaches and studying the microbiome and transcriptome in finfish aquaculture

Next-generation sequencing (NGS) approaches have altered and broadened knowledge about organism biology and their symbiotic bacteria (Lokman and Symonds, 2014), including the response of the host (transcriptomics), bacterial profiles and their gene expression

(metagenomics, meta-transcriptomics, **Figure 1.2**).

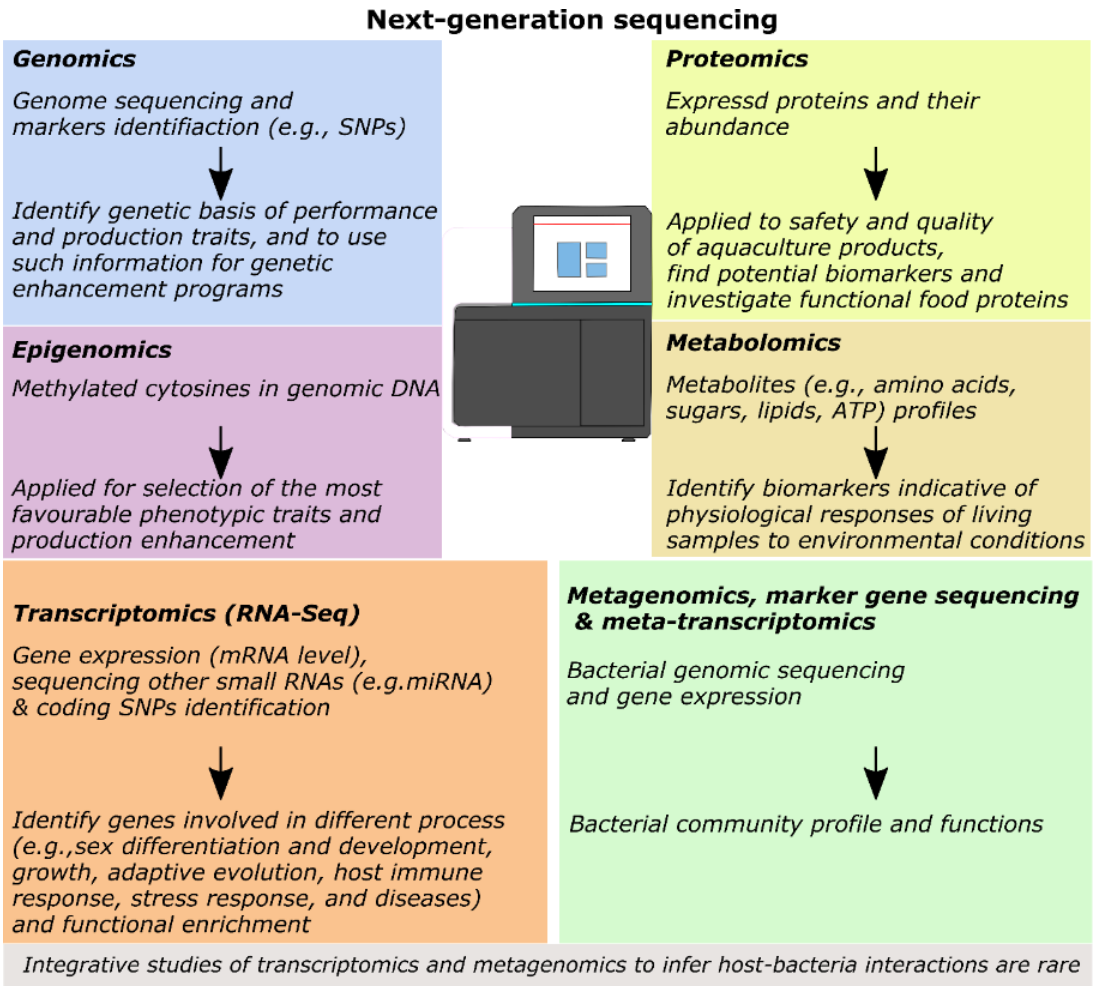


Figure 1.2. A summary of the main *omics* approaches that have been applied to aquaculture species, including next generation sequencing approaches for DNA, RNA and their modifications, proteins, and metabolites (Abdelrahman et al., 2017; Rodrigues et al., 2016; Roy et al., 2021).

Metagenomics includes molecular approaches that sample the genomes of a community of organisms inhabiting a common environment (Hugenholtz and Tyson, 2008; Sleator et al., 2008). Metagenomics approaches are culture-independent and provide a relatively unbiased view of the structure (species richness and distribution) and the function (metabolic) of microbial communities (Hugenholtz and Tyson, 2008) in different niche, including soil (Myrold et al., 2014), marine and freshwater ecosystems (Nowinski et al., 2019; Oh et al., 2011; Simon et al., 2014), plants (Fadiji and Babalola, 2020), tissues or organs in humans (e.g., gut

and skin) and other animals like fish (Carda-Diéguez et al., 2017; Ferretti et al., 2017; Johny et al., 2021; Qin et al., 2010, **Figure 1.3**).

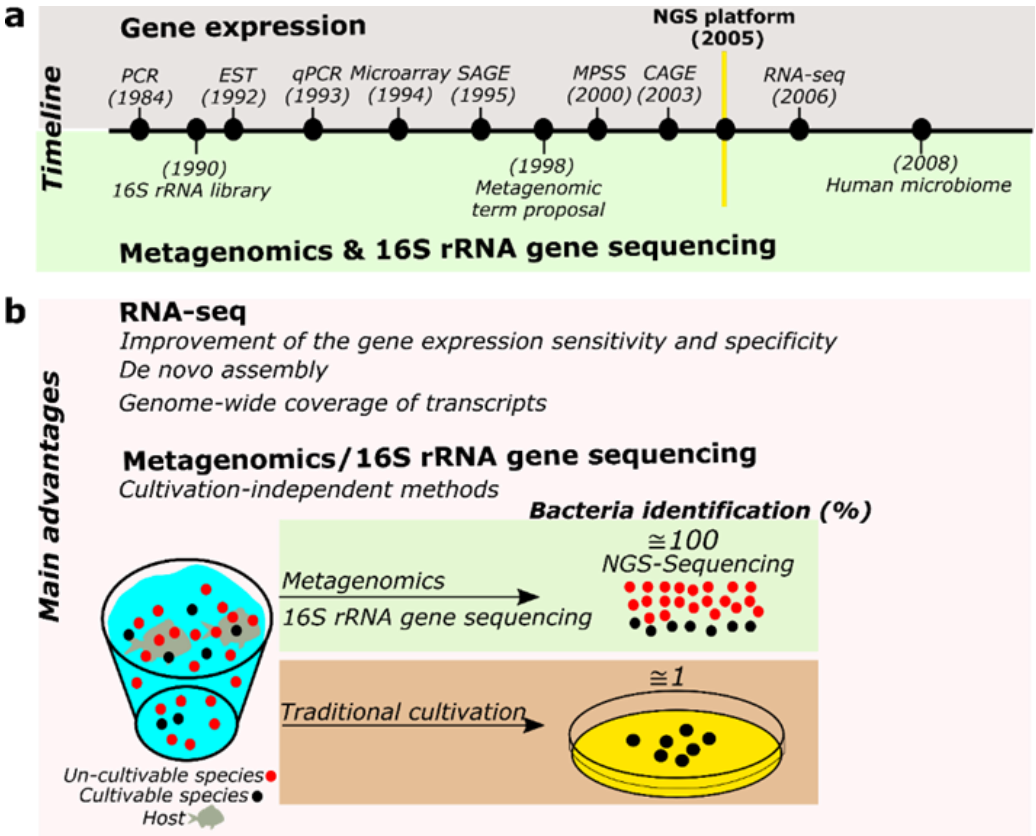


Figure 1.3. Next-generation sequencing timeline (a) and the main advantages of two high-throughput sequencing methods (RNA-seq and 16S rRNA gene sequencing) compared to traditional approaches (b). Transcriptional responses were analyzed using PCR and Sanger sequencing-based methods until 2005, and in 2006, RNA-seq, high-throughput sequencing methods started to be used and improved the sensitivity and specificity of the gene expression analysis and provided insight into the transcriptome of organisms even without the aid of a reference genome through *De novo* transcriptome assembly. The advent of 16S rRNA gene sequencing, another high-throughput sequencing method, has contributed to profile almost 100 % of the taxonomic composition of bacteria from environmental samples since it is culture-free and so overcome the significant limitations of conventional approaches.

In addition to metagenomics approaches such as shotgun metagenomics that sequence whole genome fragments, other cost-effective methods such as high-throughput sequencing of marker genes such as the 16S rRNA gene or internal transcribed spacer/ITS have

been developed to profile the taxonomic composition and predict the functional potential of microbial or fungal communities from environmental samples (Hilton et al., 2016; Kumar Awasthi et al., 2020; Quince et al., 2017).

The advent of 16S rRNA gene sequencing has expanded understanding of bacterial composition, its complexity, and how it affects the host, including fish (Ghanbari et al., 2015). 16S rRNA gene sequencing allows rapid identification of slow-growing bacteria and speeds up clinical diagnosis (Lau et al., 2015). The use of 16S rRNA genes for bacterial identification is linked to the relatively high conservation of the sequence of this gene between species except for nine hypervariable regions (v1-v9) that can be exploited for taxonomic identification (Chakravorty et al., 2007). The 16S rRNA gene sequencing pipeline starts with PCR amplification of the variable regions of the 16S rRNA gene (e.g., V3 and V4) using universal primers anchored in conserved regions. PCR amplicon libraries are generated and sequenced on an NGS platform (e.g., Illumina). The subsequent bioinformatics steps include barcode identification and sample clustering, read assembly, filtering to remove short or ambiguous reads, alignment of sequences against sequence data bases for identification (e.g., Silva), clustering of reads into operational taxonomic units (OUT), and finally phylogeny and taxonomic classification (Beiko et al., 2018; Marizzoni et al., 2020). Several tools and pipelines now exist for analysis of metagenomic data, including Galaxy, QIIME, and Mothur (Aguiar-Pulido et al., 2016). The complex interaction between fish and bacteria and the colonization of fish by bacteria during rearing are still unclear. Understanding how fish respond to bacteria at each developmental stage is essential, and since pathogens are generally found in the core bacteria of aquaculture fish this suggests boosting their response (e.g., vaccination or feed enrichment) to pathogens, coupled to control of external factors driving microbiota establishment and succession can provide alternative control strategies (Najafpour et al., 2023).

1.2. Transcriptomic modifications during early ontogeny of fish from microarray to RNA-seq

Ontogenic analysis by transcriptomics is a powerful tool to establish gene expression profiles correlated with optimal larval performance in aquaculture. Creating a link between the change in morphology, transcriptome, and microbiome may provide insight into the consequences of host-microbiota crosstalk and the biological basis of interactions with pathogenic and beneficial microbiota. Comparison of existing studies of larval transcriptomes

reveals a core of common molecular processes that occur during larval ontogeny and highlight specific gene sets crucial for normal fish development.

Knowledge about fish transcriptomes and gene expression has increased incrementally due to the availability of cost-effective NGS approaches such as RNA sequencing (RNA-seq, **Figure 1.3**). RNA-seq resolves the limitations of other transcriptomic approaches, including Sanger sequencing-based methods (e.g., SAGE, CAGE, and MPSS) that specify gene expression based on partial transcripts and do not discriminate isoforms (Anamika et al., 2016). Furthermore, RNA-seq has received more attention since it provides genome-wide coverage of transcripts, has enabled the assembly of transcriptomes even without a reference genome, and has improved gene expression sensitivity and specificity compared to other high-throughput methods like microarray (Anamika et al., 2016). RNA-seq approaches have provided valuable knowledge and extended understanding of biological processes involved in the development, growth, adaptive evolution, host immune response, stress response, and diseases and pathogen invasion in fish (Qian et al., 2014; Sudhagar et al., 2018). Meta-analysis of the PubMed database using the R packages rentrez v 1.2.3 (Winter, 2017) and easyPubMed v 2.13 (Fantini, 2019) retrieved over 200 scientific publication records related to fish RNA-seq (**Figure 1.4**, **Supplementary table 1.1**) that reveal the increasing trend of RNA-seq applications in fish, including studies of muscle growth, nutrition, evolution, stress, the immune response, pathogen challenge, and development (**Figure 1.4**). The meta-analysis of the PubMed database confirms that transcriptomic studies directed at developmental ontogeny are less common compared to other topics such as immunity and stress.

Therefore, in hatchery stages of fish during development it is essential to understand their biological capacity (e.g., immune response) and development-related requirements (e.g., nutrition, physiology etc.) so that approaches for microbiota manipulation and disease combat can be established. The availability of tools to assess whole larvae, tissues and organs and even single cell transcriptomics has the potential to bring new insights to molecular and cellular based changes during development of aquaculture species. If transcriptomics can be allied to 16S rRNA gene sequencing and metagenomics this should make it possible to decipher the crosstalk that occurs between the host and its microbiome. Since the microbiome and host transcriptome are dynamic and change during ontogeny studying their interaction is complex and has not generally been considered and for this reason is part of the present review.

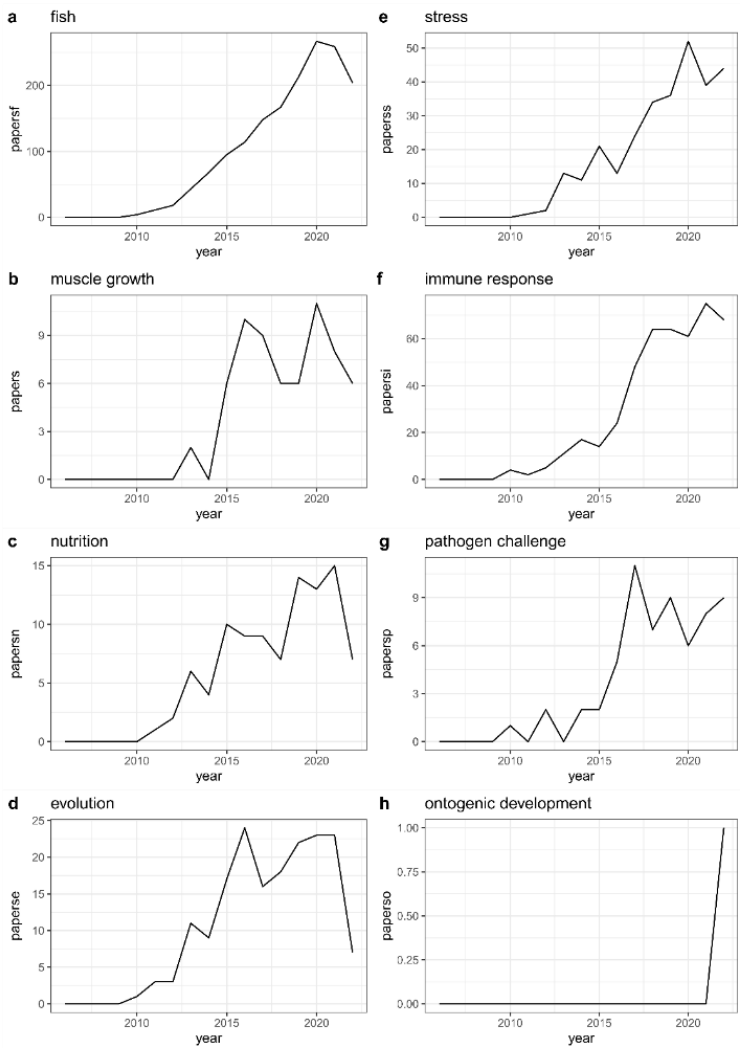


Figure 1.4. **The number of retrieved scientific publication records related to fish RNA-seq searching the PubMed database using different keywords**. The analysis was performed by the R packages rentrez v 1.2.3 (Winter, 2017) and easyPubMed v 2.13. The number of RNA-seq studies has increased significantly from 2006 to 2022. The search included the keywords presented as the titles of the graphs a-h (e.g., 'fish') plus 'RNA-seq'. The titles of all the publications collected using each keyword are provided in **Supplementary table 1.1**. The majority of RNA-seq studies in fish have focused on the important production traits, immunity, pathogens and stress in juveniles and adults.

Comparison of different developmental stages of the gilthead sea bream (neurula, hatching, pectoral budding/eye development, mouth opening, and eye pigmentation) using microarray revealed stage-specific gene clusters (Sarropoulou et al., 2005) and the genes involved in larval development from post-hatch to post-metamorphosis in Atlantic halibut

(Douglas et al., 2008). The predominant genes that changed during larval development were associated with organ development (e.g., muscular, visual, neural and ossification), the maturation of essential physiological functions (e.g., digestive), and the regulation of metabolic pathways (Darias et al., 2008; Mazurais et al., 2011). Application of *454 Pyrosequencing*, to analyze developing larval gilthead seabream identified the number of gene transcripts involved in larval development of Gilthead Sea bream, differentiation, morphology, and growth (Yúfera et al., 2012). Microarray analysis of *Solea solea* from hatching to metamorphosis revealed up-regulation of gene transcripts involved in innate (e.g., *lyz, ctsl, nlrp 1-5, c3, c5-c9, il 1b*) and adaptive immunity (e.g., *cd3z, tcra, tcrb, mhc1a, mhc1b, mhc2a, mhc2b,* Ferraresso et al., 2016).

However, RNA-seq analysis, which is more sensitive compared to previous sequencing techniques, identified differentially expressed gene isoforms between developmental stages that are involved in several pathways during larval ontogeny. Gene sets include genes of the thyroid axis (*dio1, dio2,dio3, and tg* from RNA-Seq studies in Atlantic halibut and gilthead sea bream (Alves et al., 2016; Najafpour et al. unpulished), GH-IGF axis (*gh1, igfbp1a, igfbp1b, igfr1a, igf2b, igfbp4, igfbp7,* and *ghrb* from RNA-Seq studies in zebrafish and gilthead sea bream, (Xu et al., 2017; Najafpour et al. unpulished), carbohydrate metabolism (*lct, pfkm, pygm, lct, pck1, sis, si, mgam* from RNA-Seq studies in three spot cichlid, *Amphilophus trimaculatus,* and gilthead sea bream Hilerio-Ruiz et al., 2021; Najafpour et al. unpulished), lipid metabolism (*cel.1, cel.2, pla2g12a, pla2g1b, pla2g3, fabp2, fabp6, fabp1b.1, fabp10a, agpat2, mogat2, dgat1a, apoa4a, apoa4b.2, apoa4b.3, cpt1aa, cpt1ab, cpt2, apoa1, plbd2, pla2, cel* from RNA-Seq studies in zebrafish and three spot cichlid, Xu et al., 2017; Hilerio-Ruiz et al., 2021), protein digestion (*tryp, ctr, pga* from RNA-Seq studies in three spot cichlid and gilthead sea bream, Hilerio-Ruiz et al., 2021; Najafpour et al. unpulished), cell proliferation and differentiation (*ccnd2, seh1l, kdm6a, arf4,* and *ankrd28* from RNA-Seq study in the Clearhead Icefish, *Protosalanx chinensis*, Tang et al., 2022), focal adhesion and ECM receptor interactions (*col1a1, col1a2, col5a2, ddr1, mmp13, mpk1, thbs4b, col9a2, col9a3, arhgap35b, itga8, thbs1a, frem2a, frem2b, thbs1a, thbs1b, itga8* from RNA-Seq studies in Atlantic halibut and gilthead sea bream, Najafpour et al. unpulished), and innate and adaptive immune responses (e.g., *c1qa, c1qb, c1qc, mr1, ifi44l, irf1, ifi35, ifngr1, il10ra, psme1, psme2, hsc70, hsp90aa1.2 , tap1 , b2m, b2ml, mhc1zea*; *cd74a, cd74b, ctssb, mhc2dab* from RNA-Seq studies in zebrafish and gilthead sea bream Xu et al., 2017; Najafpour et al. unpulished).

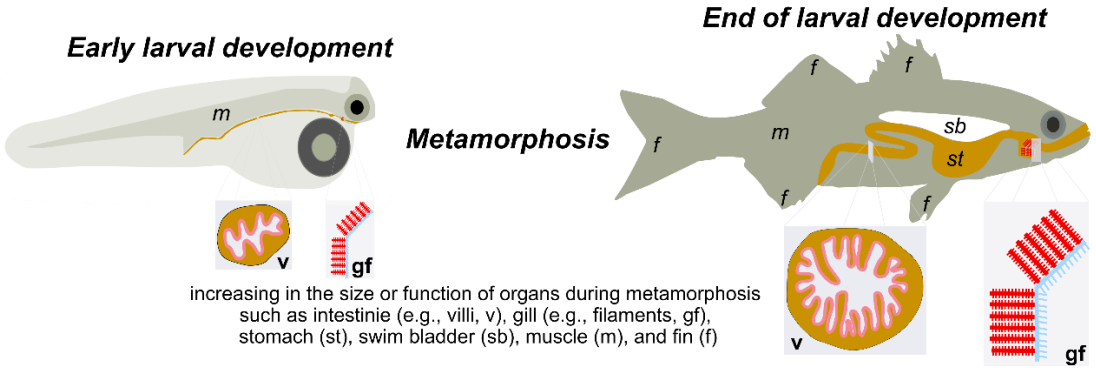
Since larvae transcriptomes are specific to each developmental stage and reflect

the developmental status and function of organs and tissues, this knowledge provides insight into stage-specific physiological competence and can be used to identify changes in larval rearing regimes that could improve larval survival and fitness. In addition to the transcriptome, the effect of age on the larvae microbiome is described in the following section (e.g., **Table 1.1**, Bakke et al., 2015; Bledsoe et al., 2016; Parata et al., 2020, Najafpour et al. unpulished). Similar transcriptional profiles for specific biological processes during larval ontogeny of different species emphasize the fundamental importance of such processes for larvae development as highlighted in **Figure 1.5**. The biological capacity and dominant pathways in larvae may influence and be influenced by microbiota during early ontogeny.

Transcriptomics highlights the substantial structural and functional changes occuring during larval ontogeny

Biological process	*Gene expression pattern*
Development	**Regulation of genes involved in muscle growth is stage-specific** *e.g., skeletal muscle cell differentiation as a stage of myogenesis up-regulated in early larval stages before metamorphosis, including genes of myogenin and myogenic factor 5*
Visual preception	**Visual perception improves in early larval stages (e.g. 7 dph) while the eye is developing during larval metamorphosis** *e.g., up-regulation of beta-crystallin B1 and gamma-crystallin M2 genes at later stages (e.g. 21 and 50 dph)*
Proteolysis	**Endopeptidase activity increases during larval ontogeny** *e.g., an increase in gene expression coding digestive enzymes of the pancreas and stomach (trypsin, chymotrypsin, pepsin) at midmetamorphosis*
Lipid homeostasis	**Increasing expression of gene sets involved in lipid digestion, absorption, and deposition during larval ontogeny** *e.g., increase in the expression of lipases (cel.1), triglyceride re-synthesis processes (e.g. dgat1a), and apolipoproteins (e.g. apoa4a) at the mouth-open stage (endogenous-exogenous feeding), and increase in lipid digestion and lipid transport at pre-metamorphosis compared to post-metamorphosis*
Tissue rearrangement	**Increased cellular fusion, communication and movement during early development and metamorphosis that is essential for morphogenetic processes** *e.g., up-regulation of genes involved in focal adhesion and extracellular matrix (ECM) during metamorphosis*
Immune response	**Immune-associated genes are mainly low abundance in very early stages. Increase expression during larval ontogeny and full activation after metamorphosis.** *e.g., low abundance of innate and adaptive genes such as galactose-specific lectin nattectin, complement, and MHCI&II in younger compared to older larvae*

The morphophysiological changes influence the biological capacity of larvae



Different aspects of larval biology are influenced by the morphophysiological changes:

Feeding and metabolism, the feeding apparatus such as the mouth or digestive system (e.g., stomach or intestine) undergo structural changes to accommodate the changes in dietary requirements.

Locomotion, the development of locomotor structures such as fins, swim bladder, and muscle affect larval swimming capabilities.

Respiration, the changes in respiratory structures, such as gills, can enhance larval oxygen uptake and support their increasing metabolic demands.

Sensory perception, the development of sensory organs, such as eyes and olfactory receptors, can impact larval perception and response to environmental cues.

Immune response, the development of immune organs (e.g., thymus, spleen, and hematopoietic tissues) leads to changes in their structure and function.

Figure 1.5. Significant changes in the expression of genes that contribute to the morpho-physiology of larvae during early development. Examples of genes with significant

differential expression during larval development is presented for each biological process and were mainly obtained from RNA-seq analysis of gilthead sea bream (flexion to mid-metamorphosis by Najafpour et al., 2023). The biological processes highlighted are common vital processes in different fish species. The modifications identified in sea bream larval ontogeny were corroborated by the identification of genes involved in similar processes in other transcriptome studies. Examples of enriched genes include genes involved in development and growth in the Atlantic halibut (e.g., myosin, Douglas et al., 2008) and zebrafish (e.g., *gh1*, *igfbp1*, Xu et al., 2017); visual perception in the Clearhead icefish (Tang et al., 2022), and Atlantic halibut (e.g., crystallin B1 protein, Douglas et al., 2008); digestion-related functions such as proteolysis in *Acipenser baeri* (Song et al., 2015), European sea bass (e.g., trypsin) (Darias et al., 2008), and the three spot cichlid, *Amphilophus trimaculatus* (Hilerio-Ruiz et al., 2021); lipid homeostasis in Atlantic halibut (Douglas et al., 2008), zebrafish (Xu et al., 2017) and the three spot cichlid (Hilerio-Ruiz et al., 2021); tissue remodeling in the Atlantic halibut (Alves et al., 2016); and the immune response in zebrafish (Xu et al., 2017) and common sole (*Solea solea*, Ferraresso et al., 2016).

1.3. An overview of the microbiome evaluation in finfish species from 16S rRNA gene sequencing

The profile of the microbiome of specific organs in fish species and predictions about the influence of a diversity of factors on the microbial community in fish has been generated by 16S rRNA gene sequencing studies (**Table 1.1**). Most fish microbiome studies are relatively recent and are mainly focused on the gut due to its important role in metabolism, growth, digestion, and the immune response (Talwar et al., 2018).

Biotic and abiotic factors that influence the fish microbiome include genetics, species, age, diet, stress, the environmental microbiota, and geographical location and water parameters (e.g., salinity and temperature). Firmicutes and Bacteroidetes are the topmost abundant phyla in the fish intestine, irrespective of the species or study and both beneficial (e.g., *Lactobacillus* spp.) and pathogenic bacteria (*Vibrio* spp.) are part of the gut microbiota. A shift in the microbial community occurs during fish larval ontogeny (**Table 1.1**, Bakke et al., 2015; Bledsoe et al., 2016; Najafpour et al., 2023; Parata et al., 2020), and several studies (Dehler et al., 2017; Najafpour et al., 2023) have shown that the environment has a strong influence on the core microbiota. More studies are needed of the microbiome during larval fish ontogeny to

determine how microbiota are established and stage specific profiles and effectors (e.g., feeding regimes and host morpho-physiology) to open the possibility of microbiome manipulation towards beneficial bacteria.

Table 1.1. A summary of microbiome studies in fish using 16S rRNA gene sequencing. Studies on larvae are less common and are highlighted in grey.

Microbiome experiment	Species	Stage	Source	Highlight	Reference
A meta-analysis of environmental and ecological factors that shape bacterial communities of fish.	Several species	-	Gut	Salinity, trophic level and possibly host phylogeny shape the composition of gut bacteria.	(Sullam et al., 2012)
The effect of feeding with plant meal proteins.	Rainbow trout[1]	Adult	Gut[RAS]	Plant meal proteins cause negative changes in the microbial community.	(Desai et al., 2012)
Analysis of the taxonomic composition and functional diversity.	Turbot[2]	Adult	Gl-tract[RAS]	Proteobacteria and Firmicutes were the dominant phyla in the GI tract.	(Xing et al., 2013)
The impact of suspended sediments.	Clownfish[3]	Larvae	Gill[I]	A shift from 'healthy' to pathogenic bacteria occurred.	(Hess et al., 2015)
The microbiota associated with developing larvae and the bacterial communities in water and live feed.	Atlantic Cod[4]	Larvae	Cod larvae, water, live feed[F]	Relative abundance was Proteobacteria > Firmicutes > Bacteroidetes, and Actinobacteria. Age influenced the composition and diversity of the microbiota. The microbiota of larvae differed from that of water and live feed, particularly in early development.	(Bakke et al., 2015)
The effects of ontogenetic development of the host on the microbiome.	Channel Catfish[5]	Early larvae	Gut[IF]	Microbial communities inhabiting the intestines of developing catfish were dynamic, with significant shifts occurring up to 125 dph. Bacteroidetes, Firmicutes,	(Bledsoe et al., 2016)

				Fusobacteria, and Proteobacteria were abundant.	
Antarctic fish microbiome	Several species	-	Gut[w]	The dominant phyla were Proteobacteria, Actinobacteria, Firmicutes, Thermi, and Bacteroidetes. Abundant genera were *Rhodococcus*, *Thermus*, *Propionibacterium*, *Acinetobacter*, *Streptococcus*, and *Mycoplasma*.	(Song et al., 2016)
The main factors driving fish gut microbiota.	Discus[6]	Larvae	Skin-mucus, gut, water[I]	Feeding on a cutaneous mucus of parents, and environmental water play a substantial role in shaping the structure of the fry gut microbiota.	(Sylvain and Derome, 2017)
The effect of environmental factors and host physiology.	Atlantic salmon[7]	Juvenile	Gut[OC, RAS]	The gut microbiota of fish reared in recirculating laboratory aquarium was significantly different from those in cages. A core microbiota was indicative of host-selective forces.	(Dehler et al., 2017)
The influence of host genetic selection for cold tolerance in shaping the microbiome.	Tilapia[8]	-	Gut[I]	The microbiomes of cold-resistant fish were more resilient to temperature changes. Indicating host selection for cold tolerance shaped the microbiome.	(Kokou et al., 2018)
The effect of selectively bred *F. psychrophilum*-resistant fish.	Rainbow trout[1]	Juvenile	Gut, gill[I]	The abundance of the opportunistic pathogen (*Brevinema andersonii*) increased in the susceptible line. Selectively bred *F. psychrophilum*-resistant trout may harness a more resilient gut microbiome. Alpha diversity did not change in the gills.	(Brown et al., 2019)

The effects of genetics, broodstock, nursery water, and age.	Channel catfish[5]	-	Eggs and gut[IF, OP]	The rearing environment is an important factor influencing the transfer of microbes from water or food to the gut.	(Abdul Razak et al., 2019)
The effect of age, diet, and region of the gut.	Surgeon fish[9]	Juvenile and adult	Gut, food[W]	A core bacterial community of *Epulopiscium* and Brevinemataceae were found in all gut and diet samples, suggesting horizontal transfer from the diet that is retained into adulthood.	(Parata et al., 2020)
The effects of thermal stress	Milkfish[10]	Juvenile	Gut[RAS]	The dominant gut bacteria differed between the gut and water, and were *Cetobacterium*, *Enterovibrio*, and *Vibrio*. Thermal stress changed the gut microbiome.	(Hassenrück et al., 2020)
The factors shaping fish microbiomes in a heterogeneous inland water system.	Several species	-	Skin[W]	Human-induced eutrophication/ sporadic nutrient pollution events promoted dysbiosis of the fish skin community.	(Krotman et al., 2020)
The influence of tank biofilm on microbial communities.	Atlantic Salmon[7]	Juvenile	Skin, gill, water and tank biofilm [RAS]	The richness of the microbiome of the water and biofilm had a positive correlation with skin and digesta. The tank biofilm directly influenced fish-associated microbial communities.	(Minich et al., 2020)
The effect of salinity.	Pike[11]	Fry	Gut[W, OP]	No significant differences in the abundance of gut bacteria species were observed. However, salinity (7 ppt) affected the abundance of some taxa, Planctomycetes, Rhodobacterales, Alphaproteobacteria	(Dulski et al., 2020)

Effects of environmental (diet and water sources) and host early ontogenetic development on microbiota.	Lake Sturgeon[12]	Larvae	Gut[IF]	Bacterial diversity declined, and community composition differed significantly among stages (pre-feeding, one and two weeks after exogenous feeding began). Water was a stronger driver of gut community members than diet.	(Razak and Scribner, 2020)
Deep-sea fish microbiome.	Several species	-	Gut[W]	The gut microbiome was dominated by the Proteobacteria (genus Photobacterium) in two fish species and by the Spirochaetes (genus Brevinema) and Tenericutes (unclassified Mycoplasmataceae) phyla in another species.	(Iwatsuki et al., 2021)
White sea fish microbiome.	Several species	-	Gut[W]	Proteobacteria, Actinobacteria, Firmicutes, Tenericutes, and Fusobacteria are abundant in the marine fish gut. Bioluminescence is a remarkable feature of some gut bacteria in marine fish.	(Burtseva et al., 2021)
The influence of gut microbiota on the innate immune responses and gut morphology.	Atlantic Cod[4]	Larvae	Gut[I]	Microbiota altered the innate immune response and gut morphology. Innate immune genes were low with non-pathogenic bacteria. Microvilli were shorter in larvae from conventional conditions with a higher microbiota density compared to germ-free larvae.	(Vestrum et al., 2021)

The effect of transition and laboratory acclimation of wild-caught fish.	Rainbow darter[13]	-	Gut[W,IF]	The gut contents of wild fish were dominated by Proteobacteria (35%) and Firmicutes (27%), while lab fish were dominated by Firmicutes (37–47%) and had lower alpha diversity. Prolonged acclimation to a new environment is required to achieve a stable gut microbiome in wild-caught fish.	(Restivo et al., 2021)
Seawater recirculation aquaculture system (RAS) microbiome.	Chinook salmon[14]	Adult	Digesta (feces), water, feed[RAS]	The dominant phyla in Chinook salmon faeces were Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria. Most taxa-dominating feces samples were also present in the surrounding water or feed, suggesting that the microbiome is shared with the ambient environment.	(Steiner et al., 2021)
The effect of commercial site, species, and type of sample.	European sea bass and gilthead sea bream	-	Eggs and water[RAS]	The dominant genera were *Vibrio* (7.1%), *Glaciecola* (4.8%), *Pseudoalteromonas* (4.4%), and *Colwellia* (4.2%) in eggs and water. A non-significant reduction occurred in the egg bacterial load by iodine-based disinfectants.	(Najafpour et al., 2021b)
The effect of benzo[a]pyrene on active gut microbiomes.	Fathead minnows[15]	Juvenile	Gut[I]	The exposure to BaP altered the community composition of the active microbiome and resulted in a change in abundance of taxa associated with hydrocarbon degradation.	(DeBofsky et al., 2022)

Latin name*: 1-* *Oncorhynchus mykiss*, 2- *Scophthalmus maximus*, 3- *Amphiprion percula*, 4- *Gadus morhua*, 5- *Ictalurus punctatus*, 6- *Symphysodon aequifasciata*, 7- *Salmo salar* L., 8- *Oreochromis aureus*, 9- *Acanthurus triostegus*, 10- *Chanos chanos*, 11- *Esox lucius*, 12- *Acipenser fulvescens*, 13- *Etheostoma caeruleum*, *14-*

Oncorhynchus tshawytscha, 15- *Pimephales promelas.* **RAS**: Recirculating aquaculture system, **IF**: indoor flowthrough water system, **I**: indoor, **OP**: outdoor pond, **OC**: outdoor cage, **W**: wild

Microbiota diversity, gene content, and prediction of their potential functions can be established using a range of approaches, but 16S rRNA gene sequencing is prevalent (Gosalbes et al., 2011, section 3). However, there are a number of limitations of 16S rRNA gene sequencing including: 1- the taxonomic classification of bacteria is difficult at the species level since most current approaches use second-generation short-read sequencing (approximately 400bp) that is most robust for bacterial taxonomy at the genus level. Nonetheless, the advent of third-generation long-read technologies (e.g., Oxford Nanopore Technologies) can overcome some of the current limitations of short-read sequencing (Meslier et al., 2022); 2- functional predictions arising from the microbiome profile of samples with large proportions of unknown organisms may not represent the functional diversity in that sample due to the low coverage of database reference profiles; 3- it is unclear whether the predicted taxa come from viable bacteria, and if they are active or silent.

While metagenomics analyzes the DNA content, meta-transcriptomics characterizes the RNA content and gives insight into microbial community gene expression, metabolic activity and taxonomic diversity (Bashiardes et al., 2016; Stewart, 2013). Meta-transcriptomics can identify active disease-driving bacteria within a microbial community. The application of metagenomics and meta-transcriptomics is advantageous since microbiota and their functional role can be determined and this approach has been used to study immune system maturation and inflammation, maintenance of the intestinal barrier (e.g., thickness and composition of the mucus layer, Rolhion and Chassaing, 2016), and the response of commensal bacteria to xenobiotics and their contribution to variations in drug efficacy and toxicity in humans (Bashiardes et al., 2016). Meta-transcriptomics techniques have still not been widely used, but the potential of this approach is shown by studies of periodontitis where the progression of the disease was associated with enrichment of specific pathways (e.g., cobalamin biosynthesis, proteolysis, and potassium transport) (Yost et al., 2015) in the whole community. However, studies to establish intra-interactions leading to the colonization by specific genera or species and the proportion of active bacteria in a given condition are still uncharacterized in fish.

1.4. The importance of the microbiome in aquaculture

The microbiome is a term that describes the genome of all symbiotic and pathogenic microorganisms, including bacteria, archaea, viruses, and fungi living in and on all vertebrates (Taneja, 2017). Generally, metagenomics approaches have been focused on the bacterial genome rather than all the microorganisms present in an ecosystem. Since much of the literature on fish microbiomes is focused on the bacterial community the present review is focused primarily on this aspect. The impact of microbiota on organisms is unsurprising considering the high loads of bacteria that reside on the internal and external surfaces of both terrestrial and aquatic organisms. This is exemplified by the estimated 3 - 10 trillion bacteria associated with animals, which is more than all the cells of the host and corresponds to 500 – 1000 bacterial species in humans and 100 - 600 bacterial species in fish and their environment (Gilbert et al., 2018; Najafpour et al., 2023; Taneja, 2017).

One of the prominent roles of the gut microbiome is its involvement in food digestion. The gut microbiota produces a range of enzymes that hydrolyze food (e.g., carbohydrate, cellulase, phosphatase, enzyme, lipase, and protease) and can also use break down products to synthesize amino acids and vitamins (e.g., B12, Balcázar et al., 2006a; Ray et al., 2012; Sugita et al., 1991; Wu et al., 2015). The protective role of the resident host bacteria is another important function of the microbiome and it can prevent colonization by pathogenic bacteria through interactions and competition and by enhancing the host immune response (Kamada et al., 2013). Commensal bacteria prevent pathogen colonization of the intestine microenvironment by releasing inhibitory factors such as bacteriocins and proteinaceous toxins. They can also alter the environmental conditions away from the pathogens' optima, such as the change in pH caused by *Bifidobacterium* (Balcázar et al., 2006a; Fukuda et al., 2011; Kamada et al., 2013; Perry et al., 2020). Furthermore, antiviral effects of bacteria are another antagonistic interaction that favours the host [e.g., the antiviral activity of two strains of *Vibrio* spp. against the infectious hematopoietic necrosis virus (IHNV) and *Oncorhynchus masou* virus (OMV)] (Direkbusarakom et al., 1998). From the studies that exist the microbiome's role in aquaculture is likely to be important and an increase in commensal bacteria can enhance fish growth and survival and dysbiosis caused by pathogen colonization inhibits these production traits (Brown et al., 2012; Butt and Volkoff, 2019; Infante-Villamil et al., 2021).

Water is one of the substantial sources of bacterial transfer to both terrestrial and aquatic organisms where multiple factors influence bacterial growth and diversity in water, including the assimilable organic carbon concentration and composition, salinity and pH value,

and temperature (Fossmark et al., 2020; Novoslavskij et al., 2016; Vital et al., 2010; Zhang et al., 2016). Two main approaches are applied to manage and limit microbial growth in human drinking water during transport to the consumer, including inclusion of a disinfectant (e.g., chlorine) and the use of high-quality source waters or multi-barrier treatments to minimize nutrient availability in the source water (Dai et al., 2020). In addition, to the effectiveness of each approach, drawbacks include, harmful disinfection by-products, the potential of loss of efficacy of some disinfectants for some pathogenic bacterial genera or species, microbial antibiotic resistance, and corrosion of facilities by repeated use of treatments as occurs in aquaculture (Dai et al., 2020; Kooij and Wielen, 2013; Li and Mitch, 2018; Najafpour et al., 2021b; Richardson, 2003; Sevillano et al., 2019; Wang et al., 2013).

Both physical filters and disinfectants are used by the aquaculture sector to control microbiota. However, despite this challenges remain and include pathogen outbreaks, larvae mortality, and the poor efficacy of disinfectants or procedures (Najafpour et al., 2021b). In addition, ethical issues have limited disinfectant usage in aquaculture (Grigorakis, 2010; Rigos and Troisi, 2005). The control of the microbiome in fish farms is more complex than for terrestrial organisms due to the impossibility of using a high doses of disinfectants in water, the low performance of treatments (e.g., UV), the large volumes of water, and low attention to the role of feeding or other human interventions as a potential source of pathogens (Najafpour et al., 2023; Olafsen, 2001). The detection of bacteria in aquaculture farms and fish mortality due to pathogens highlight the growing problem of bacterial resistance in the sector (Delphino et al., 2019; Itchell et al., 1997; Najafpour et al., 2021b, 2023). Vaccines and antibiotic usage have been effective against some acute pathogens, but they do not provide a solution so far for all pathogens due to species/strain-specific effects, lack of effective adjuvants and basic knowledge about the immune response, and antibiotic resistance (Adams, 2019; Olafsen, 2001; Preena et al., 2020; Smith, 2008).

Interventions to modulate the microbiome in hatcheries or fish farms will need much more knowledge of the factors that drive the microbiota and the fish's biological response. Larval development is a critical period during fish production due to the substantial morphophysiological changes, changing feed source and the fragility and vulnerability of larvae. Microbiomes and transcriptome can be used to determine bacterial establishment, fish development, and host-bacteria interactions in the hatchery phase during larval ontogeny. Knowledge about the proceeding factors can be linked to the identification of factors that inhibit or enhance the establishment of a favorable bacterial communities, characterization of potential

probiotics, and identification of the best approach to modulate microbiota based on fish biology.

1.4.1. Modulation of gut or water microbiota in fish aquaculture

Probiotics, prebiotics, synbiotics, para-probiotics, and postbiotics are potential therapeutic approaches to modulate fish microbiota and prevent dysbiosis in fish (Vargas-Albores et al., 2021). Fish-derived probiotics isolated from internal organs (e.g., gastrointestinal tract, gill, gonads) can be introduced as putative probiotics, and examples include the application of different lactic acid bacteria isolated from the gut microbiota of aquaculture species to impede pathogens such as *Aeromonas salmonicida* and *Vibrio anguillarum* (Alonso et al., 2019; Vargas-Albores et al., 2021). Overall, the safety and efficacy of probiotics and prebiotics in aquaculture needs further evaluation in commercial-scale systems since maintenance of high viability during their preparation and storage is still a bottleneck for commercial probiotic production (Wang et al., 2008; Vargas-Albores et al., 2021). The conditions under which probiotic usage is optimized can differ in farms across different geographical locations, and the safeness of new probiotic species/strains may not be the same as traditionally used probiotics (Wang et al., 2008). The first steps to establishing effective pro- and prebiotics include large-scale microbiome studies to screen core and divergent microbial communities, drivers of bacteria establishment, and identification of potential probiotics and the elaboration of the best strategies for modulation of microbiota based on detailed knowledge about aquaculture farm conditions (Najafpour et al., 2023). The effects of probiotics are determined by dose, treatment duration, and host age (Brugman et al., 2018). Therefore, it is essential to understand fish microbiome establishment during ontogeny. Parallel, transcriptome studies to assess the response of the host to probiotics can contribute to determine fish health and immunity at each developmental stage (Najafpour et al., 2023; Nayak, 2010).

Water-derived probiotics are also an interesting possibility for aquaculture since probiotics can be added directly to water so they modulate the microbiota of rearing water/ponds, and in this way improve water quality parameters, facilitate transfer of beneficial bacteria to fish, and eventually enhance the performance of farmed species (Cruz et al., 2012; Jahangiri and Esteban, 2018; Tabassum et al., 2021; Wang et al., 2005). Direct treatment of rearing water with probiotics is suggested as a viable and effective approach, especially for larviculture where the administration of pre- and probiotics by other approaches (e.g., feeding or injection) is limited (Jahangiri and Esteban, 2018). The effect of different factors on rearing water microbiota (e.g., live foods and source water microbiota) and the possible interactions

between bacteria add complexity to the modulation of the microbiome in aquaculture sites (Najafpour et al., 2023). Therefore, basic knowledge of microbiome establishment in aquaculture sites is the first step and then the effectiveness and cost-benefit application of therapeutic approaches need to be assessed and optimized on an industrial scale (Lauzon et al., 2014).

1.5. Microbiota colonization of fish eggs

During fish ontogeny the microbiota initially colonize eggs and can potentially influence the survival rate of embryos through modulating pathogen outbreaks and they are important for the bacterial colonization of the hatched larvae (Hansen and Olafsen, 1999; Merrifield and Rodiles, 2015; Najafpour et al., 2021b; Olafsen, 2001). The factors that influence the egg microbiome include water microbiota, genetics, disinfectants, species, and the geographic location of the production site (Abdul Razak et al., 2019; Bone et al., 2020; Najafpour et al., 2021b). Overall, the site or surrounding water was the main factor that determined the microbiota of European sea bass (*D. labrax*) and gilthead sea bream (*S. aurata*) eggs (Najafpour et al., 2021b). In sea bass and sea bream the abundance of relatively few OTUs was different between the water and egg microbiota; the higher relative abundance of some bacterial genera in eggs was proposed to most likely reflect the higher dependence of these bacteria on the egg environment (Najafpour et al., 2021b). Biofilm formation and energy needs are two predicted interactions between bacteria and surfaces (Tuson and Weibel, 2013). The basis of the interaction between the egg/embryo surface and microbiota is still unknown but is worthy of more attention in future studies.

Ontogenetic characterization of the intestinal microbiota of channel catfish, *Ictalurus punctatus*, juveniles maintained under controlled conditions, suggests simple microbial communities inhabit the catfish gut based on the low alpha diversity obtained from the OTU number and Chao1 species richness estimates (Bledsoe et al., 2016). In contrast, a higher number of OTUs were detected in another study of the gut from channel catfish sampled from a recreational fishing pond (Larsen et al., 2014). The low diversity of microbiota reported in the first study was hypothesized to be the result of Povidone-iodine treatment of eggs and the rearing of the fish under a highly controlled environment (Bledsoe et al., 2016). However, a study of sea bass and sea bream egg microbiomes revealed a non-significant effect of iodine-based disinfectants on the microbiome in hatchery conditions (Najafpour et al., 2021b). The divergent outcome of published studies on microbiota in aquaculture highlights the need for

more systematic studies of the microbiota in commercial aquaculture and the determination of how the environment (e.g., closed circuit, open circuit, under cover, open etc) can influence bacterial diversity and the colonization of larvae by pathogens.

1.6. Microbiome establishment and drivers during larval ontogeny

Ontogenic analysis of microbiota and their interaction with the host is of great interest in aquaculture as it may be useful to address bottlenecks such as poor larval quality and unpredictable mortality due to pathogens in early life stages. Furthermore, knowledge about the initial bacterial colonization of the host in early life stages is a priority since it will enable interventions to modulate microbiota and promote the development of balanced and optimal bacterial communities. For example, in several aquaculture species (channel catfish, rainbow trout *Oncorhynchus mykiss*, and coho salmon *Oncorhynchus kisutch*) the gut microbial community in different age groups differed (Bledsoe et al., 2016; Ingerslev et al., 2014; Romero and Navarrete, 2006) The change in the gut microbiota of rainbow trout occurred at the shift from endogenous (yolk sac-based) to exogenous feeding (Ingerslev et al., 2014). The significantly higher bacterial load (qPCR of 16 S rDNA) and Shannon diversity in the fish gut after 26 and 49 days post feeding onset compared to fish a day before first feeding indicates the importance of feed and age on the microbial diversity in rainbow trout (Ingerslev et al., 2014).

Feeds can influence the microbial composition of the gut in larvae in two manners: I- the feed/diet components can induce a change in the abundance of gut microbiota; this is possible when fish larvae are either fed with an artificial diet (e.g., salmonids) or fed with live feeds (e.g., sea bass and sea bream), and II- feed can act as a vector for the transfer of microorganisms to fish larvae; this is more likely when fish are fed with live feeds since feed quality is less controlled than with artificial feeds. Examples of feeding effects include the increased abundance of the *Firmicutes* phylum in rainbow trout fed with plant diets (10% of the fish meal was replaced by pea meal) compared to marine diets (fish meal and fish oil only) (Ingerslev et al., 2014). The increased abundance of *Cetobacterium* and *Mycoplasma* in intestinal microbiota of largemouth bass, *Micropterus salmoides*, fed with diets in which 30% fish meal was replaced by soybean meal and fermented soybean meal, respectively (He et al., 2020); and the increased abundance of *Lactobacillus* and *Photobacterium* and the decreased abundance of *Prevotella* and *Paraprevotella* using a plant-protein-based diet (30% fish meal replacement, Niu et al., 2020).

Internal and external factors like environment, feeding, genetics, and antibiotics

influence microbiome establishment in larvae. Colonization of the gut with unstable and dynamic microbiota at early life stages and the change toward a more stable population with development is suggested to be typical of terrestrial and aquatic organisms like humans and fish (Kumbhare et al., 2019; Llewellyn et al., 2014). Fish are directly in contact with a high bacterial load in water, and their body temperature depends on the environment since they are ectotherms. Therefore, environmental factors potentially have a bigger effect on microbial populations and microbe-host interactions in aquatic organisms (e.g., fish) compared to terrestrial organisms. The effect on the microbiome of environmental factors was suppressed under controlled hatchery conditions as revealed by a large-scale study of European sea bass and gilthead sea bream eggs and larvae from industrial hatcheries across Europe (Najafpour et al., 2023). Although, the microbiome of eggs differed between European sea bass and gilthead sea bream (Najafpour et al., 2021b). The differences in the microbiome of European sea bass and gilthead sea bream eggs even when obtained under similar hatchery conditions suggests that the egg chorion structure and chemistry may determine the microbial community at the embryonic stage. Curiously, the larval stages of gilthead sea bream and European sea bass from European hatcheries had similar microbial communities presumably due to similar management regimes, controlled environmental conditions and feed (Najafpour et al., 2023) and similar host-selective pressure.

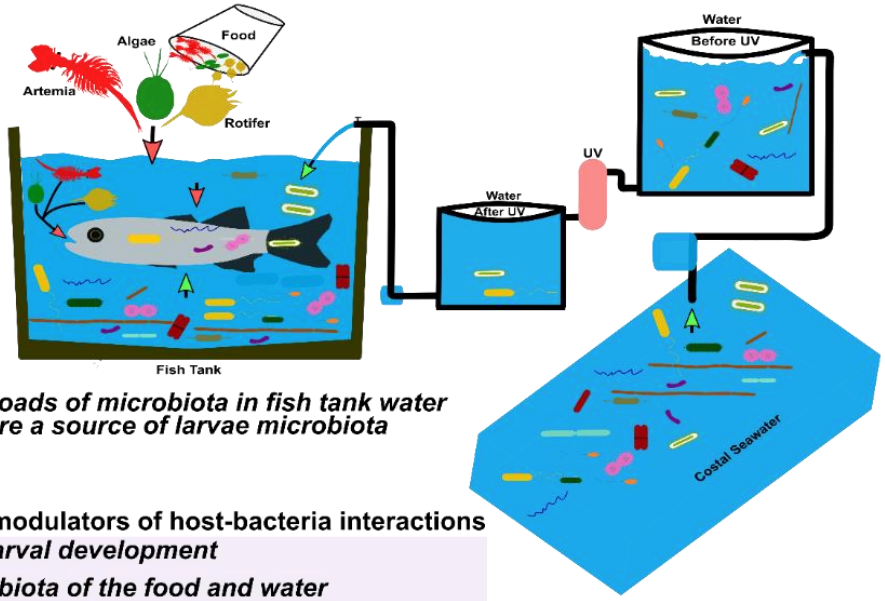
Overall, larval development, fish-microbe interactions, feeding, bacteria of water, and farm locations were suggested as the main factors establishing larvae microbiota during ontogeny by two recent large-scale studies of European sea bass and gilthead sea beam hatcheries (Najafpour et al., 2023, 2021b). The microbial community of cod larvae was proposed to change due to host selection and the development of the intestinal system (Bakke et al., 2015). Previous fish – microbiota examples have highlighted the importance of the diet for the establishment and modulation of the microbial composition of larvae (Avella et al., 2010; Borges et al., 2021; Deng et al., 2022; Dimitroglou et al., 2011; Lobo et al., 2014; Rimoldi et al., 2021). In contrast, Bakke et al. (2013) suggested live feed was not a substantial determinant of the microbiota associated with cod larvae since they did not find the modulation of larval microbiota with copepod and rotifer. The contradictory opinions regarding the effect of feeds could result from the specific conditions of each study. Nonetheless, the observation in most studies that probiotics generally influence fish microbiota, fish growth, and survival, favours the prevailing idea and raises questions about the observations on cod. Examples of the positive impact of probiotics include the enhancement of gilthead sea bream larval survival using live foods with candidate probiotics (Makridis et al., 2005), the change in the microbiota

composition of cod larvae by rotifer enrichment (Korsnes et al., 2006), and the increase in the growth and survival of tilapia *Oreochromis niloticus* L. juveniles by a diet supplemented with lactic acid bacteria (Apún-Molina et al., 2009). A large-scale study on European sea bass and gilthead sea bream larvae identified core microbiota across live feeds (algae, rotifer, and *Artemia*), water, and larvae that indicated that feeds and water affected larvae microbiota (Najafpour et al., 2023). Bacteria antagonistic and bacteria-host interactions have been less studied in fish despite their known role on the microbial community (Caruana and Walper, 2020). Microbe-microbe and microbe-larvae interactions are probably two substantial factors that influence the establishment of microbiota during larval development. For example, a high relative abundance of a given bacterial genus (e.g., *Mesoflavibacter* in rotifer) or the ability of a bacterial genus to form a biofilm (e.g., *Vibrio*) was hypothesized to be effectors of microbial composition in gilthead sea bream and European sea bass larvae (Najafpour et al., 2023). The host selection of beneficial bacteria (e.g., *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, and *Blautia*) was suggested as another factor to establish initial microbial communities at start feeding compared to mid-metamorphosis (Najafpour et al., 2023). Taking into consideration the overall high abundance of potential pathogenic bacteria (e.g., *Vibrio*) in hatcheries, there is a high risk of dysbiosis and colonization of larvae by pathogens instead of beneficial bacteria. Therefore, the first steps needed to improve colonization of larvae with a favorable community is to increase the proportion of beneficial versus pathogenic bacteria by removing the sources of pathogens and/or using selected mixtures of bacteria, that can colonize and generate biofilm in the rearing environment and outcompete harmful bacteria (Guéneau et al., 2022). The main factors influencing bacteria establishment during larval ontogeny include water, foods, and the developmental status of larvae (**Figure 1.6**).

Factors driving the microbiome during larval ontogeny

Factor	*influence*
Water	**Environmental water is a source of bacteria, and tank water microbiota influences the initial colonization and the microbial community of larvae.** *e.g., direct influence of the tank biofilm on microbial communities of the skin and gill*
Feed	**Feeds can influence the larval microbial communities through the release of nutrients and by transfer of bacteria to fish larvae or to tank water.** *e.g., Plant diets increase Firmicutes abundance or high bacterial abundance in live feeds (Mesoflavibacter in rotifer) alters the microbiota abundance in larvae*
Larval development	**Specific niches in fish larvae (e.g. gut) can determine the initial microbial colonization of larvae while morphophysiological development of larvae may change unstable initial microbiota to more stable microbiota.** *e.g., host selection of beneficial bacteria (e.g., Bifidobacterium and lactobacillus) to establish mutualistic relationships during early development, and development to juvenile may increase bacterial attachment and cause biofilm formation*
Others	**The high impact of genetics, antibiotic usage, and specific water parameters (e.g. salinity) on the microbial community may not be universal in all aquaculture sites and depends on site-specific rearing conditions.**

Feeding with live feeds may influence larvae microbiota during ontogeny directly or indirectly



High loads of microbiota in fish tank water are a source of larvae microbiota

Potential modulators of host-bacteria interactions

Stage of larval development
The microbiota of the food and water
Diet composition and nutrients
Geographical location and environmental factors

Figure 1.6. A summary of the main factors identified that influence bacteria establishment during larval ontogeny in aquaculture farms mainly under recirculating aquaculture system (RAS) or indoor flow flowthrough water system conditions in hatcheries (see **Table 1.1**). In hatcheries since the water used goes through a treatment system that includes multiple filters, UV, and ozone treatment the influence of the environmental microbiome is relatively limited unless there is a problem with the production system. The increasing tendancy for larval production is towards recirculating systems and semi-closed systems to maintain more controlled conditions, especially in countries with limited access to land and water (Martins et

al., 2010), and also due to increasingly strict regulation for water management and allowed discharges by the aquaculture sector, including EU environmental legislation (e.g., the Water Framework Directive, Directive 2000/60/EC; the Marine Strategy Framework Directive, Directive 2008/56/EC). This means that live feeds that come from a diversity of sources and have a very variable microbiome can have a large impact on the bacteria available for colonization of larvae.

1.7. Bacteria-host interactions

Microbial community formation in living organisms is complex since both bacteria-bacteria and bacteria-host interactions can lead to formation of a dominant population in a specific niche. Metabolic relationships between different microbiota are essential for sustaining a microbial community (Wolin et al., 1997). When microbes interact with a host or themselves, they may positively benefit from the association, through mutualistic, syntrophic, and commensalistic interactions, or may negatively affect each other (antagonism and competition). Antagonistic interactions mediated by marine bacteria can shape microbial community structure (Wietz et al., 2013). Fish have an active interaction with microorganisms in their environment, which sometimes includes high bacterial loads that can increase the interaction between fish and microbes. The high interaction of fish with the environmental microbiota may have shaped their immune function and over an evolutionary timeframe could explain the emergence of multiple immune gene isoforms (e.g., C3 and Cfh) that potentially accommodate the vast diversity of microbiota to which they are exposed and enhance the fish immune response (Najafpour et al., 2020). For example, fish intestinal mucosa interface with a diverse and highly abundant (10^7–10^{11} bacteria per g) community of microorganisms, (Nayak, 2010). The host has to tolerate and promote its own beneficial microflora, while still being able of recognize and trigger an immune response against pathogens. The host's innate and adaptive immunity has been proposed to regulate the symbiotic host-microbiome interactions in several tissues like the gill, skin, and gut (Yu et al., 2020). Examples of potential innate immune responses in fish that may interact with microbiota are mucous, which is secreted by intraepithelial goblet cells and a diversity of bioactive factors secreted by epithelial cells into the mucous, including antimicrobial peptides, complement, mucins, enzymes, cathelicidins, piscidins, and defensins as well as signaling molecules, and toxin-neutralizing enzymes (Lee et al., 2021; Pérez et al., 2010; Silphaduang et al., 2006; Uzzell et al., 2003; Zou et al., 2007). Adaptive immunity has also evolved in fish that may be active to modulate microbiomes (e.g.,

internal gut microbiota) such as intraepithelial lymphocytes of predominantly Ig-positive B cells in carp *Cyprinus carpio L.* and European sea bass and CD8-α-positive cells (Cytotoxic T-cells) in European sea bass (Abelli et al., 1997; Picchietti et al., 2011; Rombout et al., 1993), and leukocytes in the lamina propria including lymphocytes, macrophages, granulocytes, and dendritic-like cells constitute the gut-associated lymphoid tissues (Rombout Jan et al., 2011).

1.7.1. Probiotics as beneficial bacteria with positive effects on the host

The contribution to the host of its microbiota is evident when considered in the context of studies showing the benefits of probiotics. A range of bacteria commonly found in the microbiota of fish have been used as probiotics by inclusion in feeds, and include species of *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Enterococcus*, *Carnobacterium*, *Shewanella*, *Bacillus*, *Aeromonas*, *Vibrio*, *Enterobacter*, *Pseudomonas*, *Clostridium*, and *Saccharomyces* (Nayak, 2010). The positive effects of probiotics have been proposed to include improved nutritional uptake, growth, disease resistance, etc. A number of studies support the positive influence of probiotics and the addition of viable *Lactococcus lactis* subsp. lactis, *Lactobacillus sakei*, and *Leuconostoc mesenteroides* to the diet of rainbow trout increased phagocytosis of heat-inactivated *Aeromonas salmonicida* by leukocytes (Balcázar et al., 2006b). In the case of the European sea bass (*D. labrax*) *Lactobacillus delbrueckii*, isolated from the intestinal microbiota of adults was used as a probiotic and significantly increased T cells and acidophilic granulocytes in larvae and post-larvae (Picchietti et al., 2009). Nile tilapia, *Oreochromis niloticus* provided with a probiotic composed of *Bacillus subtilis* and *Lactobacillus acidophilus*, had an increased immune response, e.g. neutrophil adherence and lysozyme activity, and improved resistance to a bacterial challenge (Aly et al., 2008). In the case of invertebrates the provision of *Saccharomyces cerevisiae* YFI-SC2 in the diet of crayfish (*Procambarus clarkia*) improved their growth performance, and was associated with improved intestinal morphology, a modified intestinal microbiota and an improved immune response (e.g. Lysozyme and prophenoloxidase) and resistance to a disease challenge (*Citrobacter freundii*) (Xu et al., 2021).

1.7.2. The potential integration between host biological processes based on transcriptomics and the microbiota dynamic during early development

The effect of age on the host microbiota or transcriptome has been studied separately in most studies. However, analysis of the microbial community and their metabolites and host transcriptomes need to be considered together to provide a comprehensive picture of

the interaction between the host and microbiota. Case studies of the positive impact of probiotics on specific genes associated with the digestive system and immune system in the human neonate or fish give clues to a promising area of research with practical implications. Nowadays, omics approaches using high-throughput sequencing can capture host-microbe interactions since many external factors influence the establishment of the microbiome, but the host has a fundamental role in shaping the microbiome. For example, in humans analysis of the microbiome and transcriptome of human diseases such as asthma and Carotid Atherosclerosis clearly reveal the importance of the microbe-host interactions (Chan et al., 2015; Ji et al., 2021; McCauley et al., 2022).

To our knowledge, no study has analyzed transcriptome and microbiome changes simultaneously during fish larval development. In addition, the application of a Dual RNAseq approach (as a complementary omic approach, including RNA-Seq and 16S rRNA gene metabarcoding sequencing) in fish is at a preliminary stage and may still not be cost-effective for large scale aquaculture studies. This technique has unraveled host–pathogen interactions, limited to pathogens challenges in a few adult fish species, including the interaction between *Platax orbicularis* and the infectious challenge of *Tenacibaculum maritimum* (Luyer et al., 2021) or between channel catfish, *Ictalurus punctatus*, and infectious challenge of *Yersinia ruckeri* (Yang et al., 2021). In our study of gilthead seabream, the main significant modifications identified when the results of microbiome and transcriptome analyses carried out on larvae collected at the same time were compared (described in this review in the appropriate section) led us to highlight possible pathways or biological processes that may be influenced by the microbiota, which eventually suggested their potential modulation with the microbiome. During development the biological processes or pathways that are significantly modified in different fish species, include genes associated with GH-IGF axis, thyroid axis, extracellular matrix, metabolism (e.g., protein and lipid metabolism), and the immune response (**Figure 1.7**). In relation to the microbiome, during development the microbiota is dynamic and changes presumably due to feeding regime and morphophysiological development (**Table 1.1**, **Figure 1.6**). Differentially abundant bacteria lead to enrich processes or pathways at early (e.g., starch and sucrose metabolism pathway) or late (e.g., biosynthesis of amino acids) larval stages (**Figure 1.7**).

We propose a potentially significant correlation between the plasticity of specific biological processes and microbiota dynamics during larval development in **Figure 1.7**. What is currently missing are experiments that analyze the microbiome and transcriptome

simultaneously during larval development of specific species. Nonetheless, we have used the results of our recent separate studies that compared the transcriptome (using RNA-Seq, Najafpour et al. unpublished, **Figure 1.5**) and microbiome (using 16S rRNA gene sequencing, Najafpour et al. 2023, **Figure 1.6**) changes during gilthead sea bream larval development (between early larval stage and mid-metamorphic stage) as a baseline for other studies to compare and integrate transcriptome and microbiome results from oncogenic perspective. Although our studies were independent since both targeted the same species and were performed at hatchery conditions, we expect similar significant modifications in highlighted processes, possibly with minor exceptions, if we analyzed microbiome and transcriptome simultaneously in a single study using the same fish batch.

Based on transcriptomics analyses, the gene sets involved in food metabolism and energy production pathways such as the tricarboxylic acid (TCA) cycle, the final oxidative pathway to produce energy/ATP from carbohydrates, fats, and amino acids were enriched during sea bass and Clearhead Icefish larval development (Darias et al., 2008; Tang et al., 2022). The expression of gene sets involved in innate and adaptive immune responses were significantly increased during the ontogeny of zebrafish, gilthead sea bream, and Senegalese sole (Xu et al., 2017; Najafpour et al. unpublished). Although a part of the increase in the immune response and metabolism of larvae may be associated with ontogenetic maturation, which may affect microbiota colonization such as biofilm formation or microbiota composition and abundance, reciprocally, the effect of microbiota to change host responses is predictable with considering the changes in the microbial community with age which may play a fundamental role in immune and metabolic homeostasis (**Figure 1.7**). In addition to immune and metabolic processes, review of transcriptome studies using RNA-seq during larval development suggested a significant changes in genes of the thyroid axis (e.g., *dio1*, *dio2*, *dio3*, and *tg*), GH-IGF axis (e.g., *gh1*, *igfbp1*), and focal adhesion and ECM receptor interactions (e.g., *col1a1*, *col1a2*, **Figure 1.7**).

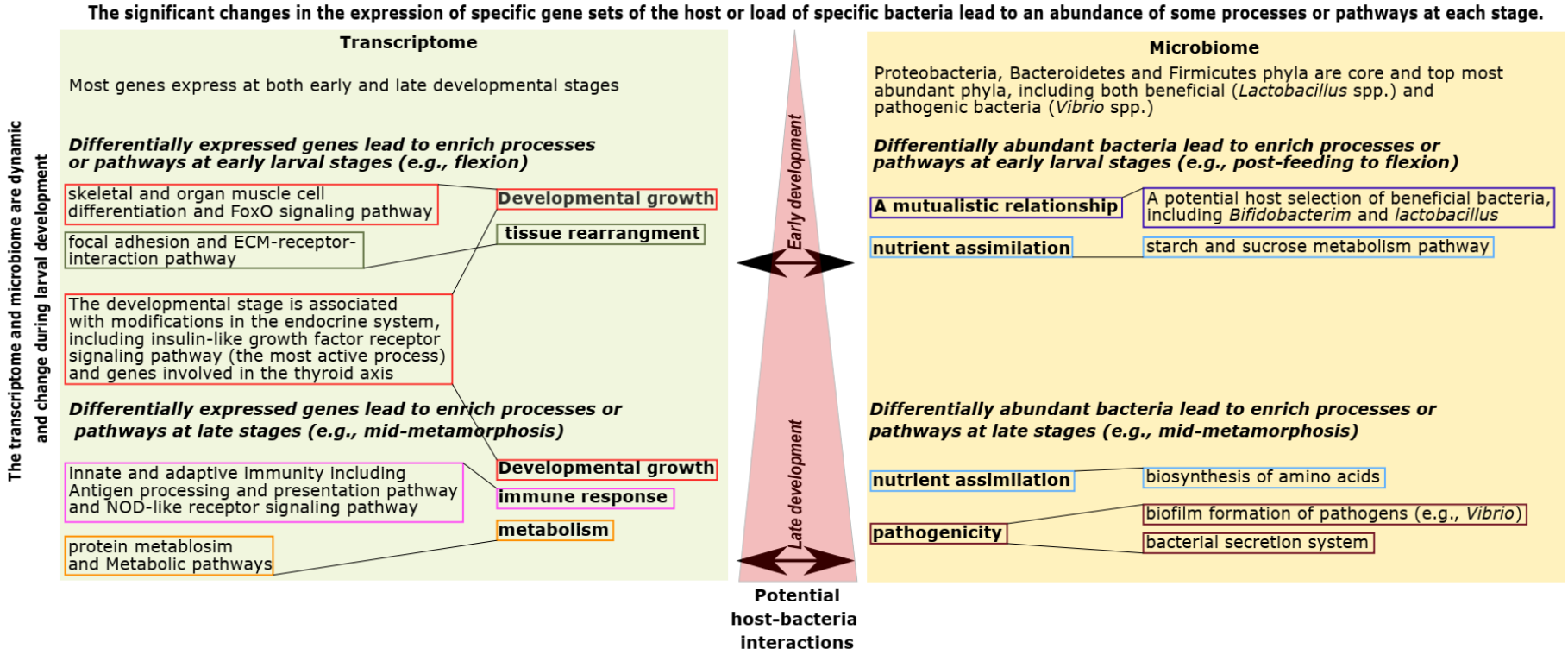


Figure 1.7. Potential significant interactions between host-bacteria during larval development. The interactions are predicted based on the highlighted significant changes in transcriptome and microbiome from early to late larval stages.

Since both pathogenic and beneficial bacteria have the potential to colonize larvae, the fate of larvae may depend on the relative abundance of different bacteria and the interaction of the larvae with them. An abundance of beneficial bacteria may promote digestive processes and improve nutrient assimilation and initiate "training" of the immune response so it is nonresponsive towards the normal microbiota and bias the energy balance towards growth and normal development. In contrast, dysbiosis may lead to poor assimilation of nutrients and expenditure of energy on the immune response and diverting it away from growth and normal development and lead to low larval quality and mortality (**Figure 1.8**).When we integrate dynamic of transcriptome and microbiome during larval development, focusing on significant changes, leads us to propose a model representing the biological processes with potential association with microbiota establishment, including the GH-IGF axis, thyroid axis, pathways involved in the extracellular matrix, and immune and metabolic processes (**Figure 1.8**).

In addition, the proposed interactions between microbiota and host in a few previous studies in fish or a mammalian model are intriguing further attention to our highlighted pathways or interactions between host-bacteria. For example, the potential effect of the gut microbiome on metabolism has been examined in a few fish species, such as changes in biosynthesis and metabolism pathways of carbohydrates, amino acids, and lipids as the composition of microbiota changes in carp (Butt and Volkoff, 2019; Ni et al., 2014). In mice, different review studies highlighted the potential link between microbiota and specific host biological processes, including the connection between intestinal microbiota and lipid and lipoprotein metabolism (Yu et al., 2019); or microbiota and IGF1, somatic and bone growth (Poinsot et al., 2018); or microbiota and skeletal muscle mass(Lustgarten, 2019); or microbiota and the immune system (Maynard et al., 2012).

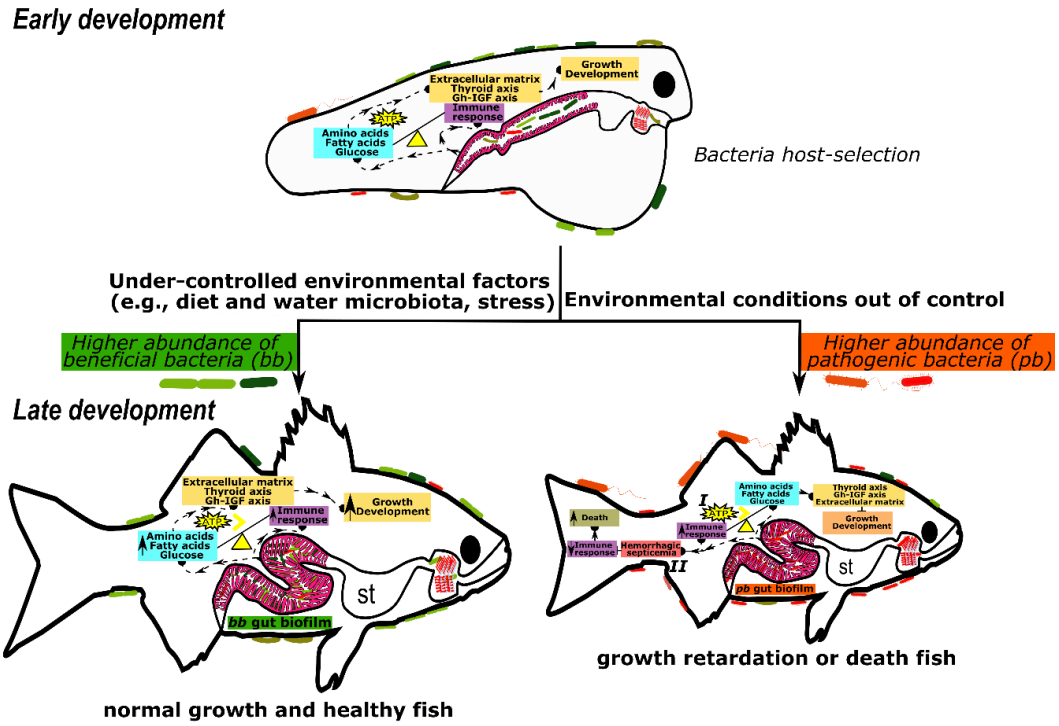


Figure 1.8. A prediction of larval development based on abundant changes in host transcriptome and bacterial community. In early development, there is a host selection of the microbiota and as the larvae develops, the microbiota can change due to external factors (e.g., diet and water) and internal factors (e.g., intestinal folds). An increase in the size of intestinal folds can provide surfaces for bacteria biofilm formation. If beneficial bacteria are established by late development, this can increase food metabolism and the immune response compared to early development due to the fish ontogenetic and developmental status. A healthy microbiome provides resistance against pathogens. As long as the energy balance is towards metabolism and an increase in the immune responses is just a training process, this leads to growth and normal development. Colonization by pathogenic bacteria has a high energy cost for the immune response (e.g., inflammation), limits energy availability for normal development, and causes low larval quality (I). In the acute phase, pathogens can impair the intestinal barrier or infect the bloodstream (septicemia) and other tissues, leading to an impaired immune responses and death (II). Activation or inhibition of genes involved in crucial pathways, including GH-IGF, thyroid axis, energy pathway, and extracellular matrix, may be correlated significantly with microbiota and determine larval quality. Identification of the host biological processes with a significant correlation with microbiota establishment in larval stages can provide a suitable basis for future studies to improve understanding of host-bacteria interactions.

1.8. Summary and future perspectives

This review has highlighted the importance of microbiomes during early larval development and described the main factors that contribute to the establishment of microbiota in aquaculture. Larval development, fish-microbe interactions, feeding, environmental bacteria, and farm location drive the establishment of the microbial community in hatcheries. The effect of each factor can depend on the specific conditions used for larval rearing and this makes studies of the microbiota of aquaculture species difficult to perform and may explain contradictory results between studies. The best alternatives to modulate the microbiota during the hatchery phase are highlighted, such as removing pathogens in live feeds or manipulating the microbiome profiles using selected mixtures of bacteria/probiotics in rearing water/feeds. Nonetheless, complementary data on hatchery site management and abiotic parameters and the developmental stage of larvae are necessary to establish effective strategies.

The change in bacterial composition and function during larval development adds complexity to the design of strategies to exploit the potential beneficial properties of the microbiome in larvae. Colonization of the gut with unstable and dynamic microbiota in early life stages suggests that microbiota modulation should start before a stable microbiota is established. Although host selection and host-bacteria interactions probably play a paramount role in microbiota establishment, the contribution of these interactions to drive specific bacterial communities is less known. Therefore, the host's biological capacity needs further investigation at the molecular level to establish basic physiology. Since transcriptional responses in fish are stage-specific, age can be a critical factor for evaluating the actual biological capacity of larvae.

Transcriptional responses and microbiota profiles are modified during larval ontogeny and parallel modifications in the transcriptome and microbiome suggest an interaction between microbiota and some developmental processes, including pathways involved in muscle development, thyroid and GH-IGF1 axes, energy homeostasis, gastrointestinal maturation, immune response, and tissue homeostasis. The transcriptome or microbiome of larvae at different developmental stages is distinct, suggesting a further adjustment of rearing practices with larval development. This review of transcriptome and microbiome can centralize and direct future studies to enhance larvae development and their interaction with microbiota. That may explain the potential interactions between fish microbiota and host immune response or larvae response to immunization by vaccines during larval ontogeny and introduce alternatives to boost immune response in larvae.

1.9. Future studies

Although NGS approaches have increased our knowledge about transcriptome and microbiome modifications, future studies are needed to obtain more in-depth results for a greater diversity of aquaculture species and to identify the mechanisms that govern fish and microbiota interactions under hatchery conditions. Most small-scale lab-based experiments on fish are under controlled lab conditions, and so their results may not be representative of what occurs under large scale hatchery production. Therefore, both lab- and hatchery-based studies are essential to resolve aquaculture bottlenecks in the future. Increasing the number of hatchery samples and applying clustering analyses is a prerequisite for reducing the bias in transcriptome and microbiome data, especially considering the multitude of hatcheries that exist and their differing management regimes and locations.

Identification of bacteria at the species level using 16S rRNA gene sequencing has limitations due to the high similarity of species within and between some bacterial genus. Sequencing longer regions of the 16S rRNA genes can overcome this limitation and the inclusion of a nuclear gene can provide more robust resolution of the taxonomy if necessary (Fuks et al., 2018). Currently, the main criteria of most 16S rRNA gene primers are to be broad scope and amplify regions that are readily sequenced by current platforms (e.g., V3-V4 primers for Illumina paired-end sequencing). Higher resolution approaches that identify bacteria are important for additional analysis, especially when follow-up studies require isolation and monoculture experiments (Fuks et al., 2018). From the pathogenicity perspective, higher resolution of sequencing to identify bacteria at the level of the species is essential since knowledge about the relative abundance of beneficial and pathogenic bacteria in the microbiome for some bacterial genera (e.g., *Vibrio*) is essential. Another point is that universal primers used in the 16S rRNA gene may not always capture accurate proportions of some genera or species. Therefore, depending on the study objectives, complimentary qPCR analysis can be used to track the modifications of specific groups of bacteria (Najafpour et al., 2022). Since pathogenic bacteria are part of the common bacterial community in aquaculture sites, one solution to reduce their risk is to include beneficial bacteria in food and water or to select environmental conditions that favour their growth. Characterization of antagonistic interactions between bacteria may lead to the detection of new probiotics that should be considered in future studies.

The interactions between microbiota and fish can change during larval development. However, the proportion of active microbiota and their metabolites that interact

with fish or the potential response of fish to microbiota and how they change during larval development is still unclear. The integrated study of metagenomics, meta-transcriptome, and host transcriptomes coupled with visualizing techniques (e.g., imaging mass spectrometry) can improve understanding of the interaction between host and microbiota. It is worth clarifying the proportion of each microbiota driving factor to establish specific bacterial genera. The *Vibrio* genus can be a typical example that colonizes larvae in high proportion and may form a biofilm with larval development. However, it is unclear if just an increase in surface attachment with larval development (e.g., intestinal folds) leads to *Vibrio* biofilm formation and establishment or if the antagonist interactions of this genus with other bacteria play a more substantial role. Determination of modifications in the larvae transcriptome and microbiome during larval development can specify targets for future research to promote larval performance and their interaction with the microbiota. Developing research on fish germ-free models has been less considered, which could be essential to understand host-microbiota interactions. For example, the study of larval immunocompetency using germ-free models may explain the increased immune response with age and larvae-juvenile transition and test whether the increasing pattern is due to less developed lymphoid organs or bacterial load at earlier stages.

1.10. See ANNEX I for Supplementary materials

Supplementary table 1.1. Retrieved scientific publication records related to fish RNA-seq from PubMed.

1. 11. References

Abdelrahman, H., ElHady, M., Alcivar-Warren, A., Allen, S., Al-Tobasei, R., Bao, L., Beck, B., Blackburn, et al., 2017. Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. BMC Genomics 18. https://doi.org/10.1186/S12864-017-3557-1

Abdul Razak, S., Griffin, M.J., Mischke, C.C., Bosworth, B.G., Waldbieser, G.C., Wise, D.J., Marsh, T.L., Scribner, K.T., 2019. Biotic and abiotic factors influencing channel catfish egg and gut microbiome dynamics during early life stages. Aquaculture 498, 556–567. https://doi.org/10.1016/j.aquaculture.2018.08.073

ABELLI, L., PICCHIETTI, S., ROMANO, N., MASTROLIA, L., SCAPIGLIATI, G., 1997. Immunohistochemistry of gut-associated lymphoid tissue of the sea bass *Dicentrarchus labrax* (L.). Fish Shellfish Immunol. 7, 235–245. https://doi.org/10.1006/FSIM.1996.0079

Adams, A., 2019. Progress, challenges and opportunities in fish vaccine development. Fish Shellfish Immunol. 90, 210–214. https://doi.org/10.1016/J.FSI.2019.04.066

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., Narasimhan, G., 2016. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. Evol. Bioinform. 12, 5. https://doi.org/10.4137/EBO.S36436

Alonso, S., Carmen Castro, M., Berdasco, M., de la Banda, I.G., Moreno-Ventas, X., de Rojas, A.H., 2019. Isolation and partial characterization of lactic acid bacteria from the gut microbiota of marine fishes for potential pplication as probiotics in aquaculture. Probiotics Antimicrob. Proteins 11, 569–579. https://doi.org/10.1007/S12602-018-9439-2

Alves, R.N., Gomes, A.S., Stueber, K., Tine, M., Thorne, M.A.S., Smáradóttir, H., Reinhard, R., Clark, M.S., Rønnestad, I., Power, D.M., 2016. The transcriptome of metamorphosing flatfish. BMC Genomics 17, 1–24. https://doi.org/10.1186/s12864-016-2699-x

Aly, S.M., Abdel-Galil Ahmed, Y., Abdel-Aziz Ghareeb, A., Mohamed, M.F., 2008. Studies on *Bacillus subtilis* and *Lactobacillus acidophilus*, as potential probiotics, on the immune response and resistance of Tilapia nilotica *(Oreochromis niloticus*) to challenge infections. Fish Shellfish Immunol. 25, 128–136. https://doi.org/10.1016/J.FSI.2008.03.013

Anamika, K., Verma, S., Jere, A., Desai, A., 2016. Transcriptomic profiling using Next Generation Sequencing - advances, advantages, and challenges. Next Gener. Seq. - Adv. Appl. Challenges. https://doi.org/10.5772/61789

Apún-Molina, J.P., Santamaría- Miranda, A., Luna-González, A., Martínez-Díaz, S.F., Rojas-Contreras, M., 2009. Effect of potential probiotic bacteria on growth and survival of tilapia *Oreochromis niloticus* L., cultured in the laboratory under high density and suboptimum temperature. Aquac. Res. 40, 887–894. https://doi.org/10.1111/J.1365-2109.2009.02172.X

Avella, M.A., Olivotto, I., Silvi, S., Place, A.R., Carnevali, O., 2010. Effect of dietary probiotics on clownfish: A molecular approach to define how lactic acid bacteria modulate development in a marine fish. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 298. https://doi.org/10.1152/ajpregu.00300.2009

Bakke, I., Coward, E., Andersen, T., Vadstein, O., 2015. Selection in the host structures the microbiota associated with developing cod larvae (*Gadus morhua*). Environ. Microbiol.

17, 3914–3924. https://doi.org/10.1111/1462-2920.12888

Bakke, I., Skjermo, J., Vo, T.A., Vadstein, O., 2013. Live feed is not a major determinant of the microbiota associated with cod larvae (*Gadus morhua*). Environ. Microbiol. Rep. 5, 537–548. https://doi.org/10.1111/1758-2229.12042

Balcázar, J.L., Blas, I. de, Ruiz-Zarzuela, I., Cunningham, D., Vendrell, D., Múzquiz, J.L., 2006a. The role of probiotics in aquaculture. Vet. Microbiol. 114, 173–186. https://doi.org/10.1016/J.VETMIC.2006.01.009

Balcázar, J.L., Vendrell, D., de Blas, I., Ruiz-Zarzuela, I., Gironés, O., Muzquiz, J.L., 2006b. Immune modulation by probiotic strains: quantification of phagocytosis of *Aeromonas salmonicida* by leukocytes isolated from gut of rainbow trout (*Oncorhynchus mykiss*) using a radiolabelling assay. Comp. Immunol. Microbiol. Infect. Dis. 29, 335–343. https://doi.org/10.1016/J.CIMID.2006.09.004

Bashiardes, S., Zilberman-Schapira, G., Elinav, E., 2016. Use of metatranscriptomics in microbiome research. bioinform. Biol. Insights 10, 19–25. https://doi.org/10.4137/BBI.S34610

Beiko, R.G., Hsiao, W., Parkinson, J. (Eds.), 2018. Microbiome analysis, methods in molecular niology. Springer, New York. https://doi.org/10.1007/978-1-4939-8728-3

Bergh, Ø., Hansen, G.H., Taxt, R.E., 1992. Experimental infection of eggs and yolk sac larvae of halibut, *Hippoglossus hippoglossus* L. J. Fish Dis. 15, 379–391. https://doi.org/10.1111/J.1365-2761.1992.TB01237.X

Bledsoe, J.W., Peterson, B.C., Swanson, K.S., Small, B.C., 2016. ontogenetic characterization of the intestinal microbiota of channel catfish through 16s rrna gene sequencing reveals insights on temporal shifts and the influence of environmental microbes. PLoS One 11. https://doi.org/10.1371/JOURNAL.PONE.0166379

Bondad-Reantaso, M.G., Subasinghe, R.P., Arthur, J.R., Ogawa, K., Chinabut, S., Adlard, R., Tan, Z., Shariff, M., 2005. Disease and health management in Asian aquaculture. Vet. Parasitol. 132, 249–272. https://doi.org/10.1016/J.VETPAR.2005.07.005

Bone, A., Bekaert, M., Papadopoulou, A., McMillan, S., Adams, A., Davie, A., Desbois, A.P., 2020. Bacterial communities of ballan wrasse (*Labrus bergylta*) eggs at a commercial marine hatchery. Curr. Microbiol. 1, 3. https://doi.org/10.1007/s00284-020-02286-8

Borges, N., Keller-Costa, T., Sanches-Fernandes, G.M.M., Louvado, A., Gomes, N.C.M., Costa, R., 2021. Bacteriome structure, function, and probiotics in fish larviculture: the good, the bad, and the gaps. Annu. Rev. Anim. Biosci. 9. https://doi.org/10.1146/annurev-animal-062920-113114

Bosco, N., Noti, M., 2021. The aging gut microbiome and its impact on host immunity. Genes Immun. 2021 225 22, 289–303. https://doi.org/10.1038/s41435-021-00126-8

Brown, K., DeCoffe, D., Molcan, E., Gibson, D.L., 2012. Diet-induced dysbiosis of the intestinal microbiota and the effects on immunity and disease. Nutr. 2012, Vol. 4, Pages 1095-1119 4, 1095–1119. https://doi.org/10.3390/NU4081095

Brown, R.M., Wiens, G.D., Salinas, I., 2019. Analysis of the gut and gill microbiome of resistant and susceptible lines of rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol. 86, 497–506. https://doi.org/10.1016/J.FSI.2018.11.079

Brugman, S., Ikeda-Ohtsubo, W., Braber, S., Folkerts, G., Pieterse, C.M.J., Bakker, P.A.H.M., 2018. A comparative review on microbiota manipulation: lessons from fish, plants, livestock, and human research. Front. Nutr. 5, 80. https://doi.org/10.3389/FNUT.2018.00080/BIBTEX

Burtseva, O., Kublanovskaya, A., Fedorenko, T., Lobakova, E., Chekanov, K., 2021. Gut microbiome of the White Sea fish revealed by 16S rRNA metabarcoding. Aquaculture 533, 736175. https://doi.org/10.1016/J.AQUACULTURE.2020.736175

Butt, R.L., Volkoff, H., 2019. Gut microbiota and energy homeostasis in fish. Front. Endocrinol. 10, 9. https://doi.org/10.3389/FENDO.2019.00009

Carda-Diéguez, M., Ghai, R., Rodríguez-Valera, F., Amaro, C., 2017. Wild eel microbiome reveals that skin mucus of fish could be a natural niche for aquatic mucosal pathogen evolution. Microbiome 5, 162. https://doi.org/10.1186/S40168-017-0376-1/FIGURES/7

Caruana, J.C., Walper, S.A., 2020. Bacterial membrane vesicles as mediators of microbe – microbe and microbe – host community interactions. Front. Microbiol. 11, 432. https://doi.org/10.3389/FMICB.2020.00432/BIBTEX

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods 69, 330. https://doi.org/10.1016/J.MIMET.2007.02.005

Chan, T.F., Ji, K.M., Yim, A.K.Y., Liu, X.Y., Zhou, J.W., Li, R.Q., Yang, K.Y., Li, J., Li, M., Law, P.T.W., Wu, Y.L., Cai, Z.L., Qin, H., Bao, Y., Leung, R.K.K., Ng, P.K.S., Zou, J., Zhong, X.J., Ran, P.X., Zhong, N.S., Liu, Z.G., Tsui, S.K.W., 2015. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. J. Allergy Clin. Immunol. 135, 539–548. https://doi.org/10.1016/J.JACI.2014.09.031

Crane, M., Hyatt, A., 2011. Viruses of Fish: An overview of significant pathogens. Viruses 3, 2025. https://doi.org/10.3390/V3112025

Cruz, P.M., Ibáñez, A.L., Hermosillo, O.A.M., Saad, H.C.R., 2012. Use of probiotics in aquaculture. ISRN Microbiol. 2012, 1–13. https://doi.org/10.5402/2012/916845

Dabrowski, K.R., 1986. Ontogenetical aspects of nutritional requirements in fish. Comp. Biochem. Physiol. A. Comp. Physiol. 85, 639–655. https://doi.org/10.1016/0300-9629(86)90272-0

Dai, Z., Sevillano-Rivera, M.C., Calus, S.T., Bautista-De Los Santos, Q.M., Eren, A.M., Van Der Wielen, P.W.J.J., Ijaz, U.Z., Pinto, A.J., 2020. Disinfection exhibits systematic impacts on the drinking water microbiome. Microbiome 8, 1–19. https://doi.org/10.1186/S40168-020-00813-0/FIGURES/6

Dang, M., Cao, T., Vasquez, I., Hossain, A., Gnanagobal, H., Kumar, S., Hall, J.R., Monk, J., Boyce, D., Westcott, J., Santander, J., 2021. Oral immunization of larvae and juvenile of lumpfish (*Cyclopterus lumpus*) against *Vibrio anguillarum* does not influence systemic immunity. Vaccines 9. https://doi.org/10.3390/VACCINES9080819/S1

Darias, M.J., Zambonino-Infante, J.L., Hugot, K., Cahu, C.L., Mazurais, D., 2008. Gene expression patterns during the larval development of European sea bass (*dicentrarchus labrax*) by microarray analysis. Mar. Biotechnol. (NY). 10, 416–428.

https://doi.org/10.1007/S10126-007-9078-1

DeBofsky, A., Xie, Y., Challis, J.K., Ankley, P.J., Brinkmann, M., Jones, P.D., Giesy, J.P., 2022. 16S rRNA metabarcoding unearths responses of rare gut microbiome of fathead minnows exposed to benzo[a]pyrene. Sci. Total Environ. 807, 151060. https://doi.org/10.1016/J.SCITOTENV.2021.151060

Dehler, C.E., Secombes, C.J., Martin, S.A.M., 2017. Environmental and physiological factors shape the gut microbiota of Atlantic salmon parr (*Salmo salar* L.). Aquaculture 467, 149–157. https://doi.org/10.1016/J.AQUACULTURE.2016.07.017

Delphino, M.K.V.C., Leal, C.A.G., Gardner, I.A., Assis, G.B.N., Roriz, G.D., Ferreira, F., Figueiredo, H.C.P., Gonçalves, V.S.P., 2019. Seasonal dynamics of bacterial pathogens of Nile tilapia farmed in a Brazilian reservoir. Aquaculture 498, 100–108. https://doi.org/10.1016/J.AQUACULTURE.2018.08.023

Deng, Y., Verdegem, M.C.J., Eding, E., Kokou, F., 2022. Effect of rearing systems and dietary probiotic supplementation on the growth and gut microbiota of Nile tilapia (*Oreochromis niloticus*) larvae. Aquaculture 546, 737297. https://doi.org/10.1016/J.AQUACULTURE.2021.737297

Desai, A.R., Links, M.G., Collins, S.A., Mansfield, G.S., Drew, M.D., Van Kessel, A.G., Hill, J.E., 2012. Effects of plant-based diets on the distal gut microbiome of rainbow trout (*Oncorhynchus mykiss*). Aquaculture 350–353, 134–142. https://doi.org/10.1016/J.AQUACULTURE.2012.04.005

Dimitroglou, A., Merrifield, D.L., Carnevali, O., Picchietti, S., Avella, M., Daniels, C., Güroy, D., Davies, S.J., 2011. Microbial manipulations to improve fish health and production - A Mediterranean perspective. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2010.08.009

Direkbusarakom, S., Yoshimizu, M., Ezura, Y., Ruangpan, L., Danayadol, Y., 1998. *Vibrio* spp., the dominant flora in shrimp hatchery against some fish pathogenic viruses. J. Mar. Biotechnol. 6, 266–267.

Douglas, S.E., Knickle, L.C., Williams, J., Flight, R.M., Reith, M.E., 2008. A first generation Atlantic halibut *Hippoglossus hippoglossus* (L.) microarray: application to developmental studies. J. Fish Biol. 72, 2391–2406. https://doi.org/10.1111/J.1095-8649.2008.01861.X

Dulski, T., Kujawa, R., Godzieba, M., Ciesielski, S., 2020. Effect of salinity on the gut microbiome of pike fry (*Esox lucius*). Appl. Sci. 10, 2506. https://doi.org/10.3390/APP10072506

Fadiji, A.E., Babalola, O.O., 2020. Metagenomics methods for the study of plant-associated microbial communities: A review. J. Microbiol. Methods 170, 105860. https://doi.org/10.1016/J.MIMET.2020.105860

Falk-Petersen, I.B., 2005. Comparative organ differentiation during early life stages of marine fish. Fish Shellfish Immunol. 19, 397–412. https://doi.org/10.1016/J.FSI.2005.03.006

Fantini, D., 2019. easyPubMed: search and retrieve scientific publication records from PubMed. R package version 2.13.

FAO, 2022. The state of world fisheries and aquaculture. Towards Blue Transformation. Rome, FAO. https://doi.org/10.4060/cc0461en

FAO, 2020. The state of world fisheries and aquaculture. https://doi.org/10.4060/CA9231EN

Ferraresso, S., Bonaldo, A., Parma, L., Buonocore, F., Scapigliati, G., Gatta, P.P., Bargelloni, L., 2016. Ontogenetic onset of immune-relevant genes in the common sole (*Solea solea*). Fish Shellfish Immunol. 57, 278–292. https://doi.org/10.1016/J.FSI.2016.08.044

Ferreira, R.B.R., Antunes, L.C.M., Brett Finlay, B., 2010. Should the human microbiome be considered when developing vaccines? PLOS Pathog. 6, e1001190. https://doi.org/10.1371/JOURNAL.PPAT.1001190

Ferretti, P., Farina, S., Cristofolini, M., Girolomoni, G., Tett, A., Segata, N., 2017. Experimental metagenomics and ribosomal profiling of the human skin microbiome. Exp. Dermatol. 26, 211–219. https://doi.org/10.1111/EXD.13210

Fossmark, R.O., Vadstein, O., Rosten, T.W., Bakke, I., Košeto, D., Bugten, A. V., Helberg, G.A., Nesje, J., Jørgensen, N.O.G., Raspati, G., Azrague, K., Østerhus, S.W., Attramadal, K.J.K., 2020. Effects of reduced organic matter loading through membrane filtration on the microbial community dynamics in recirculating aquaculture systems (RAS) with Atlantic salmon parr (*Salmo salar*). Aquaculture 524, 735268. https://doi.org/10.1016/J.AQUACULTURE.2020.735268

Froehlich, H.E., Runge, C.A., Gentry, R.R., Gaines, S.D., Halpern, B.S., 2018. Comparative terrestrial feed and land use of an aquaculture-dominant world. Proc. Natl. Acad. Sci. 115, 5295–5300. https://doi.org/10.1073/PNAS.1801692115

Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P.J., Soen, Y., Shental, N., 2018. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. Microbiome 6, 1–13. https://doi.org/10.1186/S40168-017-0396-X/FIGURES/5

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J.M., Topping, D.L., Suzuki, T., Taylor, T.D., Itoh, K., Kikuchi, J., Morita, H., Hattori, M., Ohno, H., 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nat. 2011 4697331 469, 543–547. https://doi.org/10.1038/nature09646

Ghanbari, M., Kneifel, W., Domig, K.J., 2015. A new view of the fish gut microbiome: Advances from next-generation sequencing. Aquaculture. https://doi.org/10.1016/j.aquaculture.2015.06.033

Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S. V., Knight, R., 2018. Current understanding of the human microbiome. Nat. Med. 24, 392. https://doi.org/10.1038/NM.4517

Gosalbes, M.J., Durbán, A., Pignatelli, M., Abellan, J.J., Jiménez-Hernández, N., Pérez-Cobas, A.E., Latorre, A., Moya, A., 2011. Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One 6, e17447. https://doi.org/10.1371/JOURNAL.PONE.0017447

Grigorakis, K., 2010. Ethical issues in aquaculture production. J. Agric. Environ. Ethics 23, 345–370. https://doi.org/10.1007/S10806-009-9210-5

Gudding, R., Van Muiswinkel, W.B., 2013. A history of fish vaccination: Science-based disease prevention in aquaculture. Fish Shellfish Immunol. 35, 1683–1688. https://doi.org/10.1016/J.FSI.2013.09.031

Guéneau, V., Plateau-Gonthier, J., Arnaud, L., Piard, J.C., Castex, M., Briandet, R., 2022. Positive biofilms to guide surface microbial ecology in livestock buildings. Biofilm 4, 100075. https://doi.org/10.1016/J.BIOFLM.2022.100075

Hansen, G.H., Olafsen, J.A., 1999. Bacterial interactions in early life stages of marine cold water fish. Microb. Ecol. https://doi.org/10.1007/s002489900158

Hassenrück, C., Reinwald, H., Kunzmann, A., Tiedemann, I., Gärdes, A., 2020. Effects of thermal stress on the gut microbiome of juvenile milkfish (*Chanos chanos*). Microorg. 2021, Vol. 9, Page 5 9, 5. https://doi.org/10.3390/MICROORGANISMS9010005

He, M., Li, X., Poolsawat, L., Guo, Z., Yao, W., Zhang, C., Leng, X., 2020. Effects of fish meal replaced by fermented soybean meal on growth performance, intestinal histology and microbiota of largemouth bass (*Micropterus salmoides*). Aquac. Nutr. 26, 1058–1071. https://doi.org/10.1111/ANU.13064

Hess, S., Wenger, A.S., Ainsworth, T.D., Rummer, J.L., 2015. Exposure of clownfish larvae to suspended sediment levels found on the Great Barrier Reef: Impacts on gill structure and microbiome. Sci. Reports 2015 51 5, 1–8. https://doi.org/10.1038/srep10561

Hilerio-Ruiz, A., Toledo-Solís, F.J., Peña, E., Martínez-Burguete, T., Martínez-García, R., Llera-Herrera, R., Álvarez-González, C.A., Saenz de Rodrigáñez, M., 2021. Nutrigenomic markers identified by de novo RNAseq during the early ontogeny of the three spot cichlid *Amphilophus trimaculatus*. Aquaculture 530, 735654. https://doi.org/10.1016/J.AQUACULTURE.2020.735654

Hilton, S.K., Castro-Nallar, E., Pérez-Losada, M., Toma, I., McCaffrey, T.A., Hoffman, E.P., Siegel, M.O., Simon, G.L., Johnson, W.E., Crandall, K.A., 2016. Metataxonomic and metagenomic approaches vs. culture-based techniques for clinical pathology. Front. Microbiol. 7, 484. https://doi.org/10.3389/FMICB.2016.00484/BIBTEX

Hugenholtz, P., Tyson, G.W., 2008. Metagenomics. Nat. 2008 4557212 455, 481–483. https://doi.org/10.1038/455481a

Infante-Villamil, S., Huerlimann, R., Jerry, D.R., 2021. Microbiome diversity and dysbiosis in aquaculture. Rev. Aquac. 13, 1077–1096. https://doi.org/10.1111/RAQ.12513

Ingerslev, H.C., von Gersdorff Jørgensen, L., Lenz Strube, M., Larsen, N., Dalsgaard, I., Boye, M., Madsen, L., 2014. The development of the gut microbiota in rainbow trout (*Oncorhynchus mykiss*) is affected by first feeding and diet type. Aquaculture 424–425, 24–34. https://doi.org/10.1016/J.AQUACULTURE.2013.12.032

Itchell, M., Einstein, R.W., Argaret, M., Itt, L., Ertesz, A.A.K., Hyllis, P., Yper, W., Avid, D., Ose, R., Ark, M., Oulter, C., Llison M C, A., Eer, G., Ichard, R., Acklam, F., Arola, C., Stach, O., Illey, A.M.W., Orczyk, A.L.B., Onald, D., Ow, E.L., 1997. Invasive infections due to a fish pathogen, *Streptococcus iniae*. N Engl J Med. 337, 589–594. https://doi.org/10.1056/NEJM199708283370902

Iwatsuki, T., Kanazawa, T., Ogasawara, T., Hosotani, K., Tsuchiya, K., Watanabe, S., Suzuki, T., Moriuchi, R., Kanesaki, Y., Dohra, H., 2021. 16S rRNA gene amplicon sequencing of gut microbiota in three species of deep-sea fish in suruga bay, Japan. Microbiol. Resour. Announc. 10. https://doi.org/10.1128/MRA.01260-20/ASSET/54DE02AD-0575-4F37-8AF5-82A730228D3C/ASSETS/GRAPHIC/MRA.01260-20-F0001.JPEG

Izquierdo, M.S., Socorro, J., Arantzamendi, L., Hernández-Cruz, C.M., 2000. Recent advances in lipid nutrition in fish larvae. Fish Physiol. Biochem. 22, 97–107. https://doi.org/10.1023/A:1007810506259/METRICS

Jahangiri, L., Esteban, M.Á., 2018. Administration of probiotics in the water in finfish aquaculture systems: a review. Fishes 3, 33 . https://doi.org/10.3390/FISHES3030033

Ji, L., Chen, S., Gu, G., Zhou, J., Wang, W., Ren, J., Wu, J., Yang, D., Zheng, Y., 2021. Exploration of crucial mediators for carotid atherosclerosis pathogenesis through integration of microbiome, metabolome, and transcriptome. Front. Physiol. 12, 615. https://doi.org/10.3389/FPHYS.2021.645212/BIBTEX

Johny, T.K., Puthusseri, R.M., Bhat, S.G., 2021. A primer on metagenomics and next-generation sequencing in fish gut microbiome research. Aquac. Res. 52, 4574–4600. https://doi.org/10.1111/ARE.15373

Kamada, N., Chen, G.Y., Inohara, N., Núñez, G., 2013. Control of pathogens and pathobionts by the gut microbiota. Nat. Immunol. 14, 685. https://doi.org/10.1038/NI.2608

Kamler, E., 1992. Endogenous feeding period. Early Life Hist. Fish 107–175. https://doi.org/10.1007/978-94-011-2324-2_4

Kokou, F., Sasson, G., Nitzan, T., Doron-Faigenboim, A., Harpaz, S., Cnaani, A., Mizrahi, I., 2018. Host genetic selection for cold tolerance shapes microbiome composition and modulates its response to temperature. Elife 7. https://doi.org/10.7554/ELIFE.36398

Kooij, D., Wielen, P.V.D., 2013. Microbial growth in drinking-water supplies: problems, causes, control and research needs. undefined 12. https://doi.org/10.2166/9781780400419

Korsnes, Kjetil, Nicolaisen, O., Skår, Cecilie K, Nerland, Audun H, Bergh Korsnes, Ø., Korsnes, K, Skår, C K, Nerland, A H, Bergh, Ø., 2006. Bacteria in the gut of juvenile cod *Gadus morhua* fed live feed enriched with four different commercial diets. ICES J. Mar. Sci. 63, 296–301. https://doi.org/10.1016/J.ICESJMS.2005.10.012

Krotman, Y., Yergaliyev, T.M., Alexander Shani, R., Avrahami, Y., Szitenberg, A., 2020. Dissecting the factors shaping fish skin microbiomes in a heterogeneous inland water system. Microbiome 8, 1–15. https://doi.org/10.1186/S40168-020-0784-5/FIGURES/7

Kumar Awasthi, M., Ravindran, B., Sarsaiya, S., Chen, H., Wainaina, S., Singh, E., Liu, T., Kumar, S., Pandey, A., Singh, L., Zhang, Z., 2020. Metagenomics for taxonomy profiling: tools and approaches. https://doi.org/10.1080/21655979.2020.1736238 11, 356–374. https://doi.org/10.1080/21655979.2020.1736238

Kumbhare, S. V., Patangia, D. V., Patil, R.H., Shouche, Y.S., Patil, N.P., 2019. Factors influencing the gut microbiome in children: from infancy to childhood. J. Biosci. 44, 1–19. https://doi.org/10.1007/S12038-019-9860-Z/FIGURES/1

Lam, S.H., Chua, H.L., Gong, Z., Lam, T.J., Sin, Y.M., 2004. Development and maturation of the immune system in zebrafish, *Danio rerio*: A gene expression profiling, in situ hybridization and immunological study. Dev. Comp. Immunol. 28, 9–28. https://doi.org/10.1016/S0145-305X(03)00103-4

Larsen, A.M., Mohammed, H.H., Arias, C.R., 2014. Characterization of the gut microbiota of three commercially valuable warmwater fish species. J. Appl. Microbiol. 116, 1396–1404. https://doi.org/10.1111/JAM.12475

Lau, S.K.P., Teng, J.L.L., Ho, C.C., Woo, P.C.Y., 2015. Gene amplification and sequencing for bacterial identification. Methods Microbiol. 42, 433–464. https://doi.org/10.1016/BS.MIM.2015.04.003

Lauzon, H.L., Pérez-Sánchez, T., Merrifield, D.L., Ringø, E., Balcázar, J.L., 2014. Probiotic applications in cold water fish species. Aquac. Nutr. Gut Heal. Probiotics Prebiotics 223–252. https://doi.org/10.1002/9781118897263.CH9

Luyer, J., Schull, Q., Auffret, P., Lopez, P., Crusot, M., Belliard, C., Basset, C., Carradec, Q., Poulain, J., Planes, S., Saulnier, D., 2021. Dual RNAseq highlights the kinetics of skin microbiome and fish host responsiveness to bacterial infection. Anim. Microbiome 3. https://doi.org/10.1186/s42523-021-00097-1

Leclercq, E., Taylor, J.F., Migaud, H., 2010. Morphological skin colour changes in teleosts. Fish Fish. 11, 159–193. https://doi.org/10.1111/J.1467-2979.2009.00346.X

Lee, P.T., Yamamoto, F.Y., Low, C.F., Loh, J.Y., Chong, C.M., 2021. Gut immune system and the implications of oral-administered immunoprophylaxis in finfish aquaculture. Front. Immunol. 12. https://doi.org/10.3389/FIMMU.2021.773193

Li, X.F., Mitch, W.A., 2018. Drinking water disinfection byproducts (dbps) and human health effects: multidisciplinary challenges and opportunities. Environ. Sci. Technol. 52, 1681–1689. https://doi.org/10.1021/ACS.EST.7B05440/ASSET/IMAGES/LARGE/ES-2017-05440A_0002.JPEG

Liu, W., McNeilly, T.N., Mitchell, M., Burgess, S.T.G., Nisbet, A.J., Matthews, J.B., Babayan, S.A., 2022. Vaccine-induced time- and age-dependent mucosal immunity to gastrointestinal parasite infection. npj Vaccines 2022 71 7, 1–9. https://doi.org/10.1038/s41541-022-00501-0

Llewellyn, M.S., Boutin, S., Hoseinifar, S.H., Derome, N., 2014. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front. Microbiol. 5, 1–1. https://doi.org/10.3389/FMICB.2014.00207

Llorente, I., Fernández-Polanco, J., Baraibar-Diez, E., Odriozola, M.D., Bjørndal, T., Asche, F., Guillen, J., Avdelas, L., Nielsen, R., Cozzolino, M., Luna, M., Fernández-Sánchez, J.L., 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. https://doi.org/10.1016/J.MARPOL.2020.103876

Lobo, C., Moreno-Ventas, X., Tapia-Paniagua, S., Rodríguez, C., Moriñigo, M.A., de La Banda, I.G., 2014. Dietary probiotic supplementation (*Shewanella putrefaciens* Pdp11) modulates gut microbiota and promotes growth and condition in Senegalese sole larviculture. Fish Physiol. Biochem. 40. https://doi.org/10.1007/s10695-013-9844-0

Lokman, P.M., Symonds, J.E., 2014. Molecular and biochemical tricks of the research trade: -omics approaches in finfish aquaculture. N. Z. J. Mar. Freshw. Res. 48, 492–505. https://doi.org/10.1080/00288330.2014.928333

Lustgarten, M.S., 2019. The role of the gut microbiome on skeletal muscle mass and physical function: 2019 update. Front. Physiol. 10, 1435. https://doi.org/10.3389/FPHYS.2019.01435/BIBTEX

Lynn, D.J., Benson, S.C., Lynn, M.A., Pulendran, B., 2021. Modulation of immune responses

to vaccination by the microbiota: implications and potential mechanisms. Nat. Rev. Immunol. 2021 221 22, 33–46. https://doi.org/10.1038/s41577-021-00554-7

Makridis, P., Martins, S., Vercauteren, T., Van Driessche, K., Decamp, O., Dinis, M.T., 2005. Evaluation of candidate probiotic strains for gilthead sea bream larvae (*Sparus aurata*) using an in vivo approach. Lett. Appl. Microbiol. 40, 274–277. https://doi.org/10.1111/J.1472-765X.2005.01676.X

Marizzoni, M., Gurry, T., Provasi, S., Greub, G., Lopizzo, N., Ribaldi, F., Festari, C., Mazzelli, M., Mombelli, E., Salvatore, M., Mirabelli, P., Franzese, M., Soricelli, A., Frisoni, G.B., Cattaneo, A., 2020. Comparison of bioinformatics pipelines and operating systems for the analyses of 16s rrna gene amplicon sequences in human fecal samples. Front. Microbiol. 11, 1262. https://doi.org/10.3389/FMICB.2020.01262/BIBTEX

Martínez, G., Peña, E., Martínez, R., Camarillo, S., Burggren, W., Álvarez, A., 2021. Survival, growth, and development in the early stages of the tropical gar atractosteus tropicus: developmental critical windows and the influence of temperature, salinity, and oxygen availability. Fishes 6, 5. https://doi.org/10.3390/FISHES6010005

Martins, M., Eding, H., Verdegem, J., Heinsbroek, N., Schneider, O., Blancheton, P., Roque, E., Verreth, J., Group, F., Yerseke, N.T., Exp, A., Piscicole, R., Abbreviations, D., 2010. Archimer perspective on environmental sustainability. Aquac. Eng. 43.

Mazurais, D., Darias, M., Zambonino-Infante, J.L., Cahu, C.L., 2011. Transcriptomics for understanding marine fish larval development. Can. J. Zool. 89. https://doi.org/10.1139/z11-036

McCauley, K.E., Flynn, K., Calatroni, A., DiMassa, V., LaMere, B., Fadrosh, D.W., Lynch, K. V., Gill, M.A., Pongracic, J.A., Khurana Hershey, G.K., Kercsmar, C.M., Liu, A.H., Johnson, C.C., Kim, H., Kattan, M., O'Connor, G.T., Bacharier, L.B., Teach, S.J., Gergen, P.J., Wheatley, L.M., Togias, A., LeBeau, P., Presnell, S., Boushey, H.A., Busse, W.W., Gern, J.E., Jackson, D.J., Altman, M.C., Lynch, S. V., 2022. Seasonal airway microbiome and transcriptome interactions promote childhood asthma exacerbations. J. Allergy Clin. Immunol. 150, 204-213. https://doi.org/10.1016/J.JACI.2022.01.020

Merrifield, D.L., Rodiles, A., 2015. The fish microbiome and its interactions with mucosal tissues, in: Mucosal Health in Aquaculture. Elsevier Inc., pp. 273–295. https://doi.org/10.1016/B978-0-12-417186-2.00010-8

Meslier, V., Quinquis, B., Da Silva, K., Plaza Oñate, F., Pons, N., Roume, H., Podar, M., Almeida, M., 2022. Benchmarking second and third-generation sequencing platforms for microbial metagenomics. Sci. Data 91 9, 1–9. https://doi.org/10.1038/s41597-022-01762-z

Minich, J.J., Poore, G.D., Jantawongsri, K., Johnston, C., Bowie, K., Bowman, J., Knight, R., Nowak, B., Allen, E.E., 2020. Microbial ecology of atlantic salmon (*Salmo salar*) hatcheries: Impacts of the built environment on fish mucosal microbiota. Appl. Environ. Microbiol. 86. https://doi.org/10.1128/AEM.00411-20/SUPPL_FILE/AEM.00411-20-SD002.CSV

Miquel-Clopés, A., Bentley, E.G., Stewart, J.P., Carding, S.R., 2019. Mucosal vaccines and technology. Clin. Exp. Immunol. 196, 205. https://doi.org/10.1111/CEI.13285

Mulero, I., Sepulcre, M.P., Fuentes, I., García-Alcázar, A., Meseguer, J., García-Ayala, A., Mulero, V., 2008. Vaccination of larvae of the bony fish gilthead seabream reveals a lack of correlation between lymphocyte development and adaptive immunocompetence. Mol. Immunol. 45, 2981–2989. https://doi.org/10.1016/J.MOLIMM.2008.01.017

Muniesa, A., Basurco, B., Aguilera, C., Furones, D., Reverté, C., Sanjuan-Vilaplana, A., Jansen, M.D., Brun, E., Tavornpanich, S., 2020. Mapping the knowledge of the main diseases affecting sea bass and sea bream in Mediterranean. Transbound. Emerg. Dis. 67, 1089–1100. https://doi.org/10.1111/tbed.13482

Myrold, D.D., Zeglin, L.H., Jansson, J.K., 2014. The potential of metagenomic approaches for understanding soil microbial processes. Soil Sci. Soc. Am. J. 78, 3–10. https://doi.org/10.2136/SSSAJ2013.07.0287DGS

Nafee, T.M., Farrell, W.E., Carroll, W.D., Fryer, A.A., Ismail, K.M.K., 2008. Review article: Epigenetic control of fetal gene expression. BJOG An Int. J. Obstet. Gynaecol. 115, 158–168. https://doi.org/10.1111/J.1471-0528.2007.01528.X

Najafpour, B., Cardoso, J.C.R., Canário, A.V.M., Power, D.M., 2020. Specific evolution and gene family expansion of complement 3 and regulatory factor h in fish. Front. Immunol. 11, 2945. https://doi.org/10.3389/FIMMU.2020.568631

Najafpour, B., Dorafshan, S., Heyrati, F.P., Canario, A.V.M., Power, D.M., 2021a. Comparative ontogeny of the digestive tract of *Oncorhynchus mykiss* ♀ x *Salmo trutta caspius* ♂ triploid hybrids to their parental species. Aquac. Nutr. 27, 427–438. https://doi.org/10.1111/ANU.13196

Najafpour, B., Dorafshan, S., Paykan Heyrati, F., Power, D.M., 2019. Embryonic development of the endangered Caspian brown trout, *Salmo trutta caspius* (Kessler, 1877). J. Appl. Ichthyol. 00, 1–7. https://doi.org/10.1111/jai.13874

Najafpour, B., Pinto, P., Canario, A., Power, D., 2022. Quantifying dominant bacterial genera detected in metagenomic data from fish eggs and larvae using genus-specific primers. Authorea Prepr. https://doi.org/10.22541/AU.164652122.26782894/V1

Najafpour, B., Pinto, P.I.S., Moutou, K.A., Canario, A.V.M., Power, D.M., 2021b. Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream. Microorg. 9, 2275–2275. https://doi.org/10.3390/MICROORGANISMS9112275

Najafpour, B., Pinto, P.I.S., Sanz, E.C., Martinez-Blanch, J.F., Canario, A.V.M., Moutou, K.A., Power, D.M., 2023. Core microbiome profiles and their modification by environmental, biological, and rearing factors in aquaculture hatcheries. Mar. Pollut. Bull. 193, 115218. https://doi.org/10.1016/J.MARPOLBUL.2023.115218

Naya-Català, F., do Vale Pereira, G., Piazzon, M.C., Fernandes, A.M., Calduch-Giner, J.A., Sitjà-Bobadilla, A., Conceição, L.E.C., Pérez-Sánchez, J., 2021. cross-talk between intestinal microbiota and host gene expression in gilthead sea bream (*Sparus aurata*) juveniles: insights in fish feeds for increased circularity and resource utilization. Front. Physiol. 12. https://doi.org/10.3389/FPHYS.2021.748265

Nayak, S. K., 2010. Probiotics and immunity: a fish perspective. Fish Shellfish Immunol. 29, 2–14. https://doi.org/10.1016/J.FSI.2010.02.017

Nayak, S. K., 2010. Role of gastrointestinal microbiota in fish. Aquac. Res. 41, 1553–1573. https://doi.org/10.1111/J.1365-2109.2010.02546.X

Ni, J., Yan, Q., Yu, Y., Zhang, T., 2014. Factors influencing the grass carp gut microbiome and its effect on metabolism. FEMS Microbiol. Ecol. 87, 704–714. https://doi.org/10.1111/1574-6941.12256

Niu, K.M., Lee, B.J., Kothari, D., Lee, W. Do, Hur, S.W., Lim, S.G., Kim, K.W., Kim, K.D., Kim, N.N., Kim, S.K., 2020. Dietary effect of low fish meal aquafeed on gut microbiota in olive flounder (*Paralichthys olivaceus*) at different growth stages. Microbiologyopen 9. https://doi.org/10.1002/MBO3.992

Novoslavskij, A., Terentjeva, M., Eizenberga, I., Valciņa, O., Bartkevičs, V., Bērziņš, A., 2016. Major foodborne pathogens in fish and fish products: a review. Ann. Microbiol. 66, 1–15. https://doi.org/10.1007/S13213-015-1102-5/TABLES/5

Novriadi, R., 2016. Vibriosis in aquaculture. Omni-Akuatika 12. https://doi.org/10.20884/1.OA.2016.12.1.24

Nowinski, B., Smith, C.B., Thomas, C.M., Esson, K., Marin, R., Preston, C.M., Birch, J.M., Scholin, C.A., Huntemann, M., Clum, A., Foster, Brian, Foster, Bryce, Roux, S., Palaniappan, K., Varghese, N., Mukherjee, S., Reddy, T.B.K., Daum, C., Copeland, A., Chen, I.M.A., Ivanova, N.N., Kyrpides, N.C., Glavina del Rio, T., Whitman, W.B., Kiene, R.P., Eloe-Fadrosh, E.A., Moran, M.A., 2019. Microbial metagenomes and metatranscriptomes during a coastal phytoplankton bloom. Sci. Data 2019 61 6, 1–7. https://doi.org/10.1038/s41597-019-0132-4

Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R., Konstantinidis, K.T., 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. Appl. Environ. Microbiol. 77, 6000–6011. https://doi.org/10.1128/AEM.00107-11/SUPPL_FILE/TABLES1_PFAM_COUNTS.XLS

Olafsen, J.A., 2001. Interactions between fish larvae and bacteria in marine aquaculture, in: Aquaculture. Elsevier, pp. 223–247. https://doi.org/10.1016/S0044-8486(01)00702-5

Parata, L., Nielsen, S., Xing, X., Thomas, T., Egan, S., Vergés, A., 2020. Age, gut location and diet impact the gut microbiome of a tropical herbivorous surgeonfish. FEMS Microbiol. Ecol. 96, 179. https://doi.org/10.1093/FEMSEC/FIZ179

Pérez-Sánchez, T., Mora-Sánchez, B., Balcázar, J.L., 2018. Biological approaches for disease control in aquaculture: advantages, limitations and challenges. Trends Microbiol. 26, 896–903. https://doi.org/10.1016/J.TIM.2018.05.002

Pérez, T., Balcázar, J.L., Ruiz-Zarzuela, I., Halaihel, N., Vendrell, D., De Blas, I., Múezquiz, J.L., 2010. Host–microbiota interactions within the fish intestinal ecosystem. Mucosal Immunol. 34 3, 355–360. https://doi.org/10.1038/mi.2010.12

Perissi, I., Bardi, U., El Asmar, T., Lavacchi, A., 2017. Dynamic patterns of overexploitation in fisheries. Ecol. Modell. 359, 285–292. https://doi.org/10.1016/J.ECOLMODEL.2017.06.009

Perry, W.B., Lindsay, E., Payne, C.J., Brodie, C., Kazlauskaite, R., 2020. The role of the gut

microbiome in sustainable teleost aquaculture. Proc. R. Soc. B 287. https://doi.org/10.1098/RSPB.2020.0184

Picchietti, S., Fausto, A.M., Randelli, E., Carnevali, O., Taddei, A.R., Buonocore, F., Scapigliati, G., Abelli, L., 2009. Early treatment with *Lactobacillus delbrueckii* strain induces an increase in intestinal T-cells and granulocytes and modulates immune-related genes of larval *Dicentrarchus labrax* (L.). Fish Shellfish Immunol. 26, 368–376. https://doi.org/10.1016/j.fsi.2008.10.008

Picchietti, S., Guerra, L., Bertoni, F., Randelli, E., Belardinelli, M.C., Buonocore, F., Fausto, A.M., Rombout, J.H., Scapigliati, G., Abelli, L., 2011. Intestinal T cells of *Dicentrarchus labrax* (L.): gene expression and functional studies. Fish Shellfish Immunol. 30, 609–617. https://doi.org/10.1016/J.FSI.2010.12.006

Pittman, K., Yúfera, M., Pavlidis, M., Geffen, A.J., Koven, W., Ribeiro, L., Zambonino-Infante, J.L., Tandler, A., 2013. Fantastically plastic: fish larvae equipped for a new world. Rev. Aquac. 5, S224–S267. https://doi.org/10.1111/RAQ.12034

Poinsot, P., Schwarzer, M., Peretti, N., Leulier, F., 2018. 40 YEARS OF IGF1: The emerging connections between IGF1, the intestinal microbiome, *Lactobacillus* strains and bone growth. J. Mol. Endocrinol. 61, T103–T113. https://doi.org/10.1530/JME-17-0292

Portella, M.C., Jomori, R.K., Leitão, N.J., Menossi, O.C.C., Freitas, T.M., Kojima, J.T., Lopes, T.S., Clavijo-Ayala, J.A., Carneiro, D.J., 2014. Larval development of indigenous South American freshwater fish species, with particular reference to pacu (*Piaractus mesopotamicus*): A review. Aquaculture 432, 402–417. https://doi.org/10.1016/j.aquaculture.2014.04.032

Pradeepkiran, J.A., 2019. Aquaculture role in global food security with nutritional value: a review. Transl. Anim. Sci. 3, 903–910. https://doi.org/10.1093/TAS/TXZ012

Preena, P.G., Swaminathan, T.R., Kumar, V.J.R., Singh, I.S.B., 2020. Antimicrobial resistance in aquaculture: a crisis for concern. Biol. 2020 759 75, 1497–1517. https://doi.org/10.2478/S11756-020-00456-4

Qian, X., Ba, Y., Zhuang, Q., Zhong, G., 2014. RNA-seq technology and its application in fish transcriptomics. Omi. A J. Integr. Biol. https://doi.org/10.1089/omi.2013.0110

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nat. 2010 4647285 464, 59–65. https://doi.org/10.1038/nature08821

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N., 2017. Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. 2017 359 35, 833–844. https://doi.org/10.1038/nbt.3935

Ray, A.K., Ghosh, K., Ringø, E., 2012. Enzyme-producing bacteria isolated from fish gut: a review. Aquac. Nutr. 18, 465–492. https://doi.org/10.1111/J.1365-2095.2012.00943.X

Razak, S.A., Scribner, K.T., 2020. Ecological and ontogenetic components of larval lake sturgeon gut microbiota assembly, successional dynamics, and ecological evaluation of neutral community processes. Appl. Environ. Microbiol. 86. https://doi.org/10.1128/AEM.02662-19/SUPPL_FILE/AEM.02662-19-SD004.XLSX

Restivo, V.E., Kidd, K.A., Surette, M.G., Bucking, C., Wilson, J.Y., 2021. The gut content microbiome of wild-caught rainbow darter is altered during laboratory acclimation. Comp. Biochem. Physiol. Part D Genomics Proteomics 39, 100835. https://doi.org/10.1016/J.CBD.2021.100835

Richardson, S.D., 2003. Disinfection by-products and other emerging contaminants in drinking water. TrAC Trends Anal. Chem. 22, 666–684. https://doi.org/10.1016/S0165-9936(03)01003-3

Rigos, G., Troisi, G.M., 2005. Antibacterial agents in mediterranean finfish farming: a synopsis of drug pharmacokinetics in important euryhaline fish species and possible environmental implications. Rev. Fish Biol. Fish. 2005 151 15, 53–73. https://doi.org/10.1007/S11160-005-7850-8

Rimoldi, S., Antonini, M., Gasco, L., Moroni, F., Terova, G., 2021. Intestinal microbial communities of rainbow trout (*Oncorhynchus mykiss*) may be improved by feeding a Hermetia illucens meal/low-fishmeal diet. Fish Physiol. Biochem. 47. https://doi.org/10.1007/s10695-020-00918-1

Rodrigues, P.M., Schrama, D., Campos, A., Osório, H., Freitas, M., 2016. Applications of proteomics in aquaculture. Agric. Proteomics Vol. 1 Crop. Hortic. Farm Anim. Food, Insect Microorg. 165–199. https://doi.org/10.1007/978-3-319-43275-5_10/COVER

Rolhion, N., Chassaing, B., 2016. When pathogenic bacteria meet the intestinal microbiota. Philos. Trans. R. Soc. B Biol. Sci. 371. https://doi.org/10.1098/RSTB.2015.0504

Rombout, J.H.W.M., Taverne-Thiele, A.J., Villena, M.I., 1993. The gut-associated lymphoid tissue (GALT) of carp (*Cyprinus carpio* L.): an immunocytochemical analysis. Dev. Comp. Immunol. 17, 55–66. https://doi.org/10.1016/0145-305X(93)90015-I

Rombout Jan, J.H.W.M., Abelli, L., Picchietti, S., Scapigliati, G., Kiron, V., 2011. Teleost intestinal immunology. Fish Shellfish Immunol. 31, 616–626. https://doi.org/10.1016/J.FSI.2010.09.001

Romero, J., Navarrete, P., 2006. 16S rDNA-based analysis of dominant bacterial populations associated with early life stages of coho salmon (*Oncorhynchus kisutch*). Microb. Ecol. 51, 422–430. https://doi.org/10.1007/S00248-006-9037-9

Rønnestad, I., Kamisaka, Y., Conceição, L.E.C., Morais, S., Tonheim, S.K., 2007. Digestive physiology of marine fish larvae: Hormonal control and processing capacity for proteins, peptides and amino acids. Aquaculture 268, 82–97. https://doi.org/10.1016/J.AQUACULTURE.2007.04.031

Rønnestad, I., Yúfera, M., Ueberschär, B., Ribeiro, L., Saele, Ø., Boglione, C., 2013. Feeding behaviour and digestive physiology in larval fish: current knowledge, and gaps and bottlenecks in research. Rev. Aquac. 5, S59–S98. https://doi.org/10.1111/raq.12010

Roy, S., Kumar, V., Behera, B.K., Das, B.K., 2021. Epigenetics: perspectives and potential in aquaculture. Adv. Fish. Biotechnol. 133–150. https://doi.org/10.1007/978-981-16-3215-0_9/COVER

Sarropoulou, E., Kotoulas, G., Power, D.M., Geisler, R., 2005. Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. Physiol. Genomics 23, 182–191.

https://doi.org/10.1152/physiolgenomics.00139.2005

Sbordoni, V., De Matthaeis, E., Cobolli Sbordoni, M., La Rosa, G., Mattoccia, M., 1986. Bottleneck effects and the depression of genetic variability in hatchery stocks of Penaeus japonicus (*Crustacea, Decapoda*). Aquaculture 57, 239–251. https://doi.org/10.1016/0044-8486(86)90202-4

Searcy, S.P., Sponaugle, S.U., 2001. Selective mortality during the larval-juvenile transition in two coral reef fishes. Ecology 82, 2452–2470. https://doi.org/10.1890/0012-9658

Sevillano, M., Dai, Z., Calus, S., Santos, Q.M.B. los, Eren, A.M., Wielen, P.W.J.J. van der, Ijaz, U.Z., Pinto, A.J., 2019. Disinfectant residuals in drinking water systems select for mycobacterial populations with intrinsic antimicrobial resistance. bioRxiv 675561. https://doi.org/10.1101/675561

Sifa, L., Mathias, J.A., 1987. The critical period of high mortality of larvae fish —A discussion based on current research. Chinese J. Oceanol. Limnol. 51 5, 80–96. https://doi.org/10.1007/BF02848526

Silphaduang, U., Colorni, A., Noga, E.J., 2006. Evidence for widespread distribution of piscidin antimicrobial peptides in teleost fish. Dis. Aquat. Organ. 72, 241–252. https://doi.org/10.3354/DAO072241

Silva, Y.J., Costa, L., Pereira, C., Mateus, C., Cunha, Â., Calado, R., Gomes, N.C.M., Pardo, M.A., Hernandez, I., Almeida, A., 2014. Phage therapy as an approach to prevent *Vibrio anguillarum* infections in fish larvae production. PLoS One 9. https://doi.org/10.1371/JOURNAL.PONE.0114197

Simon, H.M., Smith, M.W., Herfort, L., 2014. Metagenomic insights into particles and their associated microbiota in a coastal margin ecosystem. Front. Microbiol. 5, 466. https://doi.org/10.3389/FMICB.2014.00466/BIBTEX

Sleator, R.D., Shortall, C., Hill, C., 2008. Metagenomics. Lett. Appl. Microbiol. 47, 361–366. https://doi.org/10.1111/J.1472-765X.2008.02444.X

Smith, P., 2008. Antimicrobial resistance in aquaculture. Rev. Sci. Tech. 27, 243–264. https://doi.org/10.20506/rst.27.1.1799

Song, W., Jiang, K., Zhang, F., Lin, Y., Ma, L., Fernandes, J.M.O., 2015. Transcriptome sequencing, de novo assembly and differential gene expression analysis of the early development of *Acipenser baeri*. PLoS One 10, e0137450. https://doi.org/10.1371/JOURNAL.PONE.0137450

Song, W., Li, L., Huang, H., Jiang, K., Zhang, F., Chen, X., Zhao, M., Ma, L., 2016. The gut microbial community of antarctic fish detected by 16S rRNA gene sequence analysis. Biomed Res. Int. 2016. https://doi.org/10.1155/2016/3241529

Steiner, K., Heasman, K., Laroche, O., Pochon, X., Preece, M., Bowman, J.P., Walker, S.P., Symonds, J.E., 2021. The microbiome of Chinook salmon (*Oncorhynchus tshawytscha*) in a recirculation aquaculture system. Aquaculture 534, 736227. https://doi.org/10.1016/J.AQUACULTURE.2020.736227

Stewart, F.J., 2013. Preparation of microbial community cDNA for metatranscriptomic analysis in marine plankton. Methods Enzymol. 531, 187–218. https://doi.org/10.1016/B978-0-12-407863-5.00010-1

Sudhagar, A., Kumar, G., El-Matbouli, M., 2018. Transcriptome analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: a comprehensive review. Int. J. Mol. Sci. 19. https://doi.org/10.3390/IJMS19010245

Sugita, H., Miyajima, C., Deguchi, Y., 1991. The vitamin B12-producing ability of the intestinal microflora of freshwater fish. Aquaculture 92, 267–276. https://doi.org/10.1016/0044-8486(91)90028-6

Sullam, K.E., Essinger, S.D., Lozupone, C.A., O'Connor, M.P., Rosen, G.L., Knight, R., Kilham, S.S., Russell, J.A., 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: A meta-analysis. Mol. Ecol. 21, 3363–3378. https://doi.org/10.1111/j.1365-294X.2012.05552.x

Sylvain, F.É., Derome, N., 2017. Vertically and horizontally transmitted microbial symbionts shape the gut microbiota ontogenesis of a skin-mucus feeding discus fish progeny. Sci. Reports 71 7, 1–14. https://doi.org/10.1038/s41598-017-05662-w

Tabassum, T., Sofi Uddin Mahamud, A.G.M., Acharjee, T.K., Hassan, R., Akter Snigdha, T., Islam, T., Alam, R., Khoiam, M.U., Akter, F., Azad, M.R., Al Mahamud, M.A., Ahmed, G.U., Rahman, T., 2021. Probiotic supplementations improve growth, water quality, hematology, gut microbiota and intestinal morphology of Nile tilapia. Aquac. Reports 21, 100972. https://doi.org/10.1016/J.AQREP.2021.100972

Talwar, C., Nagar, S., Lal, R., Negi, R.K., 2018. Fish gut microbiome: current approaches and future perspectives. Indian J. Microbiol. 584 58, 397–414. https://doi.org/10.1007/S12088-018-0760-Y

Taneja, V., 2017. Microbiome: impact of gender on function & characteristics of gut microbiome. Princ. Gender-Specific Med. Gend. Genomic Era Third Ed. 569–583. https://doi.org/10.1016/B978-0-12-803506-1.00027-9

Tang, X., Jiang, S., Wang, H., Zhou, Yanfeng, Peng, F., Zhang, X., Zhou, Yifan, Guo, S., You, Y., 2022. Transcriptome sequencing analysis reveals dynamic changes in major biological functions during the early development of clearhead icefish, *Protosalanx chinensis*. Fishes 7, 115. https://doi.org/10.3390/FISHES7030115

Toranzo, A.E., Magariños, B., Romalde, J.L., 2005. A review of the main bacterial fish diseases in mariculture systems. Aquaculture 246, 37–61. https://doi.org/10.1016/J.AQUACULTURE.2005.01.002

Touraki, M., Niopas, I., Karagiannis, V., 2012. Treatment of vibriosis in European sea bass larvae, *Dicentrarchus labrax* L., with oxolinic acid administered by bath or through medicated nauplii of *Artemia franciscana* (Kellogg): efficacy and residual kinetics. J. Fish Dis. 35, 513–522. https://doi.org/10.1111/J.1365-2761.2012.01387.X

Tuson, H.H., Weibel, D.B., 2013. Bacteria-surface interactions. Soft Matter 9, 4368. https://doi.org/10.1039/C3SM27705D

Uzzell, T., Stolzenberg, E.D., Shinnar, A.E., Zasloff, M., 2003. Hagfish intestinal antimicrobial peptides are ancient cathelicidins. Peptides 24, 1655–1667. https://doi.org/10.1016/J.PEPTIDES.2003.08.024

Vadstein, O., Bergh, Ø., Gatesoupe, F.-J., Galindo-Villegas, J., Mulero, V., Picchietti, S., Scapigliati, G., Makridis, P., Olsen, Y., Dierckens, K., Defoirdt, T., Boon, N., De

Schryver, P., Bossier, P., 2013. Microbiology and immunology of fish larvae. Rev. Aquac. 5, S1–S25. https://doi.org/10.1111/j.1753-5131.2012.01082.x

Vadstein, O., Øie, G., Olsen, Y., Salvesen, I., Skjermo, J., Skjåk-Bræk, G., 1993. A strategy to obtain microbial control during larval development of marine fish. Fish Farming Technol. 69–75. https://doi.org/10.1201/9781003077770-13

Valdez, Y., Brown, E.M., Finlay, B.B., 2014. Influence of the microbiota on vaccine effectiveness. Trends Immunol. 35, 526–537. https://doi.org/10.1016/J.IT.2014.07.003

Vargas-Albores, F., Martínez-Córdova, L.R., Hernández-Mendoza, A., Cicala, F., Lago-Lestón, A., Martínez-Porchas, M., 2021. Therapeutic modulation of fish gut microbiota, a feasible strategy for aquaculture? Aquaculture 544, 737050. https://doi.org/10.1016/J.AQUACULTURE.2021.737050

Varsamos, S., Nebel, C., Charmantier, G., 2005. Ontogeny of osmoregulation in postembryonic fish: A review. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 141, 401–429. https://doi.org/10.1016/J.CBPB.2005.01.013

Vestrum, R.I., Forberg, T., Luef, B., Bakke, I., Winge, P., Olsen, Y., Vadstein, O., 2021. commensal and opportunistic bacteria present in the microbiota in Atlantic cod (*Gadus morhua*) larvae differentially alter the hosts’ innate immune responses. Microorg. 10, 24. https://doi.org/10.3390/MICROORGANISMS10010024

Vital, M., Stucki, D., Egli, T., Hammes, F., 2010. Evaluating the growth potential of pathogenic bacteria in water. Appl. Environ. Microbiol. 76, 6477–6484. https://doi.org/10.1128/AEM.00794-10

Wang, H., Pryor, M.A., Edwards, M.A., Falkinham, J.O., Pruden, A., 2013. Effect of GAC pre-treatment and disinfectant on microbial community structure and opportunistic pathogen occurrence. Water Res. 47, 5760–5772. https://doi.org/10.1016/J.WATRES.2013.06.052

Wang, Y.B., Li, J.R., Lin, J., 2008. Probiotics in aquaculture: challenges and outlook. Aquaculture 281, 1–4. https://doi.org/10.1016/J.AQUACULTURE.2008.06.002

Wang, Y.B., Xu, Z.R., Xia, M.S., 2005. The effectiveness of commercial probiotics in northern white shrimp *Penaeus vannamei* ponds. Fish. Sci. 71, 1036–1041. https://doi.org/10.1111/J.1444-2906.2005.01061.X

Wargo, J., 2020. Modulating gut microbes. Science 369, 1302–1303.

Westermann, A.J., Gorski, S.A., Vogel, J., 2012. Dual RNA-seq of pathogen and host. Nat. Rev. Microbiol. 10, 618–630. https://doi.org/10.1038/NRMICRO2852

Wietz, M., Duncan, K., Patin, N. V., Jensen, P.R., 2013. Antagonistic interactions mediated by marine bacteria: the role of small molecules. J. Chem. Ecol. 397 39, 879–891. https://doi.org/10.1007/S10886-013-0316-X

Winter, D.J., 2017. rentrez: an R package for the NCBI eUtils API. R Journal 9, 520–526.

Wolin, M.J., Miller, T.L., Stewart, C.S., 1997. Microbe-microbe interactions. Rumen Microb. Ecosyst. 467–491. https://doi.org/10.1007/978-94-009-1453-7_11

Wu, S., Ren, Y., Peng, C., Hao, Y., Xiong, F., Wang, G., Li, W., Zou, H., Angert, E.R., 2015. Metatranscriptomic discovery of plant biomass-degrading capacity from grass carp intestinal microbiomes. FEMS Microbiol. Ecol. 91.

https://doi.org/10.1093/FEMSEC/FIV107

Xiao Joe, J.T., Chiou, P.P., Kuo, C.Y., Jia Lin, J.H., Wu, J.L., Lu, M.W., 2019. The microbiota profile and transcriptome analysis of immune response during metamorphosis stages in orange spotted grouper (*Epinephelus coioides*). Fish Shellfish Immunol. 90, 141–149. https://doi.org/10.1016/J.FSI.2019.03.063

Xing, M., Hou, Z., Yuan, J., Liu, Y., Qu, Y., Liu, B., 2013. Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*). FEMS Microbiol. Ecol. 86, 432–443. https://doi.org/10.1111/1574-6941.12174

Xu, H., Liu, E., Li, Y., Li, X., Ding, C., 2017. Transcriptome analysis reveals increases in visceral lipogenesis and storage and activation of the antigen processing and presentation pathway during the mouth-opening stage in zebrafish larvae. Int. J. Mol. Sci. 18, 1634. https://doi.org/10.3390/ijms18081634

Xu, Y., Li, Y., Xue, M., Yang, T., Luo, X., Fan, Y., Meng, Y., Liu, W., Lin, G., Li, B., Zeng, L., Zhou, Y., 2021. Effects of dietary *Saccharomyces cerevisiae* YFI-SC2 on the growth performance, intestinal morphology, immune parameters, intestinal microbiota, and disease resistance of crayfish (*Procambarus clarkia*). Anim. 11. https://doi.org/10.3390/ANI11071963

Yang, Y., Zhu, X., Zhang, H., Chen, Y., Song, Y., Ai, X., 2021. Dual RNA-Seq of trunk kidneys extracted from channel catfish infected with yersinia ruckeri reveals novel insights into host-pathogen interactions. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.775708

Yoon, J. Bin, Hwang, S., Yang, J.H., Lee, S., Bang, W.Y., Moon, K.H., 2022. Dynamics of the gut microbiome and transcriptome in korea native ricefish (*Oryzias latipes*) during chronic antibiotic exposure. Genes (Basel). 13, 1243. https://doi.org/10.3390/GENES13071243/S1

Yost, S., Duran-Pinedo, A.E., Teles, R., Krishnan, K., Frias-Lopez, J., 2015. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. Genome Med. 7. https://doi.org/10.1186/S13073-015-0153-3

Yu, Y., Raka, F., Adeli, K., 2019. The role of the gut microbiota in lipid and lipoprotein metabolism. J. Clin. Med. 8. https://doi.org/10.3390/JCM8122227

Yu, Y., Wang, Q., Huang, Z., Ding, L., Xu, Z., 2020. Immunoglobulins, mucosal immunity and vaccination in teleost fish. Front. Immunol. 11, 2597. https://doi.org/10.3389/FIMMU.2020.567941/BIBTEX

Yue, K., Shen, Y., 2022. An overview of disruptive technologies for aquaculture. Aquac. Fish. 7, 111–120. https://doi.org/10.1016/J.AAF.2021.04.009

Yúfera, M., Halm, S., Beltran, S., Fusté, B., Planas, J. V., Martínez-Rodríguez, G., 2012. Transcriptomic characterization of the larval stage in gilthead seabream (*Sparus aurata*) by 454 pyrosequencing. Mar. Biotechnol. 14, 423–435. https://doi.org/10.1007/s10126-011-9422-3

Zambonino Infante, J.., Cahu, C.., 2001. Ontogeny of the gastrointestinal tract of marine fish larvae. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 130, 477–487.

https://doi.org/10.1016/S1532-0456(01)00274-5

Zhang, H., Sun, Z., Liu, B., Xuan, Y., Jiang, M., Pan, Y., Zhang, Y., Gong, Y., Lu, X., Yu, D., Kumar, D., Hu, X., Cao, G., Xue, R., Gong, C., 2016. Dynamic changes of microbial communities in *Litopenaeus vannamei* cultures and the effects of environmental factors. Aquaculture 455, 97–108. https://doi.org/10.1016/J.AQUACULTURE.2016.01.011

Zou, J., Mercier, C., Koussounadis, A., Secombes, C., 2007. Discovery of multiple beta-defensin like homologues in teleost fish. Mol. Immunol. 44, 638–647. https://doi.org/10.1016/J.MOLIMM.2006.01.012

CHAPTER 2

Transcriptomic insights into core molecular changes associated with metamorphosis in fish larvae across diverse hatcheries and species

Transcriptomic insights into core molecular changes associated with metamorphosis in fish larvae across diverse hatcheries and species

Manuscript in preparation.

Transcriptomic insights into core molecular changes associated with metamorphosis in fish larvae across diverse hatcheries and species

Babak Najafpour[1], Soraia Santos[1], Manuel Manchado[2], Aurora Vidal[3], Andreas Tsipourlianos[5], Adelino VM Canário[1,4], Katerina A Moutou[5] and Deborah M Power[1,4]*

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

[2] IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500, El Puerto de Santa María, Cádiz, Spain

[4]International Research Center for Marine Biosciences (Ministry of Science and Technology), Shanghai Ocean University, Shanghai, China

[5]Department of Biochemistry and Biotechnology, University of Thessaly, Biopolis, 45100 Larissa, Greece

CReDit statement:

Babak Najafpour: Methodology - Lead, Formal analysis, Validation, Investigation, Bioinformatics -Lead, Visualization, Writing – Original Draft, Review & Editing, Soraia Santos: Sample Organization; Manuel Manchado: Supportive Data; Aurora Vidal: Supportive Data; Andreas Tsipourlianos: Validation; Adelino VM Canário: Review & Editing; Katerina A Moutou: Review & Editing; Deborah M Power: Writing - Review & Editing, Investigation, Supervision, Project Administration, Funding Acquisition.

2.1. Abstract

Early development is a critical stage in fish since this is when the highest mortality occurs and biotic and abiotic conditions can imprint production traits in adults. In spite of the substantial phenotypic changes during early life stages understanding how settings under commercial rearing affect development is limited. Therefore, a transcriptomic study was designed to compare critical larval stages during metamorphosis using gilthead seabream and Senegalese sole larvae collected from several commercial hatcheries mainly from flexion and mid-metamorphosis stages. The clustering of gene expression profiles was driven by age and weight in a PCA analysis of the 25 gilthead sea bream libraries. Differential expressed genes were identified by comparing larvae at different developmental stages/cluster, where comparison I (flexion vs mid-metamorphosis) identified 2243 and comparison II (early-metamorphosis vs mid-metamorphosis) resulted in 2299 differentially expressed transcripts. The modified gene transcripts in larvae from different hatcheries are suggested as core plastic gene markers during metamorphosis where their association with age and weight were further identified by WGCNA analysis. The analysis of RNA-seq datasets of Senegalese sole used to extend the results to the asymmetric and benthic fish that was yielded 42651 differentially expressed transcripts through comparison III (pre-metamorphosis vs post metamorphosis) and included common and diverse transcriptional changes compared to gilthead sea bream a symmetric and pelagic fish.The enrichment of gene sets is associated with diverse biological processes during larval development, including growth, development, metabolism, nervous system, and immune system. The nervous system or FoxO signaling pathway was of enriched process and pathways at very early development, at 24 DPH. The suppression of immune-associated GOs and KEGG pathways, including antigen processing and presentation and lysosome, further reflect that the gilthead seabream (21 and 46 dph) and Senegalese sole (9 dph) larvae are not fully immunocompetent at very early stages. In line with this, the abundant of most pattern recognition receptors (PRR) genes, including lectin-like receptors with overall higher expression level, and activation of genes involved in complement system at mid-metamorphosis were suggested. Some biological processes were more activated at specific stages, including up-regulation of ‘Enteroendocrine cell differentiation’ and ‘lipid homeostasis’ at 46 dph gilthead seabream. While the expression of most endocrine genes was stable, the modification of some endocrine associated genes (*dio1*, *dio2*, *cldn1*, *ing4*, *Pou3f4 fgf22*) highlights their substantial role for metamorphosis. Transcriptomes meta-analysis of diverse species (Senegalese sole versus gilthead seabream) provided insight into common molecular expression patterns underlying larvae maturation during metamorphosis, including significant

modification in ‘endopeptidase activity’, ‘peptidase inhibitor activity’, ‘multicellular organism development’, and ‘lysosomes’.

2.2. Introduction

Early development is a critical stage in fish aquaculture since in this period the highest mortality occurs. Furthermore, biotic (e.g., larval density, food) and abiotic factors (e.g., temperature, pH) influence development and trigger epigenetic imprinting, influencing production traits in adults such as the stress response, disease resistance and growth (Burggren, 2020; Claramunt and Wahl, 2000; Cnaani, 2006; Lyimo et al., 1992; Robinson et al., 2019). Epigenetic imprinting can be transmitted between generations (Burggren et al., 2016; Ho and Burggren, 2010; Pittman et al., 2013). However, despite numerous studies using dietary additives (Chakraborty et al., 2014; Sutili et al., 2018), pre-probiotics (Brugman et al., 2018), and vaccines (Brudeseth et al., 2013) to improve fish health and survival, consistent and predictable production of high-quality larvae still remains a challenge for the aquaculture industry. Thus, understanding the regulatory factors that influence epigenetic imprinting, development and phenotypic plasticity may help establish rearing practices that enhance larval performance and favour desired adult phenotypes.

The molecular mechanisms explaining body shape during fish development are still enigmatic, although much work has been invested in characterizing this process in flatfish (Power et al., 2008). Niche occupancy, along the benthic-pelagic axis is one of the factors that has imposed body shape evolution in fish, and this is more evident in benthic compared to demersal or pelagic fish (Friedman et al., 2020). Comparative analysis of marine fish eggs and larvae, indicates that the development of flat and round fish has many similarities, such as simple V-shaped myotomes in muscle, a straight undeveloped gastrointestinal tract (GI) at hatch, increased goblet cell number and the development of the stomach as the GI tract matures, appearance of neuromasts cells with protruding kinocilia and stereocilia on the body surface, and development of fully pigmented and functional eyes within the first week of hatch (Falk-Petersen, 2005; Rønnestad et al., 2013). Numerous complex developmental changes underpin the larval – juvenile transition (metamorphosis), and the importance of a well-regulated transition for fitness and improved production traits in the grow-out stage makes this developmental transition critical for aquaculture. A better understanding of teleost metamorphosis in general can contribute to the understanding and integration of the role of endocrine regulation on development, and explain how morphologies and lifestyles have evolved and are influenced by the environment (McMenamin and Parichy, 2013).

With the advent of high throughput sequencing techniques, it has been possible to map the dynamics of gene expression during larval development (e.g., differentiation, cellular

proliferation, growth) in different species and obtain a detailed insight into these changes (Mazurais et al., 2011; Qian et al., 2014). Among the findings, it has been possible to identify gene clusters that have similar dynamics in different species, such as those implicated in digestion (e.g., trypsin) and muscle development (e.g., myosin heavy and light chain; Darias et al., 2008; Douglas et al., 2008). Groups of genes that underwent significant modifications at specific developmental stages and experimental conditions during larval ontogeny were also identified, such as genes involved in ATP synthesis in European sea bass (Darias et al., 2008) and in fat digestion and absorption in zebrafish (Xu et al., 2017). Furthermore, in Atlantic halibut during metamorphosis most transcripts (approx. 98 %) were not thyroid hormone (TH)-responsive indicating that a relatively small number of genes orchestrate this remarkable process (Alves et al., 2016).

Previous transcriptome studies of gilthead seabream have mainly focused on juvenile and adult fish due to the commercial importance of this species in the South of Europe. Production challenges are largely linked to disease, diet, growth and physicochemical conditions (Cara et al., 2005; Hampel et al., 2017; Ibeas et al., 1994; Martos-Sitcha et al., 2014; Mininni et al., 2014; Muniesa et al., 2020) and transcriptome studies have focused on adult tissue relevant for these traits like, skeletal muscle, intestine, blood, head-kidney, and the gill (Calduch-Giner et al., 2013; Garcia de la serrana et al., 2012; Louro et al., 2016; Pérez-Sánchez et al., 2019; Vieira et al., 2013). Studies of gilthead seabream hatchery stages are less common and those that exist are focused on very early development (embryo to 168 hph), or have targeted specific tissues (e.g., kidney) or experimental conditions (Sarropoulou et al., 2005, 2016). Such studies have generated important species specific molecular resources (Yúfera et al., 2012), but there are relatively few comparative studies of different developmental stages.

Common morpho-physiological changes imply sets of core genes driving larval development among different species. However, diverse conditions in aquaculture settings may lead to different developmental rates and this could contribute to variation in transcriptomics responses. The primary goal of this study was to integrate the transcriptome profiles of gilthead seabream larvae collected from diverse sites, encompassing variations in broodstock, installations, production scale, and hatchery management. Subsequently, the study aimed to identify differential transcriptomic responses occurring during metamorphosis, irrespective of the origin of the larvae. In addition, this study aimed to uncover shared gene expression patterns during transition phases in symmetric and pelagic (gilthead seabream) and asymmetric and benthic (Senegalese sole) larvae. The knowledge generated contributes to a better

understanding of metamorphosis in commercial hatcheries and generates candidate stage and quality-related gene markers.

2.3. Material and Methods

2.3.1. Larvae

Gilthead seabream larvae were collected from three different aquaculture sites located in France, Italy and Greece, from March-2018 to December-2018, preserved into RNA-later and stored at -20 ºC until RNA extraction. The larvae were from different batches collected before metamorphosis at the flexion stage (23-25 days post hatch, dph), at early- and at mid-metamorphosis (43-60 dph, **Figure 2.1**). These two stages were chosen since they are easy to identify, favouring collection of uniform samples between sites. The list of collected samples used in the analysis with variables age, weight, and length is detailed in Najafpour et al. DIB). Based on sample availability, two RNA-seq projects (A and B) were run separately, where project A included larvae from sites C (flexion, mid-metamorphosis stages) and A (mid-metamorphosis stage), and project B comprised larvae from sites B (end of larval rearing to mid-metamorphosis stage) and A (mid-metamorphosis stage).

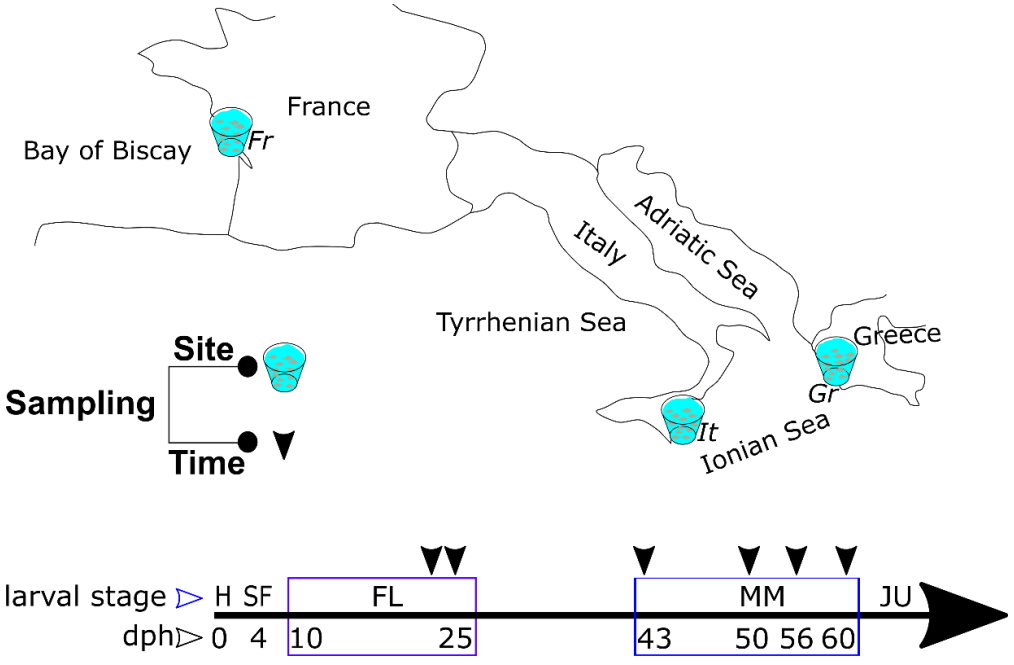


Figure 2.1. The sampling scheme of gilthead seabream during larval development from

three hatcheries (A, B, C). H = hatching; SF = start of feeding; FL = Flexion; ELR = end of larvae rearing; MM = mid-metamorphosis; MT = Metamorphosis. Arrows show sampling time points.

The larval period of the Senegalese sole (approximately 30 days) is shorter than of the gilthead seabream (~ 60 days) and is subdivided into pre-metamorphosis (7 – 12 dph, stage 0), metamorphosis (14 dph, Stage 1, 15 dph, Stage 2 and 16 dph Stage 3) and post-metamorphosis (20 dph, Stage 4, Fernández-Díaz et al., 2001). Senegalese sole (*Solea senegalensis*) larvae were collected at 9 dph (pre-metamorphosis) during the shift in the body axis and 20 dph during the shift in the position of the eyes (post-metamorphosis) from IFAPA center in Spain. All samples were washed using diethylpyrocarbonated (DEPC) water, frozen in liquid nitrogen, and stored at −80ºC until analysis.

2.3.2. RNA extraction

For RNA extraction, a gilthead seabream sample was comprised a pool of 3 larvae 23-25 dph or 1-2 larvae 43-60 dph, and a Senegalese sole sample was contained 5 larvae 9 dph (Stage 0) or 20 dph (Stage 4). Gilthead seabream larvae were processed at CCMAR and were defrosted in lysis buffer and homogenized by mechanical disruption with two iron beads (5 mm) in a Tissue lyser II (Qiagen, Germany) using 3 cycles (30 Hz) of 30 seconds at room temperature. Total RNA from whole larvae homogenates was extracted using the E.Z.N.A. Total RNA Kit I (VWR, USA) according to the manufacturer's instructions. Genomic DNA was removed by DNAse I treatment performed directly on the column during RNA isolation, following the manufacturer's instructions. Senegalese sole larvae were processed at IFAPA, and were homogenised using the Fast-prep FG120 instrument (Bio101) and Lysing Matrix D (Q- Bio-Gene) for 40 s at speed setting 6. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) following the manufacturer's indications and treated twice with DNase (RNase-Free DNase kit, Qiagen) for 30 minutes.

2.3.3. Library preparation and sequencing

The preparation and sequencing of the gilthead sea bream sequencing libraries were done by Novogene (Shanghai, China). Before sequencing library preparation RNA quality was checked using a 2100 Bioanalyser (Agilent Technologies, Inc., Santa Clara CA, USA) and only high-quality RNA samples (RIN ≥ 8.0) were used. The 25 paired-end (2 × 150bp read

length) gilthead sea bream RNA sequencing libraries, 11 libraries from project A and 14 libraries from project B, were prepared using a TruSeq TM RNA sample preparation Kit (Illumina, San Diego, CA) and sequenced using an Illumina HiSeq Xten. The Senegalese sole Illumina libraries were constructed at the MGX platform in Montpellier (Montpellier, France) using the TruSeq RNA Sample Preparation Kit v2, following the manufacturer's instructions. Each library was sequenced using TruSeq SBS Kit v3-HS, in paired end mode, 2 × 76 bp (*Solea senegalensis*) on HiSeq2000 sequencing system (Illumina, Inc) following the manufacturer's protocol (more details in Benzekri et al., 2014)

Raw RNA-seq reads data were submitted to the NCBI Short Read Archive (gilthead sea bream accession number PRJNA956882; Senegalese sole accession number PRJNA241068 for pre-metamorphic larvae at 9 dph and accession number PRJNA261151 for post-metamorphic larvae at 20 dph).

2.3.4. Quality control and differential gene expression

Quality-check of the raw reads was done with FASTQC using default parameters (Andrews, 2010) in the GALAXY (Afgan et al., 2018). After quality checks, the reads were mapped to the gilthead sea bream (NCBIassembly name: fSpa Aur1.1, RefSeq assembly accession: GCF_900880675.1) and Senegalese sole (Male_LA_Total, Claros et al., 2020) reference genomes using the Bioconductor package Rsubread (Liao et al., 2013). The number of reads mapped per gene was counted using the "*featureCounts"* function in R with default parameters and GTF.attrType = "gene transcript".

Genes with very low expression were filtered out using the *cpm* function in edgeR (Robinson et al., 2009) and only transcripts with at least one count-per-million reads (cpm) in four samples were retained. The distance between the global transcriptome profile of all the analysed samples was assessed by principal components analysis (PCA) using the *MDSplot* function in *R*. The *MDSplot* was built using normalized data with the *voom* function in the limma package. Larval transcriptomes that clustered together had the same stage classification and were considered to represent samples at the same developmental stage. To track significant changes in gene expression and their functional profiles, differentially expressed genes (DEGs) were identified between the independent clusters of the larval transcriptomes generated by PCA.

Differentially expressed genes were identified using normalized read counts and computing moderated t-statistics by empirical Bayes moderation of the standard errors in the

limma package (Ritchie et al., 2015). Based on the outcome of the PCA analysis, two comparisons (I and II) were made using the gilthead sea bream larval transcriptomes and including data from the two sequencing projects A and B. In comparison I, the transcriptomes at the flexion stage gilthead sea bream larvae (age = 24 ± 0.7 dph; weight = 4.4 ± 1.8 mg) were compared to the transcriptomes of mid-metamorphosis stage (age = 51 ± 0.5 dph; weight = 47.1 ± 9.7 mg). In comparison II, the transcriptomes of the early-metamorphosis stage larvae (age = 46 ± 1.7 dph; weight = 16.5 ± 4.4 mg) were compared to the transcriptomes of mid metamorphosis larvae (age = 54 ± 1.6 dph; weight = 49.9 ± 10.6 mg). To minimise bias, comparisons were made using only samples with a similar range of read numbers in each of the comparisons, and samples that were not clustering in the PCA were also excluded, resulting in 7 samples in comparison I and 13 samples in comparison II. For the Senegalese sole, DEGs were identified by comparing transcriptomes of 9 dph (Stage 0, n=3) and 20 dph (Stage 4, n=3) larvae (comparison III).

A library of 258 endocrine genes coding 464 transcripts extracted from the NCBI Refseq database was used to analyze the potential modifications in endocrine repertoire of the gilthead sea bream.

The common DEGs associated with metamorphosis in the two species were extracted from comparisons I and III and were used to identify genes with similar expression patterns (up-regulated or down-regulated) during metamorphosis. Orthologues of the identified DEGs were extracted from the zebrafish genome.

2.3.5. Functional analysis

DEGs were annotated by local BLAST against the Ensembl zebrafish translated proteins (Danio_rerio.GRCz11.pep.all.fa.gz) and their orthologues were extracted. The cut-off accepted for orthologue annotation was an expected p-value >$1e^{-5}$ and identity > 40 %.

The R clusterProfile package was used to carry out the Gene Ontology (GO) over-representation analysis using the *enrichGO* function and the gene set enrichment analysis (GSEA) using the *gseGO* function for each of the comparisons I-III (Yu et al., 2012). Redundant GO terms were removed, and GO terms were considered significantly enriched if $p < 0.05$ and the Benjamini-Hochberg (BH) adjusted $p < 0.05$. The Bioconductor release v 3.14 of the genome annotation for Zebrafish (org.Dr.eg.db.3.14.0) was used in R as the background for *gse*A (Carlson, 2019). The over-represented GOs were visualized based on the relative percentage of gene ratio of each GO using the R package plotrix v 3.8-2 and *pie* function

(Lemon, 2006).

KEGG Gene Set Enrichment was analysed using the the *gseKEGG* function of the R package clusterProfile v 4.2.2 (Yu et al., 2012) with the same statistical criteria as applied to GSEA (with and without a BH adjusted p-value). The classification of DEGs involved in specific processes (e.g., immune response) was further determined by manual searches of the UniProt (www.uniprot.org) and PANTHER (www.pantherdb.org) databases (Mi et al., 2019).

Weighted gene co-expression network analysis (WGCNA) was used to identify co-expressed gene modules in comparisons I and II above (Langfelder and Horvath, 2008). WGCNAs analysis was run independently for the two groups of samples (each group included samples used in the separate comparisons I and II). WGCNA I included larvae at flexion and mid-metamorphosis (24 dph versus 51 dph) and WGCNA II included larvae at early and mid-metamorphosis (46 dph versus 54 dph). Briefly, absent or low abundance genes were filtered out, and data were normalized and transformed using the *vst* function in the DESeq2 R package. The power soft threshold for WGCNA was determined using the output of *sft* object, and $R^2 > 0.85$ was applied. The gene network was constructed, and modules were identified using the *blockwiseModules* function in the WGCNA R package. Modules that were significantly associated with age and weight traits were identified by correlating each module (applying the eigengene concept) with the traits of interest (Pearson correlation method). The quantitative measure of module membership (MM) was also defined as the correlation of the module eigengene and the gene expression profile using default parameters.

WGCNA modules that had the highest association with weight and age were selected and functional information mapped. The module associated gene sets were mapped to gilthead seabream functional data through the g:Profile web server for functional enrichment analysis (https://biit.cs.ut.ee/gprofiler/gost). Gene ontology enrichment results were visualized using the python package GO-Figure! that provided a comprehensive summarized list of GO terms (Reijnders and Waterhouse, 2020).

2.3.6. Real-time quantitative RT-PCR

Real-time quantitative RT-PCR of six candidate genes (*hbae5*, *map4*, *lct*, *chia*, *rh50*, and *tcnba*) selected among the DEGs was carried out on a 10 ul final volume using KAPA SYBR® FAST qPCR Master Mix (2X) Kit (Kapa Biosystems) and the appropriate concentration of each transcript specific set of primers. Primers were designed using the Primer3 (v.0.4.0, Untergasser et al., 2012) and Beacon Designer software and all primers were

designed to span the coding sequences of the transcripts. The amplification cycle was: 5 min at 95 ºC, followed by 40 cycles of 95 ºC for 20 s and 60 ºC for 20 s, followed by the dissociation curve step to verify single product amplification. Each reaction was performed in duplicate. A standard curve using a dilution series (1:5, 1:10, 1:20, 1:50, 1:100) from pooled cDNA was prepared to estimate amplification efficiency (**Table 2.1**). Three housekeeping genes (*ef1*, *rpl13a*, and *rpl19*) were chosen as the reference genes based on geNorm rating (Vandesompele et al., 2002). From the Ct values, the starting content of the target sequence (R0) was calculated as R0 = $Threshold/(1+efficiency)^{ct}$ normalized to the geometric mean of the three housekeeping genes.

Table 2.1. The selected candidate gene set and primers used in qPCR reactions.

Gene name	Gene symbol	Primer sequence	Product size	Primer efficiency	Refseq Accession
Hemoglobin, alpha embryonic 5	*hbae5*	*Fw_CAGCGTCTATCCCCAAACCA* *Rv_ GGTCGTCGATCTTGGACACA*	131	100.85	XM_030399324
Microfibril-associated glycoprotein 4-like	*map4*	*Fv_CTCCGAGAGCAGAGAGAGTTTT* *Rv_TGCCTTTGGCAAGAACTGTG*	122	99.18	XM_030413465
Lactase-phlorizin hydrolase-like	*lct*	*Fw_GCAGAAGACGATACTCCCCA* *Rv_TGGTGCCAAATACACTTCCCT*	157	100.50	XM_030427745
Rh50-like protein	*rh50*	*Fw_TGTGTTGGTCGTCACCCTTT* *Rv_GTTTGCCGCCTACTTTGGTC*	185	103.77	XM_030420128
Transcobalamin beta a	*tcnba*	*Fw_GAGGCCATGAGTCTCTTCAGG* *Rv_TGCTTCCAGGGACATTGACT*	267	104.88	XM_030405178
Acidic mammalian chitinase-like	*chia*	*Fw_TGCTACTTCACCAACTGGGC* *Rv_TTCAGGCCATTGAAGGTCGT*	179	96.22	XM_030420597

2.4. Results

2.4.1. RNA-seq libraries and statistics

The average number of the gilthead sea bream paired ended reads per sequencing library (R1+R2) was 12,143,070 (11 samples) in project A and 46,420,480 (14 samples) in project B. The average number of merged reads that mapped to the gilthead seabream reference genome was 5,810,781 in project A and 22,018,717 in project B, or 95.1% reads mapped to the reference genome (Najafpour et al. DIB for more details). The average number of Senegalese sole paired-end reads per sequencing library was 56,451,220 (R+R2) yielding 26,369,914

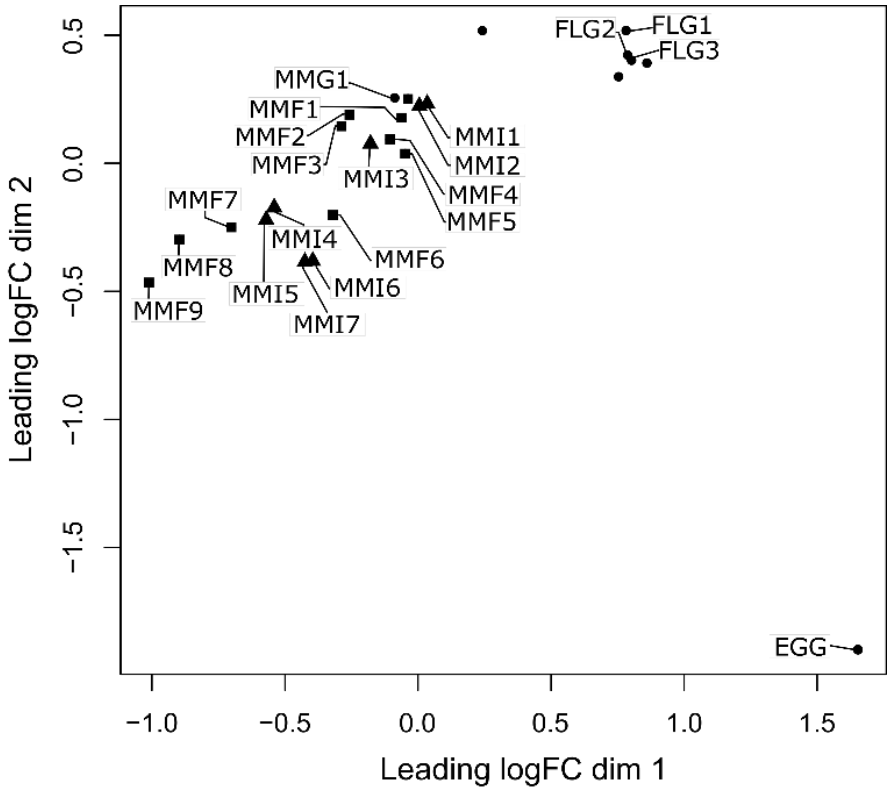
merged reads or 93.3 % were mapped to the reference genome.

2.4.2. Clustering and differential gene expression

The PCA of the gene expression profiles of the 25 gilthead sea bream sequencing libraries originated three main clusters (**Figure 2.2**). Overall, the clustering was driven by age/stage and weight rather than aquaculture site (**Figure 2.2**). One cluster contained larvae with an average age and weight of 24 dph and 5.9 mg (larvae at flexion, PCA coordinates: x-axis ~ 0.8, y-axis ~ 0.4) the second cluster contained larvae of 48 dph and 38.8 mg (mid-metamorphosis, PCA coordinates: x-axis ~ -0.01, y-axis ~ 0.2) and the third cluster contained larvae of 54 dph and 43.3 mg (metamorphosis, PCA coordinates: x-axis ~ -0.7, y-axis ~ -0.3, **Figure 2.2**).

Comparison I (flexion vs mid-metamorphosis) yielded 2243 DEGs of which 1083 were up-regulated, 1160 were down-regulated (adjusted p-value (BH) < 0.05, **Figure 2.3**). Comparison II (early-metamorphosis vs mid-metamorphosis) resulted in 2299 DEGs of which 1153 were up-regulated and 1146 were down-regulated (adjusted p-value (BH) < 0.05, **Figure 2.3**). There were 684 common DEGs between the two comparisons, 1559 specific to comparison I and 1615 specific to comparison II (**Figure 2.3**, **Supplementary tables 2.1 and 2.2**).

a)



b)

Sample	Stage	Site	Age (dph)	
FLG1	*FL*	*Gr*	*25*	comparison I 24 vs 51 (dph)
FLG2	*FL*	*Gr*	*25*	
FLG3	*FL*	*Gr*	*23*	
MMG1	*MM*	*Gr*	*52*	
MMF1	*MM*	*Fr*	*50*	
MMF2	*MM*	*Fr*	*50*	
MMF3	*MM*	*Fr*	*50*	
MMF4	*MM*	*Fr*	*50*	comparison II 46 vs 54 (dph)
MMF5	*MM*	*Fr*	*50*	
MMI1	*MM*	*It*	*43*	
MMI2	*MM*	*It*	*43*	
MMI3	*MM*	*It*	*43*	
MMI4	*MM*	*It*	*56*	
MMI5	*MM*	*It*	*56*	
MMI6	*MM*	*It*	*60*	
MMI7	*MM*	*It*	*60*	
MMF6	*MM*	*Fr*	*50*	
MMF7	*MM*	*Fr*	*50*	
MMF8	*MM*	*Fr*	*50*	
MMF9	*MM*	*Fr*	*50*	

Figure 2.2. PCA analysis and identification of the samples used for the analysis of differential gene expression (DEGs). A) Multidimensional Scaling Plot of Distances (Euclidean distance) between transcriptome profiles of 25 gilthead seabream larvae samples that differed by age and weight. The different shaped symbols represent different aquaculture sites (circle = site A; triangle = site P; rectangle = site F). Three main clusters were formed and

their associations with age and weight was identified. The three clusters comprised larvae with an average age and weight of i) 24 dph and 5.9 mg (cluster included larvae at flexion stage with an approximate average position: x-axis ≅ 0.8, y-axis ≅ 0.4); ii) 48 dph & 38.8 mg (cluster included larvae at metamorphosis with an approximate position: x-axis ≅ -0.01, y-axis ≅ 0.2); and iii) 54 dph & 43.3 mg (cluster included larvae at late metamorphosis with an approximate position: x-axis ≅ -0.7, y-axis ≅ -0.3). b) the details of the samples that were selected from different clusters to perform the two differential gene expression analyses (I & II). Seven samples with a similar range of sequencing reads were selected to perform comparison I (24 vs 51 dph) between larvae at flexion and larvae at mid-metamorphosis (MM). In comparison II (46 vs 54 dph), 13 samples with a similar range of sequencing reads were chose.

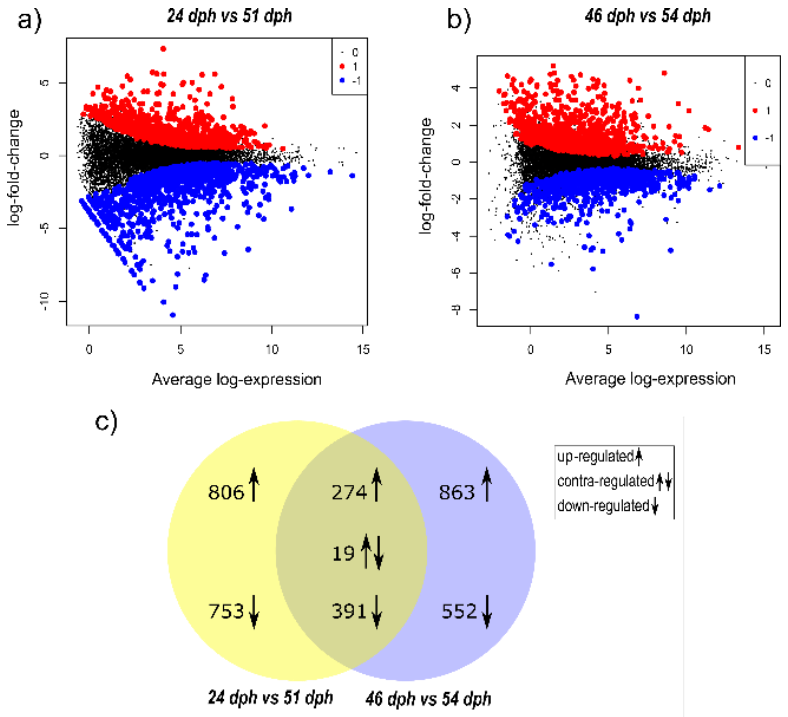


Figure 2.3. The results of the comparisons (I & II) to identify differentially expressed genes (DEG) in the selected gilthead seabream larvae samples (details in **Figure 2.2**). a) MD plot of the log-fold change of gene transcripts in the comparison of 24 dph (avg. weight = 4.6 mg) vs 51 dph (avg. weight = 43.8 mg) larvae, b) MD plot of the log-fold change of gene transcripts in the comparison of 46 dph (avg. weight = 16.54 mg) vs 54 dph (avg. weight = 41.8 mg) larvae. Significant up and down DEG transcripts are shown in red and blue, respectively. c) VennPlex diagram shows the number of common and uncommon up-regulated, down-regulated and contra-regulated DEGs from the two comparisons.

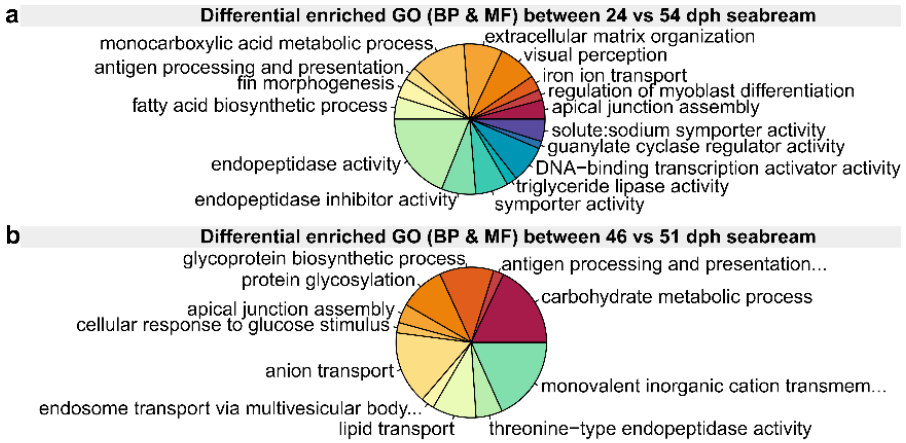
The Senegalese sole comparison III (pre-metamorphosis vs post metamorphosis) yielded 14473 unique DEGs of which 7054 were up-regulated and 7419 were down-regulated DEGs and included common and diverse DEGs compared to DEGs obtained from comparison I and II in gilthead sea bream based on the gene orthologues in zebrafish (**Supplementary table 2.3**).

2.4.3. Functional analysis

Over-represented and informative molecular functions (MF) and biological process (BP) related to each gilthead sea bream DEG comparison are summarized in **Figure 2.4**. In comparison I, the BPs showing the highest gene ratio were 'monocarboxylic acid metabolic process', visual perception, 'extracellular matrix organization', 'fatty acid biosynthetic process' and 'apical junction assembly' (**Figure 2.4**). For comparison II, they were 'carbohydrate metabolic process', 'anion transport', 'glycoprotein biosynthetic process', 'protein glycosylation', 'lipid transport' and 'apical junction assembly' (**Figure 2.4**). In Senegalese sole, comparison III yielded 'axon development', 'skeletal system development', 'ribonucleoprotein complex biogenesis', 'retina development in camera-type eye', and 'muscle cell differentiation' as top BPs (**Figure 2.4**).

GSEA identified the BP 'regulation of immune response', 'antigen processing and presentation', and 'proteolysis' to be down-regulated at gilthead seabream flexion stage (21 dph) and Senegalese sole pre-metamorphosis stage (9 dph, **Figure 2.4**). 'Enteroendocrine cell differentiation' and 'lipid homeostasis' BP were up-regulated in mid-metamorphosis (48 dph) gilthead seabream (**Figure 2.4**).

The KEGG pathways associated with 'pathogen recognition receptor' (PRR), and 'inflammatory response' (e.g., 'NOD-like receptor signalling pathway', 'Cytokine-cytokine receptor interaction'), 'lysosome', and 'phagosome' were suppressed in the younger stages compared to mid or post metamorphic larvae (**Figure 2.4**). The detailed outcome of GSEA and KEGG pathway enrichment and the full list of enriched gene sets identified are in **Supplementary figures 2.1-2.5**.



d

Enriched Biological Process	GO	9 d *Ss*	21 d *Sa*	46 d *Sa*
regulation of immune response	GO:0050776	↓	↓	-
immune system process	GO:0002376	↓	-	↓
antigen processing and presentation	GO:0019882	↓	↓	-
inflammatory response	GO:0006954	↓	-	↓
gland development	GO:0048732	-	-	↑
enteroendocrine cell differentiation	GO:0035883	-	-	↑
proteolysis	GO:0006508	↓	↓	-
lipid homeostasis	GO:0055088	↓	-	↑
Enriched pathway	**KEGG**			
Lysosome	dre04142	↓	↓	-
NOD-like receptor signaling pathway	dre04621	↓	-	↓
C-type lectin receptor signaling pathway	dre04625	↓	-	-
RIG-I-like receptor signaling pathway	dre04622	↓	-	-
Toll-like receptor signaling pathway	dre04620	↓	-	-
Cytokine-cytokine receptor interaction	dre04060	↓	-	-
Phagosome	dre04145	↓	-	↓

Figure 2.4. Gene ontology and KEGG enrichment analysis. a) enriched GO terms derived from the DEGs between 24 vs 54 dph gilthead seabream; **b**) enriched GO terms derived from the DEGs between 46 vs 51 dph gilthead seabream; **c**) enriched GO terms derived from the DEGs between 9 vs 20 dph Senegalese sole; **d**) A summary of GO terms and KEGG pathways

with enriched gene sets in less developed gilthead seabream (*Sa*) and Senegalese sole (*Ss*) larvae. The arrows pointing downwards and upwards represent up- and down-regulated gene sets related to each enriched GO term or KEGG pathway, respectively.

2.4.4. Common gilthead seabream and Senegalese sole larvae transcriptome profiles

The functional profile of common DEGs was obtained by comparing early and mid-metamorphic larvae of gilthead seabream and Senegalese sole (**Supplementary table 2.3**). 'Endopeptidase activity', 'peptidase inhibitor activity', 'multicellular organism development', 'bicellular tight junction assembly', and 'lysosome' were common enriched GO terms and included genes with similar expression pattern, such as the down-regulated *iodothyronine deiodinase type I (dio1*), *iodothyronine deiodinase II* (*dio2*), *proteolipid protein 1b* (*plp1b*), *crystallin, gamma M2b* (*crygm2b*), and the up-regulated *POU class 4 homeobox 2* (*pou4f2*), *gastrulation brain homeobox 2* (*gbx2*), and *actinodin2* (*and2*) (**Figure 2.5**, **Supplementary table 2.3**).

Overall, genes coding for proteins with endopeptidase activity (e.g., *proteasome 20S subunit beta 9a psmb9a*; *serine protease 59, prss59.2*; *chymotrypsin like elastase family member 1, cela1.6*; *trypsin I−P1−like, AL954146.1*; *complement factor I, cfi*) and peptidase inhibitor activity (e.g. *serpin peptidase inhibitor clade A, serpina1*; *complement component c3a, c3a, complement 4B, c4b*) were less expressed ($p < 0.05$) in early metamorphosis (24-day gilthead seabream and 9-day Senegalese sole) compared to mid-metamorphosis (51-day gilthead seabream and 20-day Senegalese sole, **Figure 2.5**). 'Bicellular tight junction assembly' included several claudin genes (e.g., *claudin a, cldna*; *claudin e, cldne*; *claudin k, cldnk*) in low abundance during early developmental stages in both species ($p < 0.05$, **Figure 2.5**). Several lysosome component genes (e.g., *cathepsin S, ctss2.1*; *hexosaminidase B, hexb*; *N−acetylgalactosaminidase, naga*) were down-regulated in early developmental stages of both species ($p < 0.05$, **Figure 2.5**). Phototransduction was a common KEGG pathway enriched in both species (**Supplementary table 2.3 C**), where an inverse expression pattern was identified for melanin biosynthesis and visual phototransduction genes.

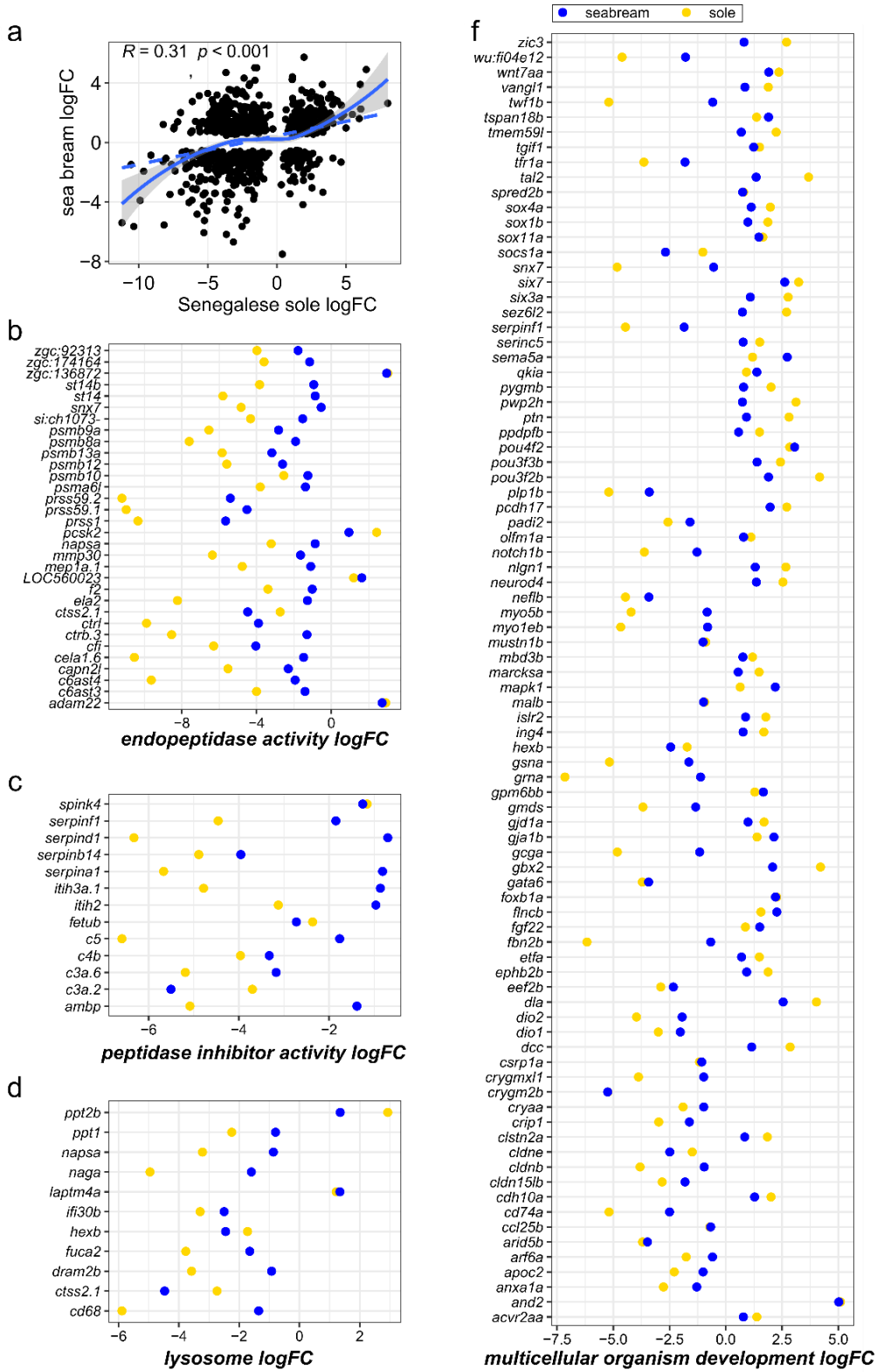


Figure 2.5. The correlation between DEGs in the gilthead seabream and Senegalese sole (a) and the expression pattern of common DEGs involved in endopeptidase activity (b), peptidase inhibitor activity (c), lysosome (d), and multicellular organism development (f) in gilthead seabream and Senegalese sole during metamorphosis. The DEGs were identified

for similar developmental stages of gilthead seabream and Senegalese sole. In the case of gilthead seabream, the transcriptome of 24 dph and 46 dph larvae were compared and in the case of Senegalese sole the transcriptome of 9 dph and 20 dph larvae were compared. The derived DEGs from gilthead seabream and Senegalese sole were compared and those that were common and had a similar expression pattern were included in the graphical representation.

2.4.5. Immune-related gene repertoire during the larval-juvenile transition

In general, immune-associated genes involved in innate and adaptive immunity were of lower abundance in younger sea bream larvae (~ 24-46 dph) compared to mid-metamorphic (~ 51-54 dph). A heatmap of 100 immune-associated DEGs with the highest levels of expression indicated that immune-related factors were up-regulated at mid-metamorphosis (51 and 54 dph) relative to the pre-metamorphic stages (24 and 46 dph, **Figure 2.6**). Several genes associated with adaptive immunity were significantly less abundant before metamorphosis in larvae at 46 dph. Some immune gene transcripts of the major histocompatibility complex (e.g., *major histocompatibility complex class I-related gene protein-like*, XM_030411748), complement (e.g., *complement C1q-like protein 4*, XM_030431421) and lectin (e.g., *galactose-specific lectin nattectin-like*, XM_030422915) were absent at flexion (~ 24 dph) compared to mid-metamorphosis (~ 50 dph).

Among the pattern recognition (PRR) genes,there was generally lower abundance in larvae at 24 and 46 dph compared to 51-54 dph, a pattern common with the detected expression of most other immune-associated genes (**Figure 2.7**, **Supplementary table 2.4 A, B**). Overall, during gilthead seabream larval development, members of the Leucine rich repeat protein (LLR) genes had the highest level of expression at all stages, followed by NOD-like receptor (NLR), and Toll-like receptor (TLR). Of the 16 unique TLR genes identified, only TLR5 was significantly modified between flexion and mid-metamorphosis (**Supplementary table 2.4 A**). 19 *nlr*-like genes were significantly changed during metamorphosis in comparisons I and II and included representatives of NLR containing CARD domain and NLR containing PYD domain sub-families (**Supplementary table 2.4 A**). Eight *nlr*-like genes were down-regulated at flexion compared to mid-metamorphosis (**Supplementary table 2.4 A**). Of the 167 unique LLR genes, 39 were significantly modified (**Supplementary table 2.4 A**). Most of the lectin genes that changed significantly were low abundance in younger larvae at 24 dph compared to older larvae at 51 dph, and in larvae at 46 dph compared to 54 dph (**Supplementary table 2.4 A**).

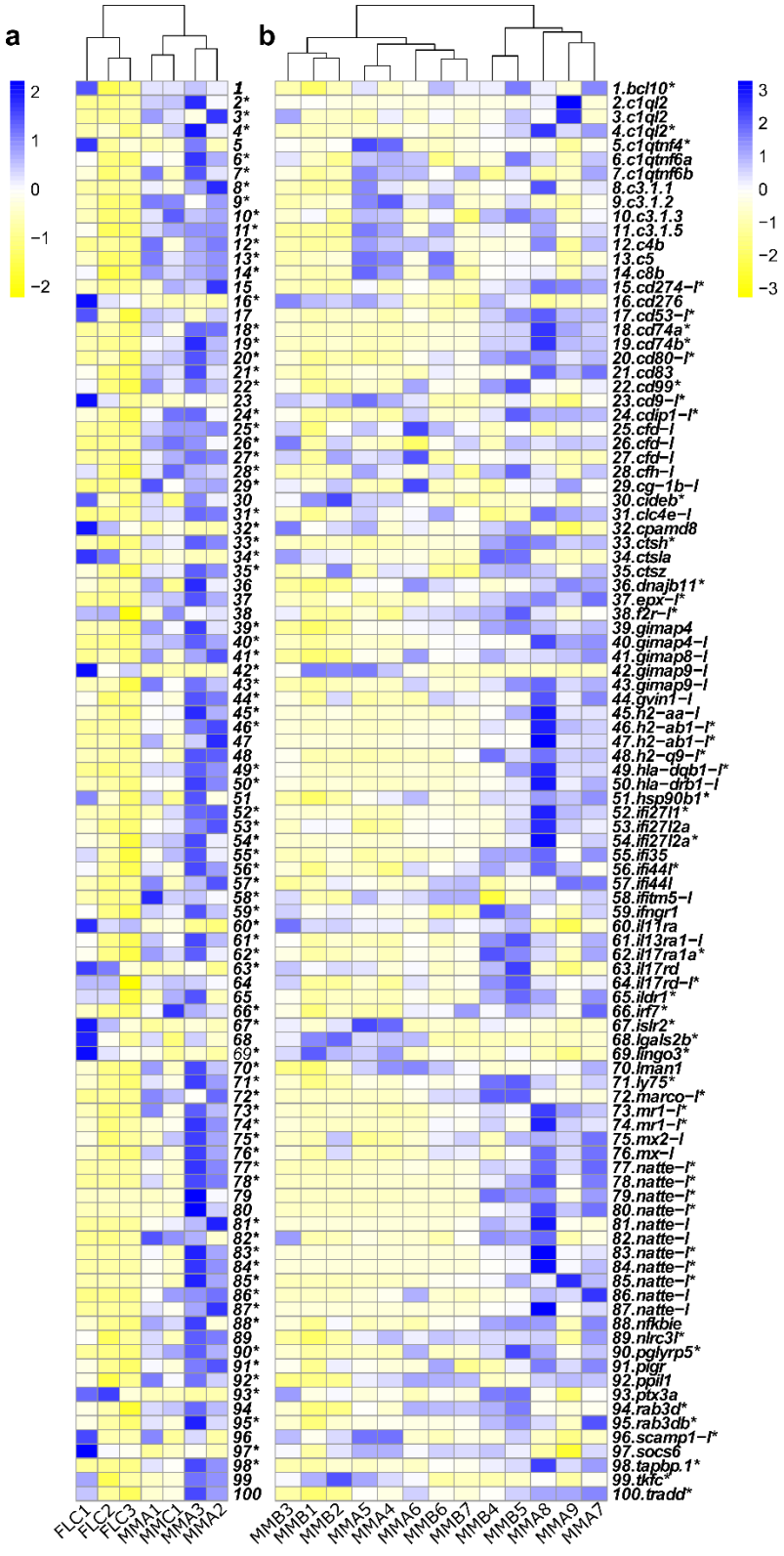


Figure 2.6. Heatmap of immune-associated gene transcripts represented in the DEGs of gilthead seabream larvae identified in comparisons I & II. a) In comparison I (24 vs 51 dph gilthead seabream larvae) seven larval samples were included and DEGs were derived by

comparing the transcriptome of 3 larvae at flexion (FL) with the transcriptome of 4 larvae at mid-metamorphosis (MM) (details in **Figure 2.2**). **b**) In comparison II (46 vs 54 dph gilthead seabream larvae) thirteen larval samples at mid-metamorphosis (MM) were analysed. The Refseq accesion of the gene transcripts used in the heatmap are presented in **Supplementary table 2.4**.

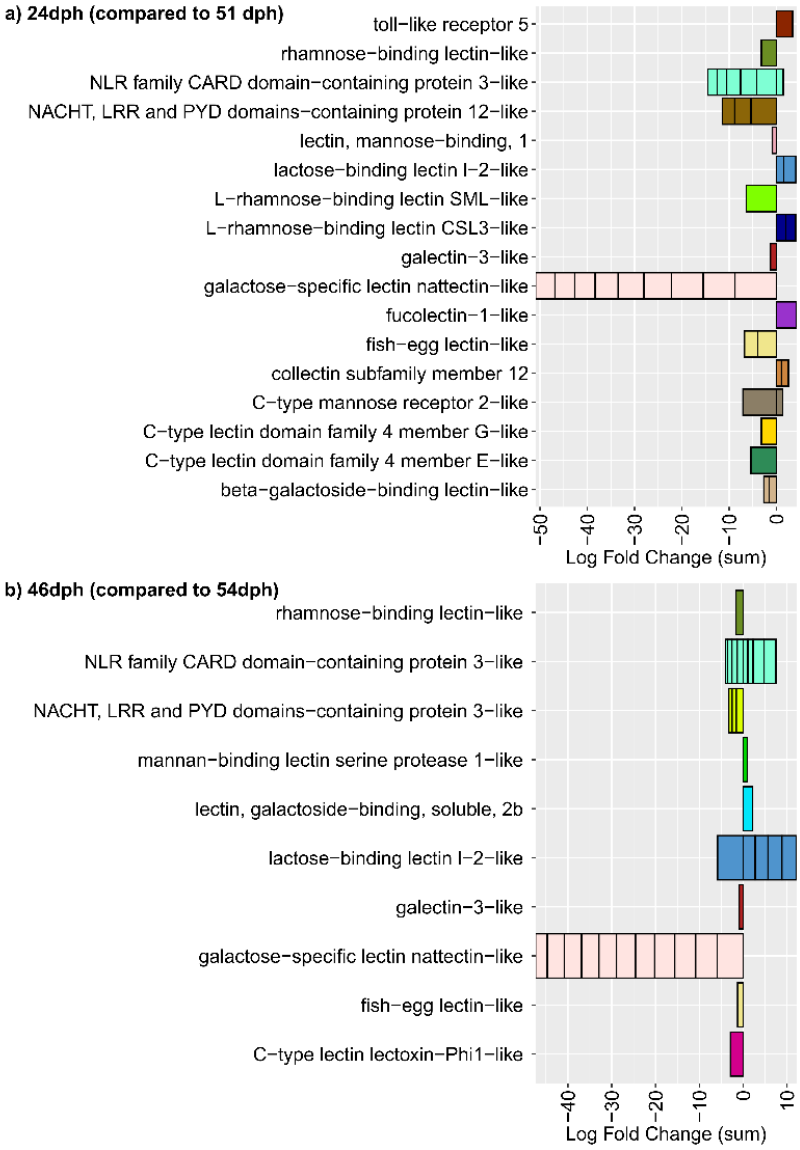


Figure 2.7. The log-fold change of pathogen recognition receptors (PRRs) during gilthead seabream larvae ontogeny. a) PRR gene transcripts that were significantly modified in larvae at 24 dph compared to larvae at 51 dph. **b**) PRR gene transcripts that were significantly changed when the transcriptome of larvae at 46 dph was compared to the transcriptome of larvae at 54 dph.

2.4.5. Transcriptome profile of the endocrine repertoire during larval development

Analysis of the global transcriptomes of gilthead seabream larvae at flexion and mid-metamorphosis (n = 20), identified 181 endocrine transcripts that were expressed in more than 50 % of the samples (**Supplementary table 2.5 A-B**). Of the 100 endocrine genes (**Supplementary table 5 C**) with the highest level of expression during flexion and mid-metamorphosis, few (16) were DEGs in comparisons I and II (**Figure 2.8**).

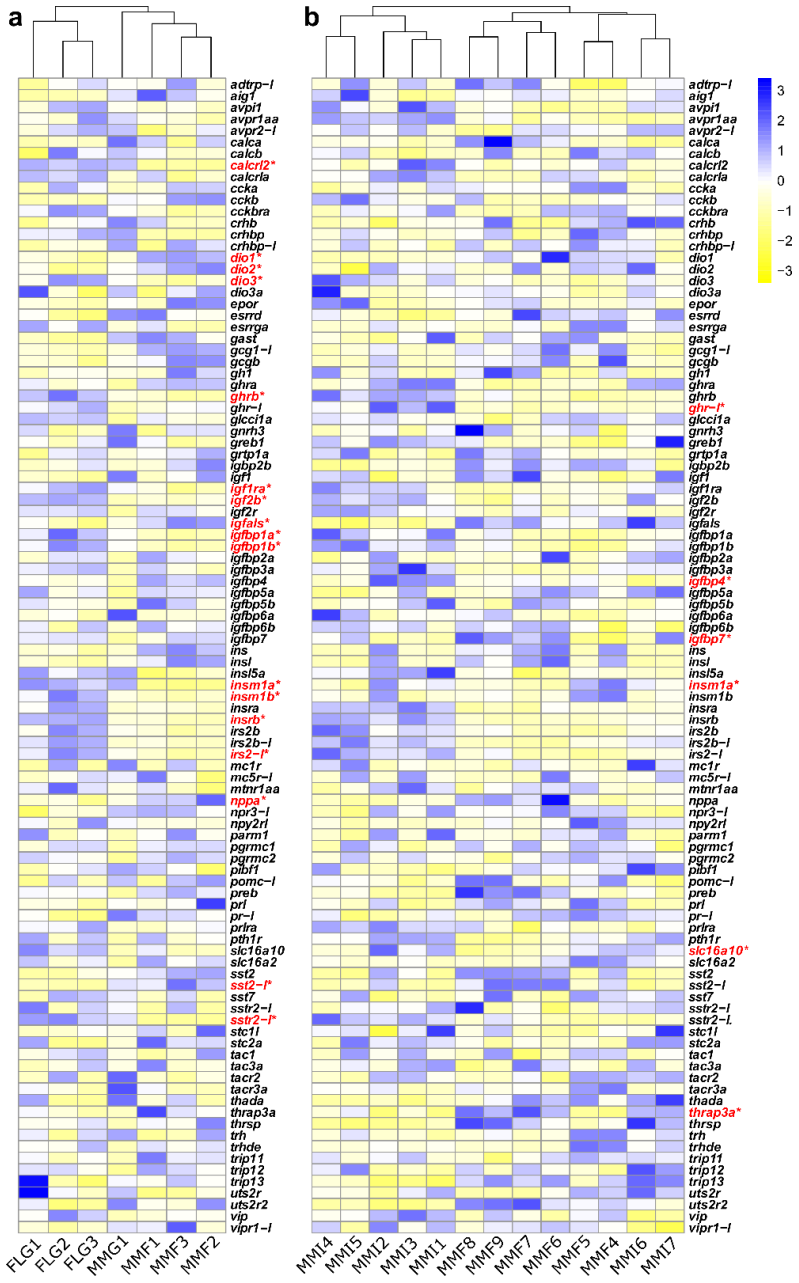


Figure 2.8. Heatmap of the top 100 endocrine-associated genes identified in gilthead seabream transcriptomes of larvae during metamorphosis. a) the gene transcripts that were

significantly different in larvae at 24 dph compared to larvae at 51 dph (comparison I) are indicated in red text and marked with an asterisk (*), and b) the gene transcripts that were significantly different in larvae at 46 dph compared to 56 dph (comparison II) are indicated in red text and marked with an asterisk (*).

The endocrine related genes differentially expressed between stages belonged to the GH/IGF axis and thyroid axis. Among the members of the GH/IGF axis, somatostatin-2-like (*sst2-l*) was low abundance and somatostatin receptor type 2-like (*sstr2-l*) was more abundant in larvae at 24 dph compared to larvae at 50 dph (**Figure 2.8**). Several igf related genes - *igfbp1a*, *igfbp1b*, *igf1ra*, i*gf2b* - and *ghrb* were significantly more abundant in larvae at flexion compared to larvae at mid-metamorphosis (**Figure 2.8**). In comparison II, *igfbp7* was down-regulated and *igfbp4* was up-regulated (**Figure 2.8**). Overall, most of the IGFBPs (including *igfbp2a*, *igfbp2b*, *igfbp3a*, *igfbp3b*, *igfbp5a*, *igfbp5b*, *igfbp6a*, *igfbp6b*) had a constant expression pattern in early development, but generally one of the isoforms of the duplicated genes had a higher expression level (*igfbp2a*, *igfbp5b*, *igfbp3a*, *igfbp6b*) during ontogeny. The IGFBPs with the highest level of expression was *igfbp5b* and *igfbp5a*, and the least expressed genes were *igfbp6a* and *igbp3b*.

In relation to the thyroid axis thyroxine 5-deiodinase-like (*Dio3*) was significantly up-regulated while iodothyronine deiodinase 1 (*Dio 1*) and iodothyronine deiodinase 2 (*Dio 2*) were low abundance in larvae at flexion (24 dph) compared to larvae at mid-metamorphosis (**Figure 2.8**). The expression of thyroid hormone receptor-associated protein 3a (*thrap3a*) was higher in more developed larvae at the mid-metamorphic stage.

Genes involved in neuroendocrine differentiation such as insulinoma-associated 1a (*insm1a*) and insulinoma-associated 1b (*insm1b*) were significantly more abundant in larvae at flexion (24 dph) compared to larvae at mid-metamorphosis (51 dph, **Figure 2.8**). *Insm1a* was also significantly more abundant in larvae at 46 dph compared to larvae at 54 dph (comparison II, **Figure 2.8**).

2.4.6. Association of co-expressed gene modules with age and weight

WGCNA I and II generated 18 gene modules and the genes with high gene significance (GS) or correlation with age and weight were analysed in more detail (**Supplementary figure 2.6** and **Supplementary table 2.6**). The green module (Megreen) of WGCNA for comparison I showed the highest positive correlation with age ($r = 0.86$, $p = 0.01$),

and blue modules (Meblue) of WGCNA comparison I and WGCNA comparison II showed the highest positive correlation with weight (r = 0.88, p =0.01; r = 0.81, p = 9e-04). The functional profile of genes in the most correlated modules with age and weight were further analyzed (Megreen and Meblue of WGCNA I, Meblue of WGCNA II). The ontology of gene sets in each module was specified and GO terms associated with age and weight were identified (**Figure 2.9**). The results showed that the biological processes like, skeletal muscle, organ development, and complement activation were strongly associated with age (**Figure 2.9 a**). Biological processes like, immune response, antigen processing and presentation, oxygen transport, peptide metabolic process, respiratory electron transport chain, and cellular respiration were associated with weight (**Figure 2.9 b**).

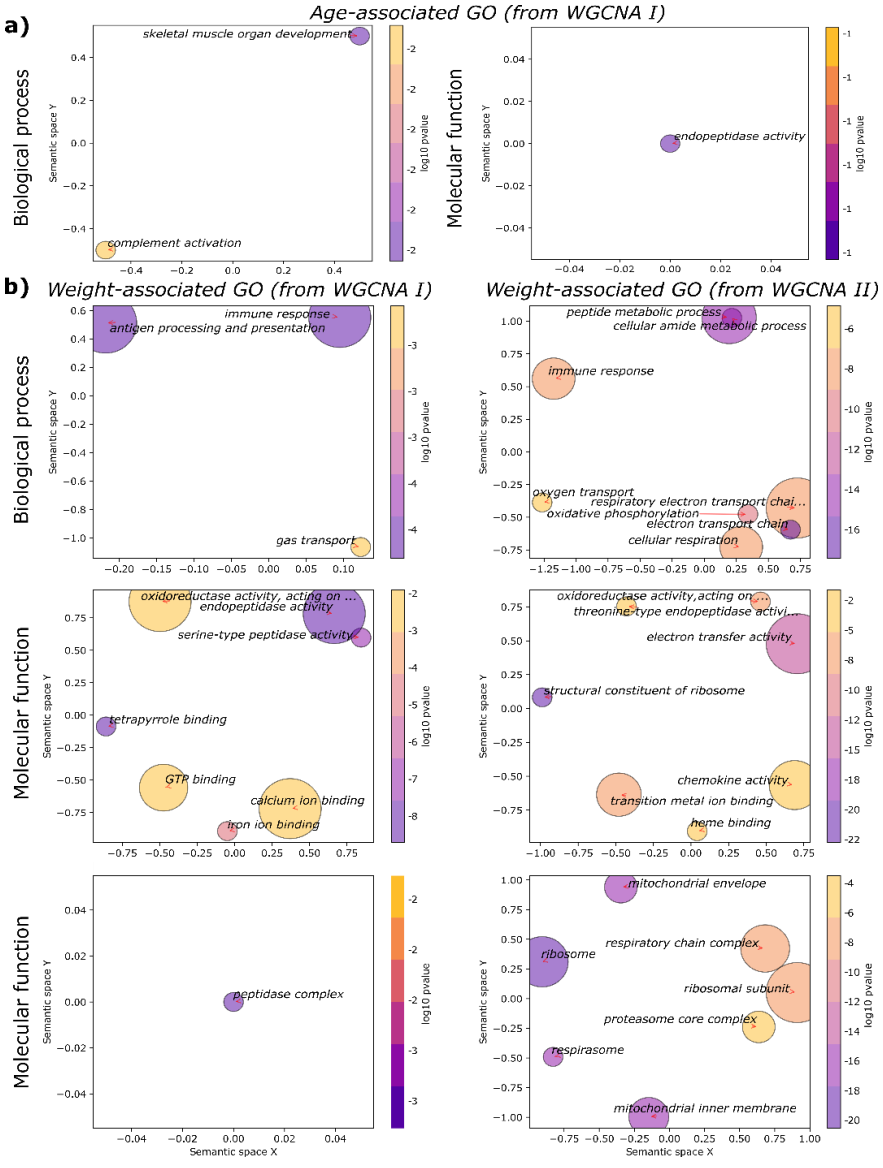


Figure 2.9. Gene ontology (GO) enrichment analysis of co-expressed genes associated with

age (a) and weight (b). The genes corresponding to the green modules that had the highest correlation with age from WGCNA I and the blue modules that had the highest correlation with weight from WGCNA I and WGCNA II were used in the GO analysis (see **Supplementary table 2.6**).

2.4.7. QPCR validation of transcriptome data

The qPCR of six candidate genes (*hbae5, map4, lct, chia, rh50, and tcnba*) selected from the DE genes and that had a high log-fold change (n = 20 samples) were found to have a significant positive correlation with the normalized RNAseq read counts and qPCR amplicons (**Figure 2.10**). Similar gene expression patterns (up- or down-regulation) of DEGs (comparisons I and II) and qPCR amplification of the candidate genes was identified, e.g. increased expression of *hbae5*, *map4*, *tcnba*, and decreased expression of *lct* and *chia* with age (**Supplementary figure 2.7**).

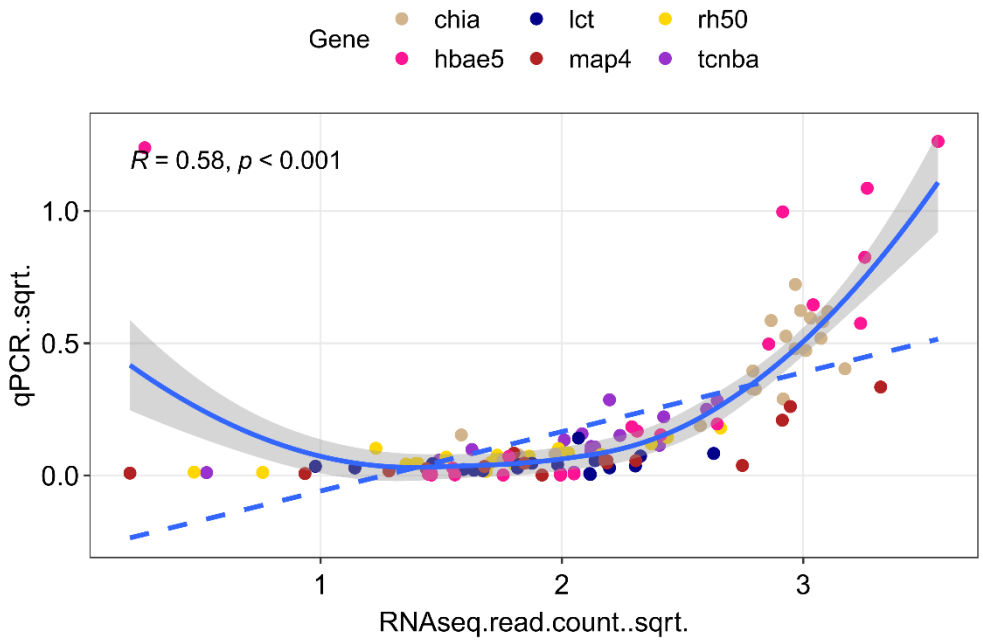


Figure 2.10. Validation of RNA-seq data using a correlation analysis between qPCR amplification and RNA-seq read counts of seven genes. The sqrt of the normalized qPCR and RNA-seq results were used in the analysis. The scatter plots were generated using the "Spearman" method in an R environment. A positive correlation was observed ($P < 0.05$).

2.5. Discussion

By integrating the transcriptomes of larvae from various hatcheries, this study suggests that the larval developmental stage, whether determined by age or weight, is the primary factor driving the transcriptional response during metamorphosis. This finding subsequently contributes to the identification of a core modification in the transcriptional responses of larvae, regardless of their origin.

The shift from larvae to juveniles in fish involves multitude alterations in their physiology, morphology, behavior, and metabolism. This critical stage encompasses the maturation of various vital systems such as the digestive tract, immune response, neuro-muscular skeletal system, skin and pigmentation, visual capabilities, and sexual maturation, ultimately leading to the establishment of a fully functional and independent organism (Manchado et al., 2016). Consistent with prior observation, our analysis revealed an enrichment of differentially expressed gene (DEG) sets associated with diverse biological processes, including development, growth, detoxification, the nervous system, the ubiquitin-proteasome system, metabolism, and phototransduction. Notably, we observed specific enrichments, such as an abundance of genes related to the nervous system or the FoxO signaling pathway during the flexion stage, as well as an enhanced immune response during mid-metamorphosis. Although the expression of the majority of endocrine genes remained stable, the notable modification of specific endocrine-associated genes (such as *dio1*, *dio2*, *cldn1*, *ing4*, *Pou3f4*, and *fgf22*) underscores their significant role in the process of metamorphosis. These findings support and reinforce existing knowledge, shedding light on the molecular mechanisms underlying crucial biological processes and highlighting the stage-specific gene expression dynamics during development and metamorphosis.

Previous transcriptome studies of fish during early development have tended to be based on small-scale or pilot scale production with low sample numbers and a single captive broodstock. Such studies reported the abundance of specific-gene sets, such as genes encoding myogenic proteins in gilthead seabream at hatch (Sarropoulou et al., 2016), homeobox genes in early stages of Atlantic bonito (*Sarda sarda*, Sarropoulou et al., 2014), genes of extracellular matrix/ECM components (e.g., *col1a1*, *col1a2*, *col5a2*, *ddr1*, *mmp13*, *matrix metallopeptidase 2* and *metalloproteinase inhibitor 3*) in flatfish (Alves et al., 2016), and genes of fat digestion and absorption and antigen processing and presentation pathways in the zebrafish mouth-opening stage (Xu et al., 2017). The extensive sampling of two crucial stages of gilthead seabream larvae in our study offers a more comprehensive perspective on the significant

biological processes occurring during metamorphosis compared to previous research. Moreover, we have successfully identified gene sets and markers that demonstrate enrichment during larval development and are consistently observed across multiple commercial hatcheries and production regimes. These findings hold great promise as potential biomarkers for monitoring both the developmental stage and production quality, regardless of variations in the management regime.

2.5.1. Stage-specific modification of gene sets and pathways crucial for morphogenesis and adaptation in larvae

Major changes in gilthead seabream morphology occur between 40 – 123 days post-hatch (dph), and at this time the head region becomes shorter, the trunk grows, and the orientation of the mouth changes and becomes more ventral (Russo et al., 2007). In gilthead seabream, metamorphosis occurs at approximately 45 dph (Manchado et al., 2016). The changes in morphology at metamorphosis are orchestrated by the activation/suppression of specific gene sets or pathways with a well described role in vertebrate development and adaptation such as focal adhesion, ECM receptor interactions (Kjær, 2004; Martins-Green and Bissell, 1995), Wnt signaling (Komiya and Habas, 2008; Logan and Nusse, 2004), the phototransduction pathways (Shimmura et al., 2017), and homeobox genes (Duverger and Morasso, 2008; Hobert and Westphal, 2000). For example, in early stages (flexion) of both gilthead seabream and Senegalese sole DE genes included those of focal adhesion such as *mpk1*, *thbs4b*, *col9a2*, *col9a3*, *arhgap35b*, *itga8*, *thbs1a*, ECM receptor interactions (*frem2a*, *frem2b*, *thbs1a*, *thbs1b*, *itga8*) and biological processes involved in cellular development, fusion and interaction. Changes in these genes are indicative of active cell differentiation, adhesion and migration, all of which are important for the development and remodeling of tissue, hallmarks of early development. DE genes encoding cadherin (*cdh17*, LI cadherin and *cdh15*, M-cadherin), which are important in development of the liver, intestine, and skeletal muscle in mouse and humans (Baumgartner, 2013; Esteves de Lima et al., 2021) were highly expressed in early stages of gilthead seabream larvae (24 and 46 dph). The high expression of homeobox genes in early stages of gilthead seabream larvae (24 and/or 46 dph) is in line with their known role as major regulators of anatomical development, including the hox genes (e.g., *hoxd4b*, *hoxa10*, *arxa*, *six1*), LIM genes (e.g., *lhx4*, *lhx1*, *lhx5*) and POU genes (e.g., *pou3f2a*, *pou3f4*, *pou4f1*, *pou4f2*), which function in brain and sensory structure development (Kitambi and Chandrasekar, 2021; Seufert et al., 2005) and neural patterning (Hobert and Westphal, 2000).

Feeding regime is crucial and determines production quality and needs to be adapted to the morphology, physiology, and nutritional needs of the larvae. In this context the DE genes identified in the gilthead seabream during metamorphosis give insight into important core metabolic processes that should be considered in the rearing regime. High expression levels of apolipoprotein genes were identified in 46 dph gilthead seabream larvae as well as upregulation of genes (e.g., *mgat*) related to lipid digestion and absorption and triacylglycerol (TAG) formation (*dgat, agpat, gpat1*) and storage. Interestingly, the histology of gilthead seabream larvae at metamorphosis reported the absorptive cells of the first intestinal segment contained large lipid droplets similar to other fish larvae (Elbal et al., 2004; García Hernández et al., 2001; Rombout et al., 1984). The present results from the transcriptome and histology are indicative of high lipid metabolism potentially due to co-feeding of live and artificial feeds rich in fats as was previously reported (Fontagné et al., 1998; Morais et al., 2007). Functional profiling of the genes involved in lipid metabolism revealed they were mostly up regulated at mid-metamorphosis in larvae with a lower growth rate and this suggests they may be interesting candidate genes for establishing larval quality. This proposal is further supported by previously published studies that noted increasing the lipid content of the diet led to reduced digestion and absorption and did not improve the growth performance of larval gilthead seabream (Bonaldo et al., 2010) or other marine fish (Gawlicka et al., 2002; Izquierdo et al., 2000; Morais et al., 2007; Olsen et al., 2000).

Carbohydrate metabolism is clearly a predominant source of metabolic energy leading up to gilthead seabream metamorphosis since in 24 dph and 46 dph larvae several genes involved in this process, *lct*, *pfkm*, and *pygm*, were up regulated. This suggests a dependence on carbohydrate metabolism in early stages and upregulation of pathways such as glycogenolysis (Migocka-Patrzałek et al., 2020), gluconeogenesis (Reyes-Ramos et al., 2018) and glycolysis (Tomasik and Horton, 2012). This is further supported by the high enrichment of the 'insulin-like growth factor receptor signaling pathway' and 'cellular response to insulin stimulus' in 24 dph larvae together with the higher expression of genes such as *insrb* and *ghrb*, that are both involved in glucose hemostasis. The greater dependence of gilthead seabream larvae on carbohydrate may be linked to the dietary preference of the larvae and their reduced capacity to digest protein as revealed by the absence of a stomach (**Figure 2.11**) and down-regulated expression of enzymes that digest protein (e.g., chymotrypsin-like, trypsin-like, and pepsin-like). The reduced capacity of early larvae to digest and use proteins seems to be a common characteristic of fish larvae (Douglas et al., 2008; Hilerio-Ruiz et al., 2021; Najafpour et al., 2021). The carnivorous preferences of gilthead seabream juveniles (Mata-Sotres et al.,

2016) is echoed by the results of the transcriptome that suggest an improved capacity to digest proteins as revealed by the up-regulation of pepsin, pepsinogen and trypsin and the appearance of a stomach (**Figure 2.11**) after metamorphosis. The increased capacity for protein digestion after metamorphosis is common in carnivorous fish species, such as white seabass (Galaviz et al., 2011), spotted rose snapper (Galaviz et al., 2012), the rainbow trout (*Oncorhynchus mykiss*) and the Caspian brown trout (*Salmo trutta caspius*, Najafpour et al., 2021).

2.5.2. The immune-related gene repertoire during the larval-juvenile transition

Overall, the immune repertoire of teleosts is more complex than other vertebrates because of the acquisition and retention of additional gene copies during the teleost specific whole genome duplication (TSWGD, Glasauer and Neuhauss, 2014) and lineage and species-specific evolution of immune-related genes (e.g., the complement genes, Najafpour et al., 2020). Studies of immune-related molecules during larval development have generally been based on a candidate gene approach, but little information is available about the acquisition of immune capacity in fish larvae. For example, in the Atlantic cod a study of IgM revealed it was absent until 8–10 weeks post-hatch (Schrøder et al., 1998), which suggests the immune repertoire and immune response differs between juvenile and adult fish. In zebrafish, the alternative pathway of the complement system is proposed to be activated in response to LPS in larval stages (Wang et al., 2008). The general mantra that teleost larvae have a less-developed adaptive immune response compared to the innate immune response (Castro et al., 2015; DeWitte-Orr et al., 2019; Vadstein et al., 2013) means studies have focused primarily on the innate immune system in larvae. Consequently most studies and experimental strategies in larvae have targeted the innate immune response and immune training in intensive-aquaculture of fish larvae using immunostimulants or molecules simulating pathogen-associated molecular patterns (PAMPs) such as ß-glucans, bacterial products, and plant constituents (Abarike et al., 2019; Bricknell and Dalmo, 2005; Citarasu, 2010; Zhang et al., 2019).

The detailed analysis of immune-related genes in the present transcriptome study contributes to improving understanding of the ontogeny of immunity in developing fish larvae. Many innate and adaptive immune-associated genes (e.g., complement, MHC, interferons and interleukins) were down-regulated in pre-metamorphic larvae (24 and 46 dph) compared to post-metamorphic larvae (51 – 54 dph). The significant changes in immune-related gene transcripts between larvae at flexion and mid-metamorphosis indicates a profound change in the immune capacity of gilthead seabream larvae during development **(Figure 2.6**). The down-

regulation of complement associated gene transcripts (e.g., *c1qa*, *c1qb*, *c1qc*, *mr1*) in gilthead seabream at flexion is similar to the reduced expression of complement system genes (C3, C1r/s, C4, Bf, MBL and MASP) in zebrafish larvae after hatching (Wang et al., 2008). Lower immune competence in early life stages was also proposed for the Indian major carp, *Labeo rohita*, which had low mean antibody levels up to 3-weeks post-hatch (Swain et al., 2006). However, studies with Atlantic cod (*Gadus morhua*) indicate the onset of innate immune gene expression at hatch and first feeding (Seppola et al., 2009). Overall, our transcriptome data agrees with that of the cod with the expression of many gene transcripts related to the immune system at flexion, although expression levels are very low, and if this is synonymous with the the functional capacity of the system remains to be established. The generally low expression of immune related gene transcripts in early larval stages of gilthead seabream reflects the global developmental status of the immune system, and corroborates histological studies of hemopoietic tissue and lymphoid organs, which develop from flexion to mid-metamorphosis (Jósefsson and Tatner, 1993). A similar situation was reported in the rock bream (*Oplegnathus fasciatus*) with the pronephric kidney (10 dph) developing first and then the thymus (15 dph) and spleen (21 dph, Xiao et al., 2013). The results of our gene transcription study corroborate previous studies and indicates that immune-related gene expression achieved significantly higher expression at mid-metamorphosis compared to earlier stages, and suggested comprehensive immune competence is acquired in late larval developmental stages.

An interesting facet of the immune-related DE genes was linked to their identity and significant change in expression of Pattern Recognition Receptor (PRR) genes were observed. PRR proteins enable organisms to recognize molecules frequently found in pathogens (Pathogen-Associated Molecular Patterns-PAMPs, Amarante-Mendes et al., 2018). The results of the gilthead seabream transcriptome analysis suggest, that contrary to the accepted dogma, the innate immune repertoire during fish development undergoes maturation from less developed in early larvae to well-developed in older larvae. Although numerous PPRs were expressed they were much less abundant in larvae at flexion compared to the larvae at metamorphosis. In general the abundance of toll-like receptors (TLR) was not significantly changed during larval development, but lectin-like PRRs and galactose-specific lectin nattectin-like (involved in neutrophil mobilization in mice, Lopes-Ferreira et al., 2011), were more abundant in metamorphic stages. Further work will be required to better characterize innate immune maturation and immunocompetence of fish larvae and potential factors that influence immune system development and the immune response.

2.5.3. Common molecular mechanisms during symmetric and asymmetric fish metamorphosis

The meta-analysis of transcriptomes from metamorphosing flatfish (Senegalese sole) and round fish (gilthead seabream) uncovered a shared set of conserved molecular modifications, despite the substantial differences in their external morphology and ecological adaptations, including significant modification in 'endopeptidase activity', 'peptidase inhibitor activity', 'multicellular organism development', and 'lysosomes'. This suggests the existence of fundamental regulatory processes that underlie metamorphosis across diverse fish species, transcending their apparent morphological and ecological disparities. Furthermore, the comparison of the Senegalese sole and gilthead seabream transcriptomes during metamorphosis suggested common molecular mechanisms underlying organ and tissue development and maturation during metamorphosis in these two species. Many genes involved in nervous system development (e.g., neuron differentiation and development) were active in both gilthead seabream and Senegalese sole, especially at an early phase of the metamorphosis transition. These results further support the extensive neurogenesis and axonogenesis, and neuron projection morphogenesis during the metamorphosis of fish that undergo external development compared to mammals (Schmidt et al., 2013). Similar expression patterns of DE genes involved in the development of the brain, head (e.g., *gbx2*, *six3a*, *fgf22*, *pou3f2*, *dcc*, *pou3f3b*), eye (e.g., *crygm2b*, *six3a*, *cryaa*) and epithelium (e.g., *fgf22*, *gata6*) were identified in both Senegalese sole and gilthead seabream.

Comparison of the Senegalese sole and gilthead seabream transcriptome during metamorphosis revealed common genes sets activated or suppressed with age in both species. Examples of activated genes and processes were the liver and hepatobiliary system development (e.g., *gata6*, *arf6b*, *cldn15lb*, *anxa1a*), circulatory system and vasculature development (e.g., *gata6*, *fbn2b*, *myo5b*, *crip1*, *notch1b*, *padi2*), muscle hyperplasia (e.g., *mustn1b*, *myo5b*, *tgif1*), immune system development (*ccl25b*, *notch1b*), thyroid hormone metabolic processes (e.g., *dio1*, *dio2*), and bicellular tight junction (e.g., *cldn15lb*, *cldne*, *cldnb*, *cldna*, *cldnk*). Examples of suppressed genes and processes were mostly involved in brain and eye development (*gbx2*, *six3a*, *pou3f2*, *pou3f3b*, *foxb1a*, *sox11a*, *six7*, *pou4f2*, *tgif1*, *tal2*, *sox4a*, *olfm1a*) and overall transcription factor activity declined significantly from 9 to 24 dph in Senegalese sole and 20 and 51 dph in the gilthead seabream. However, it should be noted that analysis based on pools of whole larvae has limited resolution and likely masks important changes in gene expression since it does not consider, the developmental status and relative contribution of specific organs and tissues to the whole-body transcriptome. To establish the

timing of important developmental processes and their responsiveness to hormones it will be important in the future to conduct tissue specific gene transcript studies.

The current study focused on common DE genes with similar expression patterns in the Senegalese sole and gilthead seabream and identified core and essential developmental changes shared in these fish larvae. Another group of genes were identified with an inverse expression pattern between Senegalese sole and gilthead seabream, which we hypothesize are linked to the unique characteristics of flatfish metamorphosis (asymmetry and ecology driven) since they included gene sets involved in melanin biosynthesis (e.g., *tyrp1b*, *qdpra*, *dct*) and visual phototransduction (*gnb1b*, *pde6gb*, *gnat1*, *bco2b*, *cnga1b*, *rpe65a*), both of which profoundly change during asymmetry acquisition in flatfish.

2.5.4. Modulation of the larval - juvenile transition by the endocrine repertoire

Overall, a stable expression pattern of most of endocrine-associated genes was identified during gilthead seabream development. This is unsurprising when the dynamic nature of endocrine systems and the associated gene expression is considered along with the relatively small contribution of endocrine-related gene transcripts to whole larvae transcriptomes. Nonetheless, several genes encoding endocrine factors were well-correlated with weight (WGCNA analysis) and specific gene isoforms of the GH-IGF and thyroid axis were associated with the larval developmental status such as activation of *igfbp7*, *igfals*, *dio1* in more developed larvae and *igfbp1a*, *dio3*, *igfbp4* in less developed larvae. The variation in the expression level of the endocrine-associated genes such as IGFBP isoforms may reflect their tissue-specific function in growth and development. In line with our results and interpretation, where a high level of *igfbp5b* was interpreted as the result of its expression in different organs, broad tissue distribution of *igfbp5b* was observed in other fish (e.g., zebrafish), while *igfbp5a* had a more specific tissue distribution (e.g., high levels in brain and gill of zebrafish, Dai et al., 2010; de la Serrana and Macqueen, 2018). Interestingly we found that genes of the GH-IGF system were the most significantly changed during metamorphosis, and this merits further attention in the future. *Igf1ra* expression pattern (up-regulated at flexion compared to mid-metamorphosis gilthead seabream larvae) is the other example that suggests the importance of the GH - IGF I system during gilthead seabream larvae metamorphosis, especially where the other associated gene (*igf1rb*) was absent or expressed at a low level. In line with our results, the significant changes in gene expression involved in the GH–IGF-I pathway (including IGF-IR and GHR transcripts) during the Atlantic halibut metamorphosis further highlighted the substantial role

of the GH-IGF I system for successful development (Hildahl et al., 2008, 2007).

Furthermore, the importance of the GH/IGF axis for immune system development needs to be determined particularly since in adult fish there is evidence that it modulates immune cell proliferation, differentiation and function (Calduch-Giner et al., 1995; Franz et al., 2016; Kajita et al., 1992; Yada, 2007).

The significant change in genes belonging to iodothyronine deiodinase family (*dio1*, *dio2*, *dio3*), in gilthead seabream during the flexion to the mid-metamorphosis transition, echoes what is seen in flatfish metamorphosis and highlights the importance of the thyroid axis in fish metamorphosis in general. Interestingly in mammals and some fish (e.g., zebrafish) *DIO3,* which regulates thyroid hormone inactivation is suggested to delay growth and has the highest expression levels in early stages (Heijlen et al., 2014; Luongo et al., 2019). Similarly, in gilthead seabream at flexion *dio3* was significantly up-regulated at flexion (24 dph), while expression of *dio1* and *dio2* were significantly increased at mid-metamorphosis (51 dph), suggesting a higher rate of thyroid hormone activation by conversion of T4 to T3 at mid-metamorphosis. The deiodinase genes (*dio1* and *dio2*) expression started to increase at 20–30 dph (beginning of metamorphosis) and reached a peak at about 45 dph in a previous study using the RT-PCR approach in gilthead seabream (Campinho et al., 2010), corroborating the current results. The enzymes *dio2* and *dio3* were suggested as two key enzymes that mediate flatfish metamorphosis (Isorna et al., 2009). This proposal is in line with the findings of this transcriptomic study and previous studies using qPCR of candidate genes in gilthead seabream, a symmetric and pelagic fish.

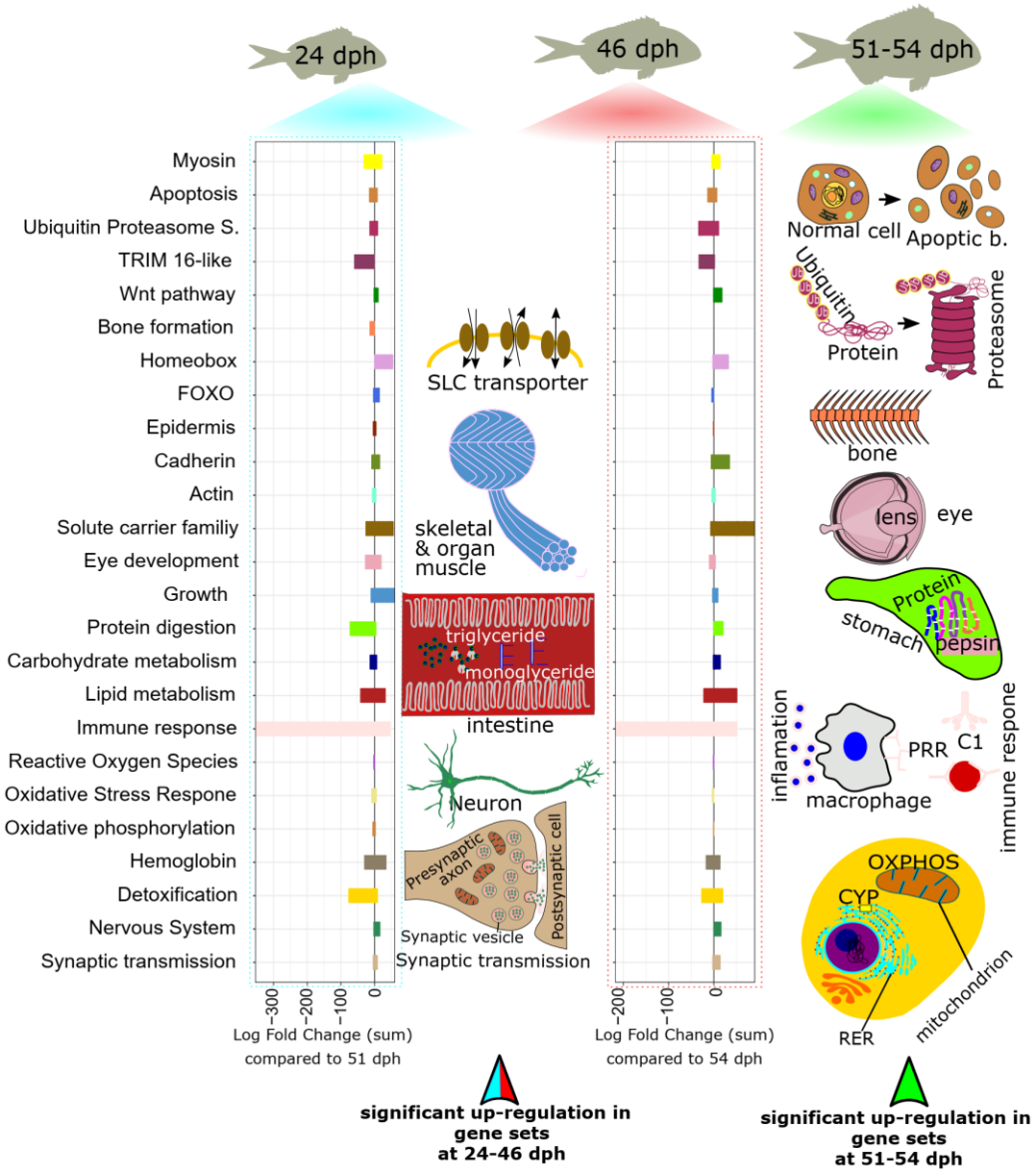


Figure 2.11. Summary of the transcriptome changes of the gilthead seabream larvae. The sum of log fold changes of the gene sets with diverse biological roles showing up- and down-regulation of expression at different larval stages (24, 46 dph) compared to older larvae (51-54 dph). The schematic synthesises the main changes in the transcriptome during the developmental stages studied. In the case of 24-46 dph larvae based on the analysis of DEGs and gene sets more activated processes compared to older larvae were skeletal and muscle development and nervous system development, while in 51-54 dph larvae DEGs and gene sets involved in the immune response were more activated than in younger larvae.

2.6. *Summary*

The principal component analysis (PCA) using different larvae samples and designing differential expression analyses based on clustering and distances between larvae samples provided unbiased gene expression profiles at each developmental stage by removing outlier samples and site-specific effects. In addition, the association of co-expressed gene modules with age and weight provided age and quality-related gene markers.

This study provided insight into core transcriptome modifications of gilthead seabream larvae during development from flexion to mid-metamorphosis under hatchery conditions. The most abundant and common differentially expressed genes during larval ontogeny were involved in development, growth, detoxification, the nervous system, the ubiquitin-proteasome system, metabolism, phototransduction, and the immune system (**Figure 2.11**). The stage-specific activation or suppression of gene sets in larvae from different sites was determined and identified robust markers for morphogenesis and adaptation in larvae. Gene set enrichment analysis suggested the enrichment of different GOs at each stage, including nervous system development at the flexion stage and pre-metamorphosis (46 dph) or muscle system process, immune system process, and proteasomal ubiquitin-independent protein catabolic process at mid-metamorphosis (51 – 54 dph). Functional profiling of the genes involved in lipid metabolism revealed they were mostly up-regulated at mid-metamorphosis in larvae with a lower growth rate and this suggests they may be interesting candidate genes for establishing larval quality. Carbohydrate metabolism is a predominant source of metabolic energy leading up to gilthead seabream metamorphosis from the pre-metamorphosis stage. This was supported by the high enrichment of the 'insulin-like growth factor receptor signaling pathway' and 'cellular response to insulin stimulus' in 24 dph larvae. The comparison of the Senegalese sole (asymmetric and benthic) and gilthead seabream (symmetric and pelagic) transcriptomes during metamorphosis suggested common molecular mechanisms underlying organ and tissue development and maturation during metamorphosis. In contrast, genes determining body pigmentation and eye symmetry had an inverse expression pattern in gilthead seabream and sole presumably reflecting gene sets important in the unique metamorphosis of flatfish.

2.7. See ANNEX II for Supplementary materials

Supplementary table 2.1. The list of significant 2166 DEG transcripts ($p < 0.05$) detected when larval gilthead seabream of 24 dph were compared to 51 dph. This is only available in digital format in Annex II because the table is very extensive.

Supplementary table 2.2. The list of significant 2205 DEGs transcripts ($p < 0.05$) detected when larval gilthead seabream of 46 dph were compared to 54 dph. This is only available in digital format in Annex II because the table is very extensive.

Supplementary table 2.3. A list of the DE gene transcripts obtained from the comparison of 24 dph and 46 dph gilthead seabream and 9 dph compared to 20 dph Senegalese sole. Gene transcripts and their functional profiles are listed. This is only available in digital format in Annex II because the table is very extensive.

Supplementary table 2.4. The pattern recognition receptor (PRR) gene transcripts identified in the transcriptomes and their expression profile during gilthead seabream larval development. This is only available in digital format in Annex II because the table is very extensive.

Supplementary table 2.5. The endocrine-associated gene transcripts identified in the transcriptomes and their expression profile during gilthead seabream larval development. This is only available in digital format in Annex II because the table is very extensive.

Supplementary table 2.6. Identification of the genes associated with the physical traits age and weight and the important modules using weighted gene co-expression network analysis (WGCNA). This is only available in digital format in Annex II because the table is very extensive.

Supplementary figure 2.1. Gene Set Enrichment (GSE) analysis at the biological process level (p value < 0.05). Down- and up-regulated genes involved in biological processes are specified based on DEGs obtained from: **a**) comparison I between younger (avg. 24 dph) and older (avg. 51 dph) gilthead seabream larvae and **b**) comparison II between 46 dph and 54 dph gilthead

seabream larvae. Gene transcripts with a significant change in were retrieved from NCBI and their orthologues identified by local blast against Ensembl protein data of zebrafish (Danio_rerio.GRCz11.pep.all.fa.gz). The bioconductor release of the genome annotation for zebrafish ("org.Dr.eg.db") was used as the background in the GSE analysis in an R environment. Asterisks (*) indicates GO terms with the p-adjusted method of Benjamini–Hochberg (BH) < 0.05. This is only available in digital format in Annex II.

Supplementary figure 2.2. Gene Set Enrichment (GSE) analysis at the levels of molecular function (p value < 0.05). Down- and up-regulated genes involved in molecular function are identified based on the DEGs retrieved from: **a**) comparison I between younger (avg. 24 dph) and older (avg. 51 dph) gilthead seabream larvae and **b**) comparison II between 46 dph and 54 dph gilthead seabream larvae. Gene transcripts with a significant change in were retrieved from NCBI and their orthologues identified by local blast against Ensembl protein data of zebrafish (Danio_rerio.GRCz11.pep.all.fa.gz). The bioconductor release of the genome annotation for zebrafish ("org.Dr.eg.db") was used as the background in the GSE analysis in an R environment. Asterisks (*) indicates GO terms with the p-adjusted method of Benjamini–Hochberg (BH) < 0.05. This is only available in digital format in Annex II.

Supplementary figure 2.3. Gene Set Enrichment (GSE) analysis at the levels of the cellular component (p value < 0.05). Down- and up-regulated genes involved classified within cellular component are specified based on the DEGs obtained from: a) comparison I between younger (avg. 24 dph) vs older (avg. 51 dph) gilthead seabream larvae and b) comparison II between 46 dph and 54 dph gilthead seabream larvae. Gene transcripts with a significant change in were retrieved from NCBI and their orthologues identified by local blast against Ensembl protein data of zebrafish (Danio_rerio.GRCz11.pep.all.fa.gz). The bioconductor release of the genome annotation for zebrafish ("org.Dr.eg.db") was used as the background in the GSE analysis in an R environment. Asterisks (*) indicates GO terms with the p-adjusted method of Benjamini–Hochberg (BH) < 0.05. This is only available in digital format in Annex II.

Supplementary figure 2.4. KEGG Gene Set Enrichment analysis (p value < 0.05). Enriched KEGG pathways that were activated (N = 8) or suppressed (N = 11) were identified. The approach taken for KEGG pathway analysis was to use the DEGs retrieved by comparing the transcriptomes of younger (avg. 24 dph) and older (avg. 51 dph) gilthead seabream larvae to

retrieve 1606 zebrafish orthologues. The zebrafish orthologues of the DE genes were identified by local blast and the full-length protein retrieved from the protein translations of Ensembl genes (Danio_rerio.GRCz11.pep.all.fa.gz). Then the zebrafish KEGG database ("dre") was used for KEGG pathway analysis of the gilthead seabream DEGs in an R environment. The asterisks (*) denotes pathways with the p-adjusted method of Benjamini–Hochberg (BH) < 0.05. This is only available in digital format in Annex II.

Supplementary figure 2.5. KEGG Gene Set Enrichment analysis (p value < 0.05). Enriched KEGG pathways that were activated (N = 14) or suppressed (N = 12) were identified using the DEGs identified from the comparison of 46 dph vs 54 dph gilthead seabream larvae to retrieve 1738 zebrafish orthologues. The zebrafish orthologues of the DE genes were identified by local blast and the full-length protein retrieved from the protein translations of Ensembl genes (Danio_rerio.GRCz11.pep.all.fa.gz). Then the zebrafish KEGG database ("dre") was used for KEGG pathway analysis of the identified gilthead seabream DEGs in an R environment. The asterisks (*) denotes pathways with the p-adjusted method of Benjamini–Hochberg (BH) < 0.05. This is only available in digital format in Annex II.

Supplementary figure 2.6. Gene modules and their association with larvae age and weight traits. Each row corresponds to a module eigengene, and each column presents the associated trait. Each cell contains the corresponding correlation and p-value. The columns of the table (Age_dph and Weight_mg) are colour-coded to give a visual representation of the intensity of the correlation obtained (intense green -1 and intense red +1). **a**) transcriptome data of seven samples with a similar range of reads were used as the input for weighted gene co-expression network analysis (WGCNA), the sample codes are provided in figure 2b and marked as comparison I, **b**) The expression data of 13 samples with a similar range of reads were used as the input for weighted gene co-expression network analysis (WGCNA), the sample codes are provided in **Figure 2.2 b** and marked as comparison II. The names and accession number of the co-expressed genes giving rise to each module are provided in Supplementary table 2.6.

Supplementary figure 2.7. QPCR amplification of seven genes in gilthead seabream at different developmental stages. For each gene, part **a** represents the results for larvae samples that were used in comparison I (flexion, 24 dph vs mid-metamorphosis, 51 dph, for more details see **Figure 2.2**), and part **b** represents the results for larvae samples that were used in

comparison II (46 vs 54 dph larvae at mid-metamorphosis; for more details see **Figure 2.2**). COMI = comparison I, COMII = comparison II. Similar expression patterns (down-regulated or up-regulated) that were specified in comparison I and II were also identified by qPCR amplification of the selected genes. This is only available in digital format in Annex II.

2.8. References

Abarike, E.D., Kuebutornye, F.K.A., Jian, J., Tang, J., Lu, Y., Cai, J., 2019. Influences of immunostimulants on phagocytes in cultured fish: a mini review. Rev. Aquac. 11, 1219–1227. https://doi.org/10.1111/RAQ.12288

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., 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. https://doi.org/10.1093/NAR/GKY379

Alves, R.N., Gomes, A.S., Stueber, K., Tine, M., Thorne, M.A.S., Sráradóttir, H., Reinhard, R., Clark, M.S., Rønnestad, I., Power, D.M., 2016. The transcriptome of metamorphosing flatfish. BMC Genomics 17, 1–24. https://doi.org/10.1186/s12864-016-2699-x

Amarante-Mendes, G.P., Adjemian, S., Branco, L.M., Zanetti, L.C., Weinlich, R., Bortoluci, K.R., 2018. Pattern recognition receptors and the host cell death molecular machinery. Front. Immunol. 9, 2379. https://doi.org/10.3389/FIMMU.2018.02379

Andrews, S., 2010. Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data.

Baumgartner, W., 2013. Possible roles of LI-Cadherin in the formation and maintenance of the intestinal epithelial barrier. Tissue Barriers 1, e23815. https://doi.org/10.4161/TISB.23815

Benzekri, H., Armesto, P., Cousin, X., Rovira, M., Crespo, D., Merlo, M.A., Mazurais, D., Bautista, R., Guerrero-Fernández, D., Fernandez-Pozo, N., Ponce, M., Infante, C., Zambonino, J.L., Nidelet, S., Gut, M., Rebordinos, L., Planas, J. V., Bégout, M.L., Claros, M.G., Manchado, M., 2014. De novo assembly, characterization and functional annotation of Senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: Integration in a database and design of a microarray. BMC Genomics 15, 1–18. https://doi.org/10.1186/1471-2164-15-952/TABLES/4

Bonaldo, A., Isani, G., Fontanillas, R., Parma, L., Grilli, E., Gatta, P.P., 2010. Growth and feed utilization of gilthead sea bream (*Sparus aurata*, L.) fed to satiation and restrictively at increasing dietary energy levels. Aquac. Int. 18, 909–919. https://doi.org/10.1007/S10499-009-9312-0/TABLES/4

Bricknell, I., Dalmo, R.A., 2005. The use of immunostimulants in fish larval aquaculture. Fish Shellfish Immunol. 19, 457–472. https://doi.org/10.1016/J.FSI.2005.03.008

Brudeseth, B.E., Wiulsrød, R., Fredriksen, B.N., Lindmo, K., Løkling, K.E., Bordevik, M., Steine, N., Klevan, A., Gravningen, K., 2013. Status and future perspectives of vaccines for industrialised fin-fish farming. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2013.05.029

Brugman, S., Ikeda-Ohtsubo, W., Braber, S., Folkerts, G., Pieterse, C.M.J., Bakker, P.A.H.M., 2018. A comparative review on microbiota manipulation: lessons from fish, plants, livestock, and human research. Front. Nutr. 5, 80. https://doi.org/10.3389/FNUT.2018.00080/BIBTEX

Burggren, W., O'callaghan, C., Finne, J., Torday, J.S., 2016. Epigenetic inheritance and its role

in evolutionary biology: re-evaluation and new perspectives. Biol. 5, 24. https://doi.org/10.3390/BIOLOGY5020024

Burggren, W.W., 2020. Phenotypic switching resulting from developmental plasticity: fixed or reversible? Front. Physiol. 10, 1634. https://doi.org/10.3389/FPHYS.2019.01634/XML/NLM

Calduch-Giner, J.A., Bermejo-Nogales, A., Benedito-Palos, L., Estensoro, I., Ballester-Lozano, G., Sitjà-Bobadilla, A., Pérez-Sánchez, J., 2013. Deep sequencing for de novo construction of a marine fish (*Sparus aurata*) transcriptome database with a large coverage of protein-coding transcripts. BMC Genomics 14, 1–11. https://doi.org/10.1186/1471-2164-14-178/TABLES/3

Calduch-Giner, J.A., Sitja-Bobadilla, A., Alvarez-Pellitero, P., Perez-Sanchez, J., 1995. Evidence for a direct action of GH on haemopoietic cells of a marine fish, the gilthead sea bream (*Sparus aurata*). J. Endocrinol. 146, 459–467. https://doi.org/10.1677/JOE.0.1460459

Campinho, M.A., Galay-Burgos, M., Sweeney, G.E., Power, D.M., 2010. Coordination of deiodinase and thyroid hormone receptor expression during the larval to juvenile transition in sea bream (*Sparus aurata*, Linnaeus). Gen. Comp. Endocrinol. 165, 181–194. https://doi.org/10.1016/J.YGCEN.2009.06.020

Cara, J.B., Aluru, N., Moyano, F.J., Vijayan, M.M., 2005. Food-deprivation induces HSP70 and HSP90 protein expression in larval gilthead sea bream and rainbow trout. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 142, 426–431. https://doi.org/10.1016/j.cbpb.2005.09.005

Castro, R., Jouneau, L., Tacchi, L., Macqueen, D.J., Alzaid, A., Secombes, C.J., Martin, S.A.M., Boudinot, P., 2015. Disparate developmental patterns of immune responses to bacterial and viral infections in fish. Sci. Reports 51, 1–13. https://doi.org/10.1038/srep15458

Chakraborty, S.B., Horn, P., Hancz, C., 2014. Application of phytochemicals as growth-promoters and endocrine modulators in fish culture. Rev. Aquac. 6, 1–19. https://doi.org/10.1111/raq.12021

Citarasu, T., 2010. Herbal biomedicines: a new opportunity for aquaculture industry. Aquac. Int. 18, 403–414. https://doi.org/10.1007/S10499-009-9253-7

Claramunt, R.M., Wahl, D.H., 2000. The effects of abiotic and biotic factors in determining larval fish growth rates: a comparison across species and reservoirs. Trans. Am. Fish. Soc. 129, 835–851. https://doi.org/10.1577/1548-8659

Claros, M.G., Pedro, S., Manchado, M., 2020. Sequences and annotations of a provisional genome draft of a Senegalese sole female (Sosen1) and a male (Sse05_10M). figshare. Dataset. https://doi.org/10.6084/m9.figshare.12472100.v3.

Cnaani, A., 2006. Genetic perspective on stress response and disease resistance in aquaculture. Isr. J. Aquac. 58, 375–383. https://doi.org/hdl.handle.net/10524/19199

Dai, W., Kamei, H., Zhao, Y., Ding, J., Du, Z., Duan, C., 2010. Duplicated zebrafish insulin-like growth factor binding protein-5 genes with split functional domains: evidence for evolutionarily conserved IGF binding, nuclear localization, and transactivation activity.

FASEB J. 24, 2020–2029. https://doi.org/10.1096/FJ.09-149435

Darias, M.J., Zambonino-Infante, J.L., Hugot, K., Cahu, C.L., Mazurais, D., 2008. Gene expression patterns during the larval development of European sea bass (*dicentrarchus labrax*) by microarray analysis. Mar. Biotechnol. (NY). 10, 416–428. https://doi.org/10.1007/S10126-007-9078-1

Garcia de la Serrana, D.G., Macqueen, D.J., 2018. Insulin-like growth factor-binding proteins of teleost fishes. Front. Endocrinol. (Lausanne). 9, 80. https://doi.org/10.3389/FENDO.2018.00080/BIBTEX

DeWitte-Orr, S., Edholm, E.S., Grayfer, L., 2019. Editorial: innate immunity in aquatic vertebrates. Front. Immunol. 10, 2959. https://doi.org/10.3389/FIMMU.2019.02959/BIBTEX

Douglas, S.E., Knickle, L.C., Williams, J., Flight, R.M., Reith, M.E., 2008. A first generation Atlantic halibut *Hippoglossus hippoglossus* (L.) microarray: application to developmental studies. J. Fish Biol. 72, 2391–2406. https://doi.org/10.1111/J.1095-8649.2008.01861.X

Duverger, O., Morasso, M.I., 2008. Role of homeobox genes in the patterning, specification and differentiation of ectodermal appendages in mammals. J. Cell. Physiol. 216, 337. https://doi.org/10.1002/JCP.21491

Elbal, M.T., García Hernández, M.P., Lozano, M.T., Agulleiro, B., 2004. Development of the digestive tract of gilthead sea bream (*Sparus aurata* L.). Light and electron microscopic studies. Aquaculture 234, 215–238. https://doi.org/10.1016/J.AQUACULTURE.2003.11.028

Esteves de Lima, J., Bou Akar, R., Mansour, M., Rocancourt, D., Buckingham, M., Relaix, F., 2021. M-Cadherin is a PAX3 target during myotome patterning. Front. Cell Dev. Biol. 9, 755. https://doi.org/10.3389/FCELL.2021.652652/BIBTEX

Falk-Petersen, I.B., 2005. Comparative organ differentiation during early life stages of marine fish. Fish Shellfish Immunol. 19, 397–412. https://doi.org/10.1016/J.FSI.2005.03.006

Fernández-Díaz, C., Yýfera, M., Cañavate, J.P., Moyano, F.J., Alarcón, F.J., Díaz, M., 2001. Growth and physiological changes during metamorphosis of Senegal sole reared in the laboratory. J. Fish Biol. 58, 1086–1097. https://doi.org/10.1111/J.1095-8649.2001.TB00557.X

Fontagné, S., Geurden, I., Escaffre, A.M., Bergot, P., 1998. Histological changes induced by dietary phospholipids in intestine and liver of common carp (*Cyprinus carpio* L.) larvae. Aquaculture 161, 213–223. https://doi.org/10.1016/S0044-8486(97)00271-8

Franz, A.C., Faass, O., Kollner, B., Shved, N., Link, K., Casanova, A., Wenger, M., D'cotta, H., Baroiller, J.F., Ullrich, O., Reinecke, M., Eppler, E., 2016. Endocrine and Local IGF-I in the Bony Fish Immune System. Biology (Basel). 5. https://doi.org/10.3390/BIOLOGY5010009

Friedman, S.T., Price, S.A., Corn, K.A., Larouche, O., Martinez, C.M., Wainwright, P.C., 2020. Body shape diversification along the benthic–pelagic axis in marine fishes. Proc. R. Soc. B Biol. Sci. 287, 20201053. https://doi.org/10.1098/RSPB.2020.1053

Galaviz, M.A., García-Gasca, A., Drawbridge, M., Álvarez-González, C.A., López, L.M., 2011. Ontogeny of the digestive tract and enzymatic activity in white seabass,

Atractoscion nobilis, larvae. Aquaculture 318, 162–168. https://doi.org/10.1016/J.AQUACULTURE.2011.05.014

Galaviz, M.A., García-Ortega, A., Gisbert, E., López, L.M., Gasca, A.G., 2012. Expression and activity of trypsin and pepsin during larval development of the spotted rose snapper *Lutjanus guttatus*. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 161, 9–16. https://doi.org/10.1016/J.CBPB.2011.09.001

Garcia de la serrana, D., Estévez, A., Andree, K., Johnston, I.A., 2012. Fast skeletal muscle transcriptome of the Gilthead sea bream (*Sparus aurata*) determined by next generation sequencing. BMC Genomics 13, 1–17. https://doi.org/10.1186/1471-2164-13-181/FIGURES/4

García Hernández, M.P., Lozano, M.T., Elbal, M.T., Agulleiro, B., 2001. Development of the digestive tract of sea bass (*Dicentrarchus labrax* L). Light and electron microscopic studies. Anat. Embryol. 2001 2041 204, 39–57. https://doi.org/10.1007/S004290100173

Gawlicka, A., Herold, M.A., Barrows, F.T., De la Noüe, J., Hung, S.S.O., 2002. Effects of dietary lipids on growth, fatty acid composition, intestinal absorption and hepatic storage in white sturgeon (*Acipenser transmontanus* R.) larvae. J. Appl. Ichthyol. 18, 673–681. https://doi.org/10.1046/J.1439-0426.2002.00371.X

Glasauer, S.M.K., Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics 289, 1045–1060. https://doi.org/10.1007/S00438-014-0889-2

Hampel, M., Blasco, J., Babbucci, M., Ferraresso, S., Bargelloni, L., Milan, M., 2017. Transcriptome analysis of the brain of the sea bream (*Sparus aurata*) after exposure to human pharmaceuticals at realistic environmental concentrations. Mar. Environ. Res. 129, 36–45. https://doi.org/10.1016/J.MARENVRES.2017.04.012

Heijlen, M., Houbrechts, A.M., Bagci, E., Van Herck, S.L.J., Kersseboom, S., Esguerra, C. V., Blust, R., Visser, T.J., Knapen, D., Darras, V.M., 2014. Knockdown of type 3 iodothyronine deiodinase severely perturbs both embryonic and early larval development in zebrafish. Endocrinology 155, 1547–1559. https://doi.org/10.1210/EN.2013-1660

Hildahl, J., Galay-Burgos, M., Sweeney, G., Einarsdóttir, I.E., Björnsson, B.T., 2007. Identification of two isoforms of Atlantic halibut insulin-like growth factor-I receptor genes and quantitative gene expression during metamorphosis. Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 147, 395–401. https://doi.org/10.1016/J.CBPB.2007.02.006

Hildahl, J., Power, D.M., Björnsson, B.T., Einarsdóttir, I.E., 2008. Involvement of growth hormone-insulin-like growth factor I system in cranial remodeling during halibut metamorphosis as indicated by tissue- and stage-specific receptor gene expression and the presence of growth hormone receptor protein. Cell Tissue Res. 332, 211–225. https://doi.org/10.1007/S00441-007-0568-2

Hilerio-Ruiz, A., Toledo-Solís, F.J., Peña, E., Martínez-Burguete, T., Martínez-García, R., Llera-Herrera, R., Álvarez-González, C.A., Saenz de Rodrigáñez, M., 2021. Nutrigenomic markers identified by de novo RNAseq during the early ontogeny of the three spot cichlid *Amphilophus trimaculatus*. Aquaculture 530, 735654. https://doi.org/10.1016/J.AQUACULTURE.2020.735654

Ho, D.H., Burggren, W.W., 2010. Epigenetics and transgenerational transfer: a physiological perspective. J. Exp. Biol. 213, 3–16. https://doi.org/10.1242/JEB.019752

Hobert, O., Westphal, H., 2000. Functions of LIM-homeobox genes. Trends Genet. 16, 75–83. https://doi.org/10.1016/S0168-9525(99)01883-1

Ibeas, C., Izquierdo, M.S., Lorenzo, A., 1994. Effect of different levels of n-3 highly unsaturated fatty acids on growth and fatty acid composition of juvenile gilthead seabream (*Sparus aurata*). Aquaculture 127, 177–188. https://doi.org/10.1016/0044-8486(94)90424-3

Isorna, E., Obregon, M.J., Calvo, R.M., Vázquez, R., Pendón, C., Falcón, J., Muñoz-Cueto, J., 2009. Iodothyronine deiodinases and thyroid hormone receptors regulation during flatfish (*Solea senegalensis*) metamorphosis. J. Exp. Zool. B. Mol. Dev. Evol. 312B, 231–246. https://doi.org/10.1002/JEZ.B.21285

Izquierdo, M.S., Socorro, J., Arantzamendi, L., Hernández-Cruz, C.M., 2000. Recent advances in lipid nutrition in fish larvae. Fish Physiol. Biochem. 22, 97–107. https://doi.org/10.1023/A:1007810506259/METRICS

Jósefsson, S., Tatner, M.F., 1993. Histogenesis of the lymphoid organs in sea bream (*Sparus aurata* L.). Fish Shellfish Immunol. 3, 35–49. https://doi.org/10.1006/FSIM.1993.1004

Kajita, Y., Sakai, M., Kobayashi, M., Kawauchi, H., 1992. Enhancement of non-specific cytotoxic activity of leucocytes in rainbow trout *Oncorhynchus mykiss* injected with growth hormone. Fish Shellfish Immunol. 2, 155–157. https://doi.org/10.1016/S1050-4648(05)80044-0

Kjær, M., 2004. Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading. Physiol. Rev. 84, 649–698. https://doi.org/10.1152/PHYSREV.00031.2003/ASSET/IMAGES/LARGE/Z9J0020403060009.JPEG

Komiya, Y., Habas, R., 2008. Wnt signal transduction pathways. Organogenesis 4, 68. https://doi.org/10.4161/ORG.4.2.5851

Langfelder, P., Horvath, S., 2008. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9, 1–13. https://doi.org/10.1186/1471-2105-9-559/FIGURES/4

Lemon, J., 2006. Plotrix: a package in the red light district of R. R-News 6, 8–12.

Liao, Y., Smyth, G., Shi, W., 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-voteI. Nucleic Acids Res. 41, e108.

Logan, C.Y., Nusse, R., 2004. The Wnt signaling pathway in development and disease. Annu. Rev. Cell Dev. Biol. 20, 781–810. https://doi.org/10.1146/ANNUREV.CELLBIO.20.010403.113126

Lopes-Ferreira, M., Magalhães, G.S., Fernandez, J.H., Junqueira-De-Azevedo, I.D.L.M., Le Ho, P., Lima, C., Valente, R.H., Moura-Da-Silva, A.M., 2011. Structural and biological characterization of Nattectin, a new C-type lectin from the venomous fish *Thalassophryne nattereri*. Biochimie 93, 971–980. https://doi.org/10.1016/J.BIOCHI.2011.03.001

Louro, B., Marques, J.P., Power, D.M., Canário, A.V.M., 2016. Having a BLAST: Searchable transcriptome resources for the gilthead sea bream and the European sea bass. Mar.

Genomics 30, 67–71. https://doi.org/10.1016/j.margen.2016.10.004

Luongo, C., Dentice, M., Salvatore, D., 2019. Deiodinases and their intricate role in thyroid hormone homeostasis. Nat. Rev. Endocrinol. 2019 158 15, 479–488. https://doi.org/10.1038/s41574-019-0218-2

Lyimo, E.O., Takken, W., Koella, J.C., 1992. Effect of rearing temperature and larval density on larval survival, age at pupation and adult size of *Anopheles gambiae*. Entomol. Exp. Appl. 63, 265–271. https://doi.org/10.1111/J.1570-7458.1992.TB01583.X

M, C., 2019. org.Dr.eg.db: Genome wide annotation for Zebrafish. https://doi.org/10.18129/B9.bioc.org.Dr.eg.db

Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L., Claros, M.G., 2016. Current status in other finfish species: Description of current genomic resources for the gilthead seabream (*Sparus aurata*) and soles (*Solea senegalensis* and *Solea solea*). Genomics Aquac. 195–221. https://doi.org/10.1016/B978-0-12-801418-9.00008-1

Martins-Green, M., Bissell, M.J., 1995. Cell-ECM interactions in development. Semin. Dev. Biol. 6, 149–159. https://doi.org/10.1016/S1044-5781(06)80024-1

Martos-Sitcha, J.A., Wunderink, Y.S., Straatjes, J., Skrzynska, A.K., Mancera, J.M., Martínez-Rodríguez, G., 2014. Different stressors induce differential responses of the CRH-stress system in the gilthead sea bream (*Sparus aurata*). Comp. Biochem. Physiol. -Part A Mol. Integr. Physiol. 177, 49–61. https://doi.org/10.1016/j.cbpa.2014.07.021

Mata-Sotres, J.A., Martos-Sitcha, J.A., Astola, A., Yúfera, M., Martínez-Rodríguez, G., 2016. Cloning and molecular ontogeny of digestive enzymes in fed and food-deprived developing gilthead seabream (*Sparus aurata*) larvae. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 191, 53–65. https://doi.org/10.1016/J.CBPB.2015.09.006

Mazurais, D., Darias, M., Zambonino-Infante, J.L., Cahu, C.L., 2011. Transcriptomics for understanding marine fish larval development. Can. J. Zool. https://doi.org/10.1139/z11-036

McMenamin, S.K., Parichy, D.M., 2013. Metamorphosis in Teleosts. Curr. Top. Dev. Biol. 103, 127. https://doi.org/10.1016/B978-0-12-385979-2.00005-8

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., Thomas, P.D., 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14, 703–721. https://doi.org/10.1038/s41596-019-0128-8

Migocka-Patrzałek, M., Lewicka, A., Elias, M., Daczewska, M., 2020. The effect of muscle glycogen phosphorylase (Pygm) knockdown on zebrafish morphology. Int. J. Biochem. Cell Biol. 118, 105658. https://doi.org/10.1016/J.BIOCEL.2019.105658

Mininni, A.N., Milan, M., Ferraresso, S., Petochi, T., Di Marco, P., Marino, G., Livi, S., Romualdi, C., Bargelloni, L., Patarnello, T., 2014. Liver transcriptome analysis in gilthead sea bream upon exposure to low temperature. BMC Genomics 15, 1–12. https://doi.org/10.1186/1471-2164-15-765

Morais, S., Conceição, L.E.C., Rønnestad, I., Koven, W., Cahu, C., Zambonino Infante, J.L., Dinis, M.T., 2007. Dietary neutral lipid level and source in marine fish larvae: Effects on digestive physiology and food intake. Aquaculture 268, 106–122.

https://doi.org/10.1016/J.AQUACULTURE.2007.04.033

Muniesa, A., Basurco, B., Aguilera, C., Furones, D., Reverté, C., Sanjuan-Vilaplana, A., Jansen, M.D., Brun, E., Tavornpanich, S., 2020. Mapping the knowledge of the main diseases affecting sea bass and sea bream in Mediterranean. Transbound. Emerg. Dis. 67, 1089–1100. https://doi.org/10.1111/tbed.13482

Najafpour, B., Dorafshan, S., Heyrati, F.P., Canario, A.V.M., Power, D.M., 2021. Comparative ontogeny of the digestive tract of *Oncorhynchus mykiss* ♀ x *Salmo trutta caspius* ♂ triploid hybrids to their parental species. Aquac. Nutr. 27, 427–438. https://doi.org/10.1111/ANU.13196

Olsen, A.I., Attramadal, Y., Reitan, K.I., Olsen, Y., 2000. Food selection and digestion characteristics of Atlantic halibut (*Hippoglossus hippoglossus*) larvae fed cultivated prey organisms. Aquaculture 181, 293–310. https://doi.org/10.1016/S0044-8486(99)00230-6

Pérez-Sánchez, J., Naya-Català, F., Soriano, B., Piazzon, M.C., Hafez, A., Gabaldón, T., Llorens, C., Sitjà-Bobadilla, A., Calduch-Giner, J.A., 2019. Genome sequencing and transcriptome analysis reveal recent species-specific gene duplications in the plastic gilthead sea bream (*Sparus aurata*). Front. Mar. Sci. 6, 760. https://doi.org/10.3389/FMARS.2019.00760/BIBTEX

Pittman, K., Yúfera, M., Pavlidis, M., Geffen, A.J., Koven, W., Ribeiro, L., Zambonino-Infante, J.L., Tandler, A., 2013. Fantastically plastic: fish larvae equipped for a new world. Rev. Aquac. 5, S224–S267. https://doi.org/10.1111/RAQ.12034

Power, D.M., Einarsdóttir, I.E., Pittman, K., Sweeney, G.E., Hildahl, J., Campinho, M.A., Silva, N., Saele, Ø., Galay-Burgos, M., Sráradóttir, H., Björn, &, Björnsson, T., Einarsd´ottir, E, Einarsd´ottir, Einarsd´, Heidd´, H., Sm´arad´ottir, H., Sm´arad, S., Sm´arad´ottir, S., Bj¨ Bj¨orn, B., Bj¨ornsson, T., Bj¨ornsson, B., 2008. The molecular and endocrine basis of flatfish metamorphosis. Rev. Fish. Sci. 93–109. https://doi.org/10.1080/10641260802325377

Qian, X., Ba, Y., Zhuang, Q., Zhong, G., 2014. RNA-Seq technology and its application in fish transcriptomics. Omi. A J. Integr. Biol. 18, 98–110. https://doi.org/10.1089/omi.2013.0110

Reijnders, M.J., Waterhouse, R.M., 2020. Summary visualisations of gene ontology terms with GO-Figure! Front. Bioinform. 1, 638255. https://doi.org/10.1101/2020.12.02.408534

Reyes-Ramos, C.A., Peregrino-Uriarte, A.B., Cota-Ruiz, K., Valenzuela-Soto, E.M., Leyva-Carrillo, L., Yepiz-Plascencia, G., 2018. Phosphoenolpyruvate carboxykinase cytosolic and mitochondrial isoforms are expressed and active during hypoxia in the white shrimp Litopenaeus vannamei. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 226, 1–9. https://doi.org/10.1016/J.CBPB.2018.08.001

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, N.A., Johnsen, H., Moghadam, H., Andersen, Ø., Tveiten, H., 2019. Early developmental stress affects subsequent gene expression response to an acute stress in atlantic salmon: an approach for creating robust fish for aquaculture? G3 Genes Genomes Genetics 9, 1597–1611. https://doi.org/10.1534/G3.119.400152

Rombout, J.H.W.M., Stroband, H.W.J., Taverne-Thiele, J.J., 1984. Proliferation and differentiation of intestinal epithelial cells during development of *Barbus conchonius* (Teleostei, Cyprinidae). Cell Tissue Res. 236, 207–216. https://doi.org/10.1007/BF00216533

Rønnestad, I., Yúfera, M., Ueberschär, B., Ribeiro, L., Saele, Ø., Boglione, C., 2013. Feeding behaviour and digestive physiology in larval fish: current knowledge, and gaps and bottlenecks in research. Rev. Aquac. 5, S59–S98. https://doi.org/10.1111/raq.12010

Russo, T., Costa, C., Cataudella, S., 2007. Correspondence between shape and feeding habit changes throughout ontogeny of gilthead sea bream *Sparus aurata* L., 1758. J. Fish Biol. 71, 629–656. https://doi.org/10.1111/J.1095-8649.2007.01528.X

Sarropoulou, E., Kotoulas, G., Power, D.M., Geisler, R., 2005. Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. Physiol. Genomics 23, 182–191. https://doi.org/10.1152/physiolgenomics.00139.2005

Sarropoulou, E., Moghadam, H.K., Papandroulakis, N., De La Gándara, F., Garcia, A.O., Makridis, P., 2014. The Atlantic bonito (*Sarda sarda*, Bloch 1793) transcriptome and detection of differential expression during larvae development. PLoS One 9, 87744. https://doi.org/10.1371/journal.pone.0087744

Sarropoulou, E., Tsalafouta, A., Sundaram, A.Y.M., Gilfillan, G.D., Kotoulas, G., Papandroulakis, N., Pavlidis, M., 2016. Transcriptomic changes in relation to early-life events in the gilthead sea bream (*Sparus aurata*). BMC Genomics 17, 1–14. https://doi.org/10.1186/s12864-016-2874-0

Schmidt, R., Strähle, U., Scholpp, S., 2013. Neurogenesis in zebrafish - from embryo to adult. Neural Dev. 8, 1–13. https://doi.org/10.1186/1749-8104-8-3/FIGURES/2

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. https://doi.org/10.1016/J.DCI.2009.06.013

Shimmura, T., Nakayama, T., Shinomiya, A., Fukamachi, S., Yasugi, M., Watanabe, E., Shimo, T., Senga, T., Nishimura, T., Tanaka, M., Kamei, Y., Naruse, K., Yoshimura, T., 2017. Dynamic plasticity in phototransduction regulates seasonal changes in color perception. Nat. Commun. 2017 81 8, 1–7. https://doi.org/10.1038/s41467-017-00432-8

Sutili, F.J., Gatlin, D.M., Heinzmann, B.M., Baldisserotto, B., 2018. Plant essential oils as fish diet additives: benefits on fish health and stability in feed. Rev. Aquac. 10, 716–726. https://doi.org/10.1111/raq.12197

Swain, P., Dash, S., Bal, J., Routray, P., Sahoo, P.K., Sahoo, S.K., Saurabh, S., Gupta, S.D., Meher, P.K., 2006. Passive transfer of maternal antibodies and their existence in eggs, larvae and fry of Indian major carp, *Labeo rohita* (Ham.). Fish Shellfish Immunol. 20, 519–527. https://doi.org/10.1016/J.FSI.2005.06.011

Tomasik, P., Horton, D., 2012. Enzymatic conversions of starch. Adv. Carbohydr. Chem. Biochem. 68, 59–436. https://doi.org/10.1016/B978-0-12-396523-3.00001-4

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40, e115. https://doi.org/10.1093/NAR/GKS596

Vadstein, O., Bergh, Ø., Gatesoupe, F.-J., Galindo-Villegas, J., Mulero, V., Picchietti, S., Scapigliati, G., Makridis, P., Olsen, Y., Dierckens, K., Defoirdt, T., Boon, N., De Schryver, P., Bossier, P., 2013. Microbiology and immunology of fish larvae. Rev. Aquac. 5, S1–S25. https://doi.org/10.1111/j.1753-5131.2012.01082.x

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, 1–12. https://doi.org/10.1186/GB-2002-3-7-RESEARCH0034/COMMENTS

Vieira, F.A., Thorne, M.A.S., Stueber, K., Darias, M., Reinhardt, R., Clark, M.S., Gisbert, E., Power, D.M., 2013. Comparative analysis of a teleost skeleton transcriptome provides insight into its regulation. Gen. Comp. Endocrinol. 191, 45–58. https://doi.org/10.1016/J.YGCEN.2013.05.025

Wang, Z., Zhang, S., Wang, G., 2008. Response of complement expression to challenge with lipopolysaccharide in embryos/larvae of zebrafish *Danio rerio*: acquisition of immunocompetent complement. Fish Shellfish Immunol. 25, 264–270. https://doi.org/10.1016/J.FSI.2008.05.010

Xiao, Z., He, T., Li, J., Gao, T., 2013. Ontogeny of the immune system in rock bream *Oplegnathus fasciatus*. Chinese J. Oceanol. Limnol. 2013 315 31, 1028–1035. https://doi.org/10.1007/S00343-013-2278-9

Xu, H., Liu, E., Li, Y., Li, X., Ding, C., 2017. Transcriptome analysis reveals increases in visceral lipogenesis and storage and activation of the antigen processing and presentation pathway during the mouth-opening stage in zebrafish larvae. Int. J. Mol. Sci. 18, 1634. https://doi.org/10.3390/ijms18081634

Yada, T., 2007. Growth hormone and fish immune system. Gen. Comp. Endocrinol. 152, 353–358. https://doi.org/10.1016/J.YGCEN.2007.01.045

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi. A J. Integr. Biol. 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Yúfera, M., Halm, S., Beltran, S., Fusté, B., Planas, J. V., Martínez-Rodríguez, G., 2012. Transcriptomic Characterization of the Larval Stage in Gilthead Seabream (*Sparus aurata*) by 454 Pyrosequencing. Mar. Biotechnol. 14, 423–435. https://doi.org/10.1007/s10126-011-9422-3

Zhang, Z., Chi, H., Dalmo, R.A., 2019. Trained innate immunity of fish is a viable approach in larval aquaculture. Front. Immunol. 10, 42. https://doi.org/10.3389/FIMMU.2019.00042/XML/NLM

CHAPTER 3

Specific evolution and gene family expansion of complement 3 and regulatory factor H in fish

Specific evolution and gene family expansion of complement 3 and regulatory factor H in fish

Manuscript published in Frontiers in Immunology 11 (2020)

https://doi.org/10.3389/fimmu.2020.568631

Acknowledgements

Thanks to Sandra Silva for help with the tissue expression studies. This work was financed by the European Union Horizon2020 Programme (PERFORMfish, grant nº 727610) and the Portuguese Foundation for Science and Technology (FCT) project to CCMAR (UIDB/04326/2020).

Specific evolution and gene family expansion of complement 3 and regulatory factor H in fish

Babak Najafpour+, João CR Cardoso*, Adelino VM Canário and Deborah M Power*

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

Running title: complement gene evolution in fish

* Corresponding authors

CReDit statement:

Babak Najafpour: Methodology, Bioinformatics, Validation, Investigation, Visualization, Writing – Original Draft, Review & Editing, João CR Cardoso: Methodology, Review & Editing, Adelino VM Canário: Review & Editing; Deborah M Power: Writing - Review & Editing, Methodology, Supervision, Project administration, Funding acquisition.

3.1. Abstract

The complement system comprises a large family of plasma proteins that play a central role in innate and adaptive immunity. To better understand the evolution of the complement system in vertebrates and the contribution of complement to fish immunity comprehensive *in silico* and expression analysis of the gene repertoire was made. Particular attention was given to C3 and the evolutionary related proteins C4 and C5 and to one of the main regulatory factors of C3b, factor H (Cfh). Phylogenetic and gene linkage analysis confirmed the standing hypothesis that the ancestral *c3*/*c4*/*c5* gene duplicated early. The duplication of *C3* (*C3.1* and *C3.2*) and *C4* (*C4.1* and *C4.2*) was likely a consequence of the (1R and 2R) genome tetraploidization events at the origin of the vertebrates. In fish, gene number was not conserved and multiple *c3* and *cfh* sequence related genes were encountered, and phylogenetic analysis of each gene generated two main clusters. Duplication of *c3* and *cfh* genes occurred across the teleosts in a species-specific manner. In common, with other immune gene families the *c3* gene expansion in fish emerged through a process of tandem gene duplication. Gilthead sea bream (*Sparus aurata*), had nine *c3* gene transcripts highly expressed in liver although as reported in other fish, extra-hepatic expression also occurs. Differences in the sequence and protein domains of the nine deduced C3 proteins in the gilthead sea bream and the presence of specific cysteine and N-glycosylation residues within each isoform was indicative of functional diversity associated with structure. The diversity of C3 and other complement proteins as well as Cfh in teleosts suggests they may have an enhanced capacity to activate complement through direct interaction of C3 isoforms with pathogenic agents.

Keywords: Complement system, environment, evolution, fish, innate immunity, liver, skin

3.2. Introduction

In vertebrates, the innate and adaptive immune response provides protection from pathogens. The innate immune system was the first defence mechanism to evolve and includes a range of non-specific mechanical and physiological barriers (e.g. skin, mucous, pH, lysozyme, etc), and granulocytes, macrophage and dendritic cells that are responsible for identification and elimination of pathogens (Dzik, 2010; Medzhitov and Janeway, 2000; Netea et al., 2008). The adaptive immune response mediated by B and T lymphocytes, first emerged in vertebrates and is characterised by its specific recognition of pathogens and antigens, immunological memory and "self and non-self" recognition (Flajnik and Kasahara, 2010). In fish, innate immunity appears to have a dominant role in combatting pathogens (Bergljot, 2006).

The complement system is part of the innate immune system and consists of a group of plasma proteins produced by the liver that may also act as an effector and signalling mechanism for adaptive immunity (Riera Romo et al., 2016). More than 30 components of the complement system have been identified in vertebrates and their expression is regulated by a range of stimuli including cytokines and hormones. The complement system can be activated by the antibody dependent-classical pathway, the mannose binding protein (MBP)-lectin pathway and the alternative pathway (Merle et al., 2015; Ricklin et al., 2010; Whaley et al., 1993). The classical complement pathway is triggered by interaction between the antigen-antibody complex and complement 1q (C1q) and two other effector proteins, complement 4 (C4) and complement 2 (C2) (Petersen et al., 2000). The MBP-lectin pathway is triggered by carbohydrates such as mannans or N-acetylglucosamine (GlcNAc) (Sunyer et al., 2003), which are found in the bacterial cell wall (Sakai, 1992). The alternative pathway (AP) does not require specific molecular recognition and is activated by hydrolysis of complement 3 (C3) by factor B, factor D and properdin and this "C3 tick-over" mechanism forms the basis of the rapid activation of all the complement pathways (Harboe et al., 2009, 2004). Other complement components (C5, C6, C7, C8, and C9) form the membrane-attack complex (MAC) that causes cell lysis.

All three complement pathways converge at C3, which has a central role in complement system function. C3 when enzymatically cleaved generates two protein subunits C3a and C3b, with the latter attaching to the pathogen surface and activating the lytic pathway. C3b is required for the sequential junction of C5b, C6, C7, C8 and C9 proteins to form the membrane-attack complex (MAC), which provokes cell lysis (Sarma and Ward, 2011). C3, C4 and C5 share sequence similarity and they belong to the α2-macroglobulin (α2M) family

(Sottrup-Jensen et al., 1985). The proteins possess a variable central region associated with pathogen recognition and a thioester motif for attachment of complement to the target cells (Ferreira et al., 2010). Several plasma and membrane-bound proteins or complement control proteins (CCP) regulate the activity of the complement system. In mammals, complement factor H (CFH) is an important complement system regulator that binds C3b and inhibits the alternative pathway (Ferreira et al., 2010; Rebl and Goldammer, 2018). In humans, five CFH-related plasma proteins exist (CFHR1 to 5, Skerka et al., 2013) and CFH and CFH-related proteins have also been described in a few teleost species (Qi et al., 2018; Sun et al., 2010).

From teleost fish to mammals, the complement pathway appears to have been conserved, although in teleosts it functions at a lower temperatures and the titre of complement components of the alternative pathway in plasma is higher (Boshra et al., 2006; Hawlisch and Köhl, 2006). The complement pathway has been characterized in relatively few fish species (trout, *Oncorhynchus mykiss*; sea bream, *Sparus aurata*; carp, *Cyprinus carpio*; medaka, *Oryzias latipes*; two Antarctic teleosts, *Trematomus bernacchii* and *Chionodraco hamatus*; and zebrafish, *Danio rerio*) and studies mainly focussed on the multiple copies of the *c3* gene (Forn-Cuní et al., 2014; Kuroda et al., 2000; Mauri et al., 2011b; Melillo et al., 2015; Nakao et al., 2000; J O Sunyer et al., 1997; Sunyer et al., 1996; Zhang and Cui, 2014) and in zebrafish they were shown to have a different transcriptional response to LPS injection (Forn-Cuní et al., 2014). There is even less information about complement regulatory factors and so far they have only been characterized in rainbow trout, yellow croaker (*Larimichthys crocea*) and zebrafish and in common with humans are present in multiple gene copies (Anastasiou et al., 2011; Qi et al., 2018; Sun et al., 2010). Studies of immune challenged zebrafish, rainbow trout, yellow croaker, winter flounder (*Pseudopleuronectes americanus*) and Japanese flounder (*Paralichthys olivaceus*) indicate *cfh* responds (Anastasiou et al., 2011; Qi et al., 2018; Straub et al., 2004; Sun et al., 2010; Wang et al., 2017).

It seems likely that the evolution of the immune system in fish was shaped by their contact with the greater number and diversity of microbes and viruses found in water compared to other vertebrates (Groff, 2001). The barrier function of epithelia such as the skin in vertebrates has a major role in protection and impedes the entry of pathogens. The complement system rapidly responds to pathogens that breach the damaged barrier or adsorb to and colonize the surface (Dovezenski et al., 1992). To lay the foundation for a better understanding of the physiology of the complement system in teleosts the evolution of complement proteins and Cfh was studied in deuterostomes paying particular attention to the teleosts. The wealth of available

fish genomes was exploited to pinpoint specific events in complement evolution by including in the analysis a diversity of phylogenetically informative vertebrate taxa. Considering the proposed importance of the alternative pathway in teleosts (Yano et al., 1988) and the key role of C3 in the lytic pathway the evolution of C3 isoforms was targeted. The results of the global analysis of the genes of the complement system in vertebrates indicate it was well conserved during evolution. Of note was the highly variable number of *c3* genes in different fish and so this and the related *c4* and *c5* genes were examined in more detail. By conducting a multi-species analysis of *c3*/*c4*/*c5* using evolutionary informative species, with particular emphasis on the most successful group of vertebrates, the teleosts, (Nelson et al., 2016) a consensus evolutionary model was proposed. The model reveals that duplication of *c3* is common across the fishes and that the complement system in fish has been under different evolutionary pressure compared to other vertebrates. Our data confirm that the *c3* gene expansion previously identified (Forn-Cuní et al., 2014; Kuroda et al., 2000; Mauri et al., 2011b; Melillo et al., 2015; Nakao et al., 2000; J O Sunyer et al., 1997; Sunyer et al., 1996; Zhang and Cui, 2014) occurred across the teleosts and that the deduced proteins form two main clusters, C3.1 and C3.2. In common, with other immune gene families the *c3* gene expansion in fish emerged through a process of tandem gene duplication, a process associated with an unstable genomic organization and the appearance of new genes (Gemayel et al., 2010; Nydam and DeTomaso, 2011). Furthermore, retention of multiple *c3* gene copies was not associated with specialization through tissue specific expression. However, analysis of the differences in the sequence and domains of the nine deduced C3 proteins in the gilthead sea bream (*Sparus aurata*), and the presence of specific cysteine and N-glycosylation residues within each isoform was indicative of potential functional diversity associated with structure. In vertebrates in which *c3* gene expansion occurred, cfh-like the C3b regulatory factor also underwent expansion so that a species-specific gene repertoire emerged. Overall, the results indicate that in fish complement signalling and regulatory proteins shared common ancestry with the tetrapod homologues but evolved under distinct pressures.

3.3. Material and methods

3.3.1. Screening for the complement system and complement regulatory factor H in ray-finned fish

In order to analyse and understand the evolution of the complement system (C1-C9) in fish and its regulation, orthologues of human C1QA, C1QB, C1QC, C2, C3, C4 (C4A/C4B), C5, C6, C7, C8 (α/β/δ) and C9 genes. Complement and complement regulatory factor (CFH and CFHR) genes were procured in the genomes of several ray-finned fish using tBLASTn (**Supplementary table 3.1**, Altschul et al., 1990). The species analysed included, the spotted gar (*Lepisosteus oculatus*), a basal ray-finned fish that radiated prior to the teleost expansion, and eleven teleost genomes: the Amazon molly (*Poecilia formosa*), cave fish (*Astyanax mexicanus*), cod (*Gadus morhua*), Fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), platyfish *(Xiphophorus maculatus),* stickleback (*Gasterosteus aculeatus*), Tetraodon, (*Tetraodon nigroviridis*), tilapia (*Oreochromis niloticus*), European sea bass (*Dicentrarchus labrax*), Gilthead sea bream (*Sparus aurata)* and zebrafish (*Danio rerio*) (**Supplementary table 3.1**). Sequences were identified based on their high sequence similarity (*e-value* $\leq$ 1e $^{-40}$) with human homologues and *in silico* genome annotations. The zebrafish complement component and CFH/CFHL sequences retrieved were subsequently used to search again the ray-finned fish genomes to retrieved missing/non-annotated hits. Sequence identity was confirmed by searching the NCBI non-redundant protein sequence database using the human filter (taxid:9606) and the isolated fish genes and also by phylogeny (see methods below). The results from the general analysis of the complement system in fish revealed the *c3* genes and the *c4* genes, which belong to the alpha-2 macroglobulin superfamily of thioester containing proteins, underwent a large expansion. For this reason, the members of the alpha-2 macroglobulin superfamily of thioester containing proteins, C3, C4 and C5 were selected for more in-depth analysis. The fish *cfh* and putative *cfhr* family genes, that encode complement regulator proteins, were also further analysed as multiple genes were found. All searches were performed against the most recent annotated fish genome assemblies available from ENSEMBL or NCBI (**Supplementary table 3.1**).

3.3.2. In depth analysis of c3, c4 and c5 and cfh genes

The deduced proteins of human C3, C4, C5, CFH and CFHRs were used as the bait for database searches which included the fish listed above and other teleost fish genomes; flatfish smooth tongue sole (*Cynoglossus semilaevis*), Japanese flounder (*Paralichthys olivaceus*), Antarctic black rockcod (*Notothenia coriiceps*) and Atlantic salmon (*Salmo salar*). We also interrogated genomes of fish species that diverged earlier in the vertebrate radiation and thus are considered to possess less rearranged genomes: the lobe-finned coelacanth

(*Latimeria chalumnae*), that diverged basal to the tetrapods; two cartilaginous fishes, the elephant shark (*Callorhinchus milii)* and whale shark (*Rhincodon typus*), that are basal to the bony vertebrates and two Agnathans (jawless fish), the sea lamprey (*Petromyzon marinus*) and the inshore hagfish (*Eptatretus burgeri*) which diverged prior to the gnathostomes (**Supplementary table 3.1**). For comparative analysis the mouse (*Mus musculus*), the chicken (*Gallus gallus*), the reptile (*Anolis carolinensis*) and the amphibian clawed African toadfish (*Xenopus tropicalis*) were also included in the analysis as well as two basal deuterostomes, the urochordate Ciona (*Ciona intestinalis*) and the cephalochordate Amphioxus (*Branchiostoma floridae*) genomes to infer the likely evolutionary origin of the vertebrate complement proteins and CFH-family members (**Supplementary table 3.1**).

3.3.3. Sequence comparisons and phylogenetic analysis

Multiple sequence alignments were performed using the MUSCLE algorithm (Edgar, 2004) in the Aliview platform (Larsson, 2014) and conserved regions were identified. Searches in ray-finned fish genomes and transcriptomes identified many incomplete genes/transcripts that encoded for complement protein fragments. When multiple transcripts (of variable sizes) for the same gene were found they were aligned and sequences merged (> 98% sequence identity) to obtain a full-length protein sequence. GeneDoc (http://www.nrbsc.org/gfx/genedoc) was used to calculate identity/similarity between the sequences. Only the deduced protein sequences that encoded full-length C3 proteins were considered for phylogenetic analysis (**Supplementary table 3.1**).

Phylogenetic trees were constructed based on alignments of the deduced amino acid (aa) sequence of the fish and tetrapod proteins including the invertebrate deuterostome genes (urochordate and cephalochordate). Sequences were aligned using the MUSCLE algorithm in the AliView platform and the alignment was edited to remove sequence gaps and poorly aligned regions. Phylogenetic trees were constructed using Maximum-Likelihood (ML) and Bayesian Inference (BI) methods and the models that best fit the data were calculated in model test-ng 0.1.5. ML and BI trees were constructed in the CIPRES Science Gateway V.3.3 (Miller et al., 2010) and run on XSEDE. ML trees were built with the RAxML v8.2.12 (Stamatakis, 2014) method with a WAG matrix and 1000 bootstrap replicates and BI trees using MrBayes (Ronquist et al., 2012) with a WAG matrix and 1.000.000 generation sampling and probability values to support tree branching. For the fish Cfh tree a similar strategy to that for complement proteins (outlined above) was used, other Cfh related proteins, CFHR from human

and Cfhl from zebrafish (Sun et al., 2010) were also included in the alignment; the sequence file was manually edited to delete gaps and misaligned sequences before tree building using the BI method with a VT matrix. All trees were displayed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The C3/C4/C5 trees were rooted with the deduced protein sequences of human α-macroglobulin (A2M, ENST00000318602.12) and cluster of Differentiation 109 (CD109, ENST00000437994.6). The CFH/CFHR tree was rooted with the predicted protein sequence of human coagulation factor XIII B (ENST00000367412.1). Trees for the fish C1, C2/Cfb, C6, C7, C8 and C9 were also built to confirm sequence identity and were constructed using the BI method and default settings and they were rooted using the human orthologue.

3.3.4. Neighbouring gene analysis

To better characterize the evolution of the multiple gene copies of *c3* and *c4*, the neighbouring gene environment of these genes in teleosts along with *c5* and *cfh* genes was characterized. Homologues of human *C3*, *C4* and *C5* neighbouring genes were procured in the coelacanth, spotted gar, elephant shark and four teleost genomes (tetraodon, stickleback, medaka and zebrafish). The species selection was based on the quality of the available genomes and distinctive evolutionary patterns suggested by the number of genes identified. For comparison, the neighbouring gene environment of the human, lizard or chicken were also characterised. Fifteen genes upstream and downstream of human *C3*, *C4* and *C5* gene loci were retrieved and homologues were identified in the other species analysed using the genome annotations provided by Genomicus software (http://www.genomicus.biologie.ens.fr) and by homology sequence searches in the genomes available from ENSEMBL. The gene environment of the *C4* gene in the chicken was characterized in-depth as phylogenetic analysis suggested that an extra gene copy exists.

3.3.5. Protein motif annotation

To identify changes indicative of potential functional divergence between the vertebrate orthologues, the fish paralogues and the species-specific duplicates the deduced proteins sequence of selected fish C3, C4 and C5 and also of the multiple C3 isoforms from the gilthead sea bream were annotated and compared with the homologues in human and chicken. Sequence alignments were performed using the MUSCLE algorithm in the Aliview platform and edited in GeneDoc software (http://www.nrbsc.org/gfx/genedoc). The signal

peptide was deduced using SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) identification of protein domains was carried out based on the crystallographic structure of the human C3 (Janssen et al., 2005) and homology modelling of an Antarctic teleost C3 (Melillo et al., 2015). Sequence alignments of the thioester-bound domain of fish C3 and C4 deduced proteins and the catalytic site for C4 were analysed in detailed. The amino acid residues associated with protein structure, conformation and signalling such as the cysteine residues responsible for disulphide bridges and N-glycosylation sites (N-X-T/S) were mapped in human C3 using Uniprot annotation (https://www.uniprot.org) and identified manually in the gilthead sea bream.

3.3.6. Transcriptome analysis

To increase the number of teleost complement sequences available for analysis and to infer the importance and functional divergence of the complement components in teleost skin, expression data was analysed. Databases of gene transcripts of the European sea bass and gilthead sea bream (*Sparus aurata*) (Louro et al., 2016) and RNA-seq transcriptome (Illumina platform) assemblies of the European sea bass skin (GFJW00000000) (Pinto et al., 2017), Senegalese sole skin (PRJEB29449) (Pinto et al., 2019) and the *de novo* transcriptome assemblies of the intestine, skin and head-kidney of two Antarctic fish, the black rockcod and marbled rockcod (*Notothenia rossii*, **Supplementary table 3.1**), were interrogated. The nucleotide sequences (*e-value* $\leq 1e^{-30}$) were retrieved and translated into proteins using the Expasy translation tool (https://web.expasy.org/translate/) and annotation was assigned using blastp against the human (taxid:9606) and confirmed by phylogenetic analysis. Skin transcriptome data was complemented with data retrieved from previously published transcriptome and proteome studies of fish skin and mucous.

3.3.7. RNA extraction and cDNA synthesis

To further confirm *c3* gene expression and extend understanding of the persistence in the genome of the multiple gene isoforms, total RNA (tRNA) from the gills, spleen and liver (n= 3) of gilthead sea bream (average weight = 87.09 ± 5.54 g) that was already available in the laboratory and stored at -80ºC in the context of a previous study of skin regeneration was used (Mateus et al., 2017). Reactions for cDNA synthesis contained 500ng of the DNase treated tRNA (denatured at 65ºC for 5 min), 10 ng of pd(N)6 random hexamers (Jena Bioscience, Germany), 2 mM dNTPs (ThermoScientific, USA), 100 U of RevertAid Reverse Transcriptase

and 8 U Ribolock RNAse inhibitor (ThermoScientific) in a final reaction volume of 20 µl. The time and temperature cycle of the cDNA synthesis reaction was 10 min at 20 ºC; 50 min at 42 ºC and 5 min at 70 ºC. The integrity and quality of the synthetized cDNA was assessed by amplification of the sea bream ribosomal subunit *18s* rRNA using the following thermocycle: 95°C for 10 min followed by 25 cycles of [95°C for 20 sec; 60 °C for 20 sec; 72 °C for 20 sec] and a final cycle at 72 °C for 5 min.

3.3.8. Quantitative expression of sea bream c3 isoforms

Specific primers for each gilthead sea bream *c3* gene isoform were designed and the amplified PCR products were sequenced to confirm their identity. Five of the 9 identified gilthead sea bream *c3* gene transcripts (*c3.1.1, c3.1.2, c3.1.3, c3.1.4 and c3.*2) were successfully amplified. Difficulties with design of specific and efficient qPCR primers meant that *c3.1.5, c3.1.6, c3.1.7* and *c3.1.8* were not quantified. The tissue distribution of the *c3* gene transcripts was analysed in the gill, spleen and liver (**Table 3.1**). Quantitative real-time PCR (qPCR) reactions were performed in duplicate (< 5% variation between replicates) using a BioRad CFX Connect Real Time System and SsoFast EvaGreen supermix (Bio-Rad, Portugal) and 96-well plates (Axygen). The gilthead sea bream 18S ribosomal subunit (*18s*) was used as the reference gene as it showed stable expression levels in all samples. The final qPCR reaction volume was 10 µl and contained 200 nM of each primer, 2 µl of the template cDNA (diluted 1:5 target gene, 1:5000 for *18s*) and 8 µl of EvaGreen Supermix (Bio-Rad, Portugal). Thermocycling conditions were 95 ºC for 30 sec and followed by 39 cycles of [95 ºC for 5 sec and 60 °C for 10 sec, **Table 3.1**]. To detect non-specific amplification products and primer dimers melting curves were performed. Standard curves were included in PCR plates for each *c3* gene isoform and prepared from serial dilutions of the sequenced and quantified amplicons. Control reactions were included in all runs to confirm the absence of qPCR or genomic contamination. qPCR reaction efficiencies and r2 (coefficient of determination) were all > 90% for each target gene transcript. Expression normalization was performed using *18s* ribosomal RNA (**Table 3.1**).

Table 3.1. **List of primers used to amplify C3 and 18S ribosomal RNA genes in gilthead sea bream.** The annealing temperature, the efficiency (%) of the primer pairs and the linearity R2 of the standard curve are indicated.

Name	Sequence (5'-3')	Temp(º C)	Efficiency (%)	R2
Complement gene				
C3.1.1Fwd	GATCAGGTTGGAGAACCCAG	60	101.4	0.99
C3.1.1Rev	GACCCTGTCTCCTTCAGAAC			
C3.1.2Fwd	GTTAAGGTCACTGAGTGATGC	60	92.7	0.99
C3.1.2Rev	GCTCACTAAAGTGCCTTTACTC			
C3.1.3Fwd	GCTTTGAAACATAAGAACTGCACA	60	91.6	0.99
C3.1.3Rev	TCCATTTGCCAGAGGTTAATTTGT			
C3.1.4Fwd	GATGAACAGAGTCAGGCGTAC	62	94.3	1
C3.1.4Rev	CATCCGTGTTGCGTGTACTTC			
C3.2Fwd	GACCTGAGGGACACAGTCAG	60	97.6	0.99
C3.2Rev	CACGGTGGACTTGCTGAAGT			
Reference gene				
18S Fwd	TGACGGAAGGGCACCACCAG	60	90.3	1
18S Rev	AATCGCTCCACCAACTAAGAACGG			

3.4. Results

3.4.1. Complement system and regulatory proteins in ray-finned fish

Initial screening for members of the complement cascade revealed that in ray-finned fish some genes of the signalling pathway have duplicated but a larger expansion of the complement 3 (*c3*) gene occurred when compared to human (**Table 3.2**). In fish, three *c3* genes were retrieved from the spotted gar genome but the number of *c3* genes varied across the analysed teleost genomes (**Figure 3.1**). A large expansion of *c3* genes was found in the sea bass genome and many hits corresponded to small incomplete genes. Duplicates of complement genes encoding *c4* and *c6* were identified in some fish genomes and in cod four *c4* genes were found and in all the teleost genomes analysed duplicate *c7* genes (*c7a*, *c7b*) persisted (**Figure 3.1**, **Table 3.2**, **Supplementary figure 3.1**). For other complement members the gene number was similar between ray-finned fish and human with some exceptions such as the identification of four *c2/cfb* in zebrafish (*c2/cfb.1.1*, *c2/cfb.1.2*, *c2/cfb.1.3*, *c2/cfb.2*) and two *c9* (*c9.1*, *c9.2*) in the cod genome (**Table 3.2**, **Supplementary figure. 3.1**). The *cfh* gene family, in common

with the *c3* genes, underwent a large expansion in ray-finned fishes and gene number ranged from a single gene copy in the spotted gar, two copies in the sea bass to five copies in cavefish and six in the zebrafish and Atlantic salmon. This suggests that in addition to the expansion of the *c3* gene, expansion of *cfh* its regulatory factor also occurred in teleosts (**Table 3.2**).

Table 3.2. Number of genes of the complement system found in ray-finned fishes. The identified human gene copies are indicated for comparison. The number of incomplete C3 genes found in teleost genomes is indicated within brackets. Searches included the genes of the three chains of C1Q (A, B, C) and of C8 (α, β, δ).

	C1Q (A, B, C)	C2/CFB	C4	C3	C5	C6	C7	C8 (α , β, δ)	C9	CFH
Human	**3**	**2**	**2**	**1**	**1**	**1**	**1**	**3**	**1**	**1**
Spotted gar	**3**	**2**	**1**	**3**	**1**	**1**	**1**	**3**	**1**	**1**
Amazon molly	**2 (A, C)**	**2**	**2**	**7**	**1**	**2**	**2**	**3**	**1**	**1**
Cavefish	**3**	**3**	**2**	**3**	**1**	**1**	**2**	**3**	**1**	**5**
Cod	**3**	**2**	**4**	**4**	**1**	**1**	**2**	**2 (α, β)**	**2**	**3**
Fugu	**3**	**2**	**1**	**5**	**1**	**1**	**2**	**3**	**1**	**2**
Medaka	**3**	**2**	**2**	**5 (1)**	**1**	**1**	**2**	**3**	**1**	**2**
Platyfish	**3**	**2**	**1**	**5**	**1**	**1**	**2**	**3**	**1**	**3**
Stickleback	**1 (C)**	**2**	**1**	**5 (3)**	**1**	**1**	**2**	**3**	**1**	**3**
Tetraodon	**3**	**2**	**2**	**4**	**1**	**1**	**2**	**3**	**1**	**4**
Tilapia	**3**	**2**	**2**	**5**	**1**	**1**	**2**	**3**	**1**	**4**
Zebrafish	**3**	**4**	**1**	**8**	**1**	**2**	**2**	**3**	**1**	**6**
Sea bass	**3**	**2**	**2**	**4 (10)**	**1**	**1**	**2**	**3**	**1**	**2**
Seabream	**3**	**2**	**1**	**9**	**1**	**1**	**2**	**3**	**1**	**4**

The number of incomplete C3 genes found in teleost genomes is indicated within brackets.

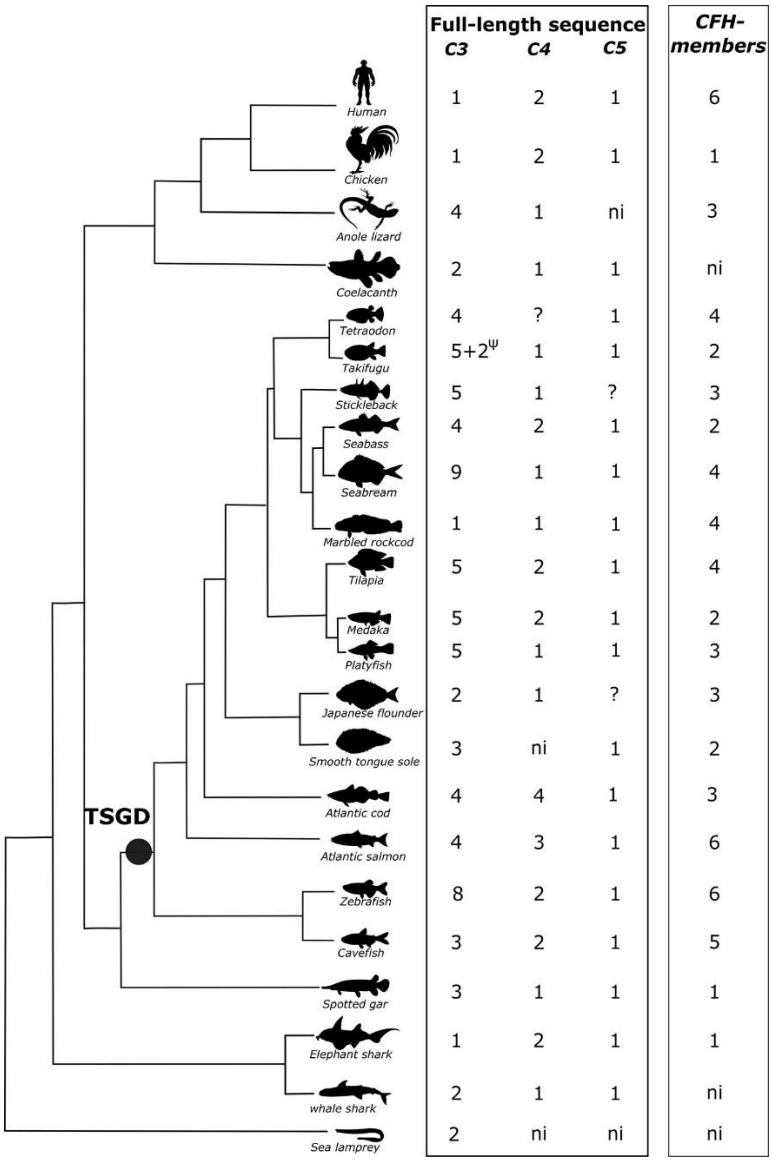


Figure 3.1. Dendrogram showing the number of predicted C3, C4, C5, and CFH genes and transcripts identified in fish and other vertebrates. The number of full-length/complete C3, C4, and C5 sequences are indicated. Incomplete C3 sequences were also found for stickleback, sea bass, marbled rockcod, medaka, smooth tongue sole and Atlantic salmon (**Supplementary table 3.1**). C4 incomplete sequences were found in the whale shark (**Supplementary table 3.1**). ?, only incomplete sequences found. ψ, pseudogene; ni, not identified; TSGD, teleost specific genome duplication.

3.4.2. The c3, c4, c5 and cfh family genes in fish

3.4.2.1. Agnathan and cartilaginous fishes

In the sea lamprey, two *c3* genes (*c3.x.1* and *c3.x.2*) were found and they mapped

in tandem in chromosome 14 but in the inshore hagfish genome assembly four putative *c3* genes were identified (**Supplementary table 3.1**). The two predicted lamprey C3 proteins are likely to be complete but in hagfish the predicted proteins from the four genes are incomplete and one gene (*c3.y1*) encoded the transcript that was previously reported (Ishiguro et al., 1992). In hagfish two of the duplicate c3 genes (*c3.y*1 and *c3.y2*) also mapped in close proximity in the same genome region (FYBX02010427.1) suggesting a similar genome structure to that found in the sea lamprey.

For cartilaginous fish, in the elephant shark a single *c3* gene was retrieved but for the whale shark two genes were found (**Figure 3.1**, **Supplementary table 3.1 C3**). In lamprey and hagfish, no gene homologues of human *C4* or *C5* were identified. In the elephant shark two *c4* genes were found but in the whale shark three hits for putative *c4* genes were obtained but only one (*c4.1.1*) coded for a full-length *c4* protein (Figure 3.1). The remaining sequences (*c4.1.2* and *c4.1.3*) were shorter and they aligned to different regions of the full-length C4 protein. It was not possible to establish if they correspond to different genes or are fragments of the same gene (**Supplementary table 3.1 C4**). For *c5* in cartilaginous fish only a single gene hit was obtained and the deduced protein was full-length (**Supplementary table 3.1 C5**).

A homologue of the human *CFH* gene was retrieved from the elephant shark but not from the lamprey or whale shark genome assemblies.

3.4.2.2. Ray-finned fish

Searches in ray-finned fish genomes identified multiple putative c3 genes. In spotted gar, three c3 genes and single c4 and c5 gene were retrieved, which were of a similar length to the human orthologues (**Figure 3.1**). In teleosts, the number of *c3* genes varied from one in the marbled rockcod to nine in the gilthead sea bream. Sequence comparisons revealed that several of the deduced teleost *c3* genes coded for incomplete C3 proteins and that most of them lacked the protein beta-chain domain. However, searches in the NCBI transcript database identified full-transcripts for some of the *c3* genes that were incomplete in the genome assembly of the Amazon molly (*c3.1.3*, *c3.1.4* and *c3.1.5*), cavefish (*c3.1.2*), medaka (*c3.1.4*), zebrafish (*c3.1.5*) and smooth tongue sole (*c3.2*, **Supplementary table 3.1**). In the Fugu genome *c3.1.5* and *c3.1.6* genes are annotated as pseudogenes, if the incomplete genes found in other teleosts are also pseudogenes or the result of poor sequence assembly remains to be established (**Supplementary table 3.1**). In the sea bass genome only two full-length *c3* genes were found (*c3.1.1* and *c3.2*) and it is in this species that the largest number of incomplete *c3* genes were

retrieved. Searches in the sea bass transcriptome and NCBI database retrieved three full-length transcripts (**Supplementary table 3.1 C3**) of which one is the transcript for one of the predicted full-length genes. The other two transcripts did not align (> 98% nucleotide identity) with any genome region suggesting there may be errors in the genome or transcriptome assemblies (**Supplementary table 3.1 C3**).

For *c4* and *c5* a single gene was retrieved in most species except for tilapia, medaka, sea bass, cavefish and zebrafish where two *c4* genes were identified and the salmon and cod that had three and four genes, respectively (**Figure 3.1**). In tetraodon, two putative genes for *c4* (one that maps to chromosome 8 and another that maps to a non-annotated chromosome) were retrieved but sequence alignment of the translated proteins with the full-length C4 from human/fish revealed that they do not overlap and aligned to different regions. If they correspond to two independent genes or are fragments of the same gene remains to be established. In the Japanese flounder genome three hits for a putative *c5* gene were retrieved that most likely represent fragments of a single gene since although they did not overlap, they aligned with different regions of the human/fish full-length homologue proteins.

In the ray-finned fish a single gene for a putative *cfh* family member was found in the spotted gar but in teleosts the gene number varied from two in fugu, medaka, sea bass and smooth tongue sole to five and six in cavefish and zebrafish genomes, respectively (**Figure 3.1**).

3.4.2.3. Lobe-finned fish and tetrapods

Two *c3* and a single *c4* and *c5* gene were found in the coelacanth but our sequence searches failed to retrieve a putative *cfh* gene from the lobe-finned fish genome. In tetrapods multiple *C3* genes and *Cfh* genes exist. In the amphibian, *Xenopus* five full-length *C3* genes and a single *Cfh* gene were found but for the reptile, the anole lizard four *C3* gene copies were retrieved and three gene copies of *Cfh* also exist (**Supplementary table 3.1**, **Figure 3.1**). In the chicken gene number was similar to human and single *C3* and *C5* genes and duplicate *C4* genes were found but only a single *Cfh* family member was retrieved (**Figure 3.1**, **Supplementary table 3.1**).

3.4.3. Phylogenetic analysis

3.4.3.1. C3, C4 and C5 phylogeny

Phylogenetic trees of C3, C4 and C5 suggest that they emerged prior to the

vertebrate radiation from a common ancestral gene via gene duplications (**Figure 3.2A** and **3.2B** and **Supplementary figure 3.2**). According to the tree topology, C5 was the first member to diverge and vertebrate C3 and C4 emerged subsequently. The C3 clade possess the largest number of members (**Figure 3.1**). The C3 clade was rooted with the lamprey and hagfish C3 proteins and two main sequence clusters were found: one cluster contained C3 from tetrapods, coelacanth and sharks and included C3 from most of the teleosts and was designated C3.1 and the second cluster contained only ray-finned fish sequences and was named C3.2 (**Figure 3.2A** and **Supplementary figure 3.2**). The two spotted gar C3 (C3.1.1 and C3.1.2) sequences clustered within the C3.1 clade and the third clustered within the C3.2 clade indicating that the C3 gene duplication that originated *c.3.1* and *c.3.2* genes occurred prior to the ray-finned fish radiation. The clustering of the multiple ray-finned fish sequences within the C3.1 and C3.2 clusters suggests that they arose through lineage or species-specific duplication events. The tree topology also suggests that the multiple *c3* gene copies found in lizard and *Xenopus* and the tunicate *c3/4/5* sequences also arose through species-specific duplication events.

For C4, the tree topology indicated that two distinct protein clusters existed, and they were designated C4.1 and C.4.2 (**Figure 3.2B**, **Supplementary figure 3.2**). The C4.1 cluster contained sequences from tetrapods and most of the teleost C4 sequences including 4 cod members, C4.1.1, C4.1.2, C4.1.3, C4.1.4, which based on clustering probably arose from a species-specific duplication. The C4.2 cluster contained a C4 sequence from the chicken along with the spotted gar C4 and the deduced protein of one gene from zebrafish and cavefish and two genes from the Atlantic salmon (C4.2.1 and C4.2.2). The lobe-finned fish C4 and the two cartilaginous fish C4s tend to group within the C4.2 cluster but analysis of their neighbouring gene environment suggests that they are *c4.1* genes (see short-range gene linkage section). Clustering of the fish C4 sequences suggested that in common with the fish *c3* genes they expanded by lineage and species-specific duplication events. The existence of two *C4* genes in the chicken, which clustered independently in sister clades along with the teleost paralogues suggests that the C4.1 and C4.2 genes resulted from a gene duplication that occurred early in the vertebrate radiation and that the duplicate gene copies only persisted in the genomes of a few ray-finned fish species and in the chicken (**Figure 3.2B**).

Within the C5 cluster two main branches exist one containing C5 from ray-finned fishes and the other cluster containing C5 from the cartilaginous fish, the coelacanth and the tetrapods (**Figure 3.2B**, **Supplementary figure 3.2**).

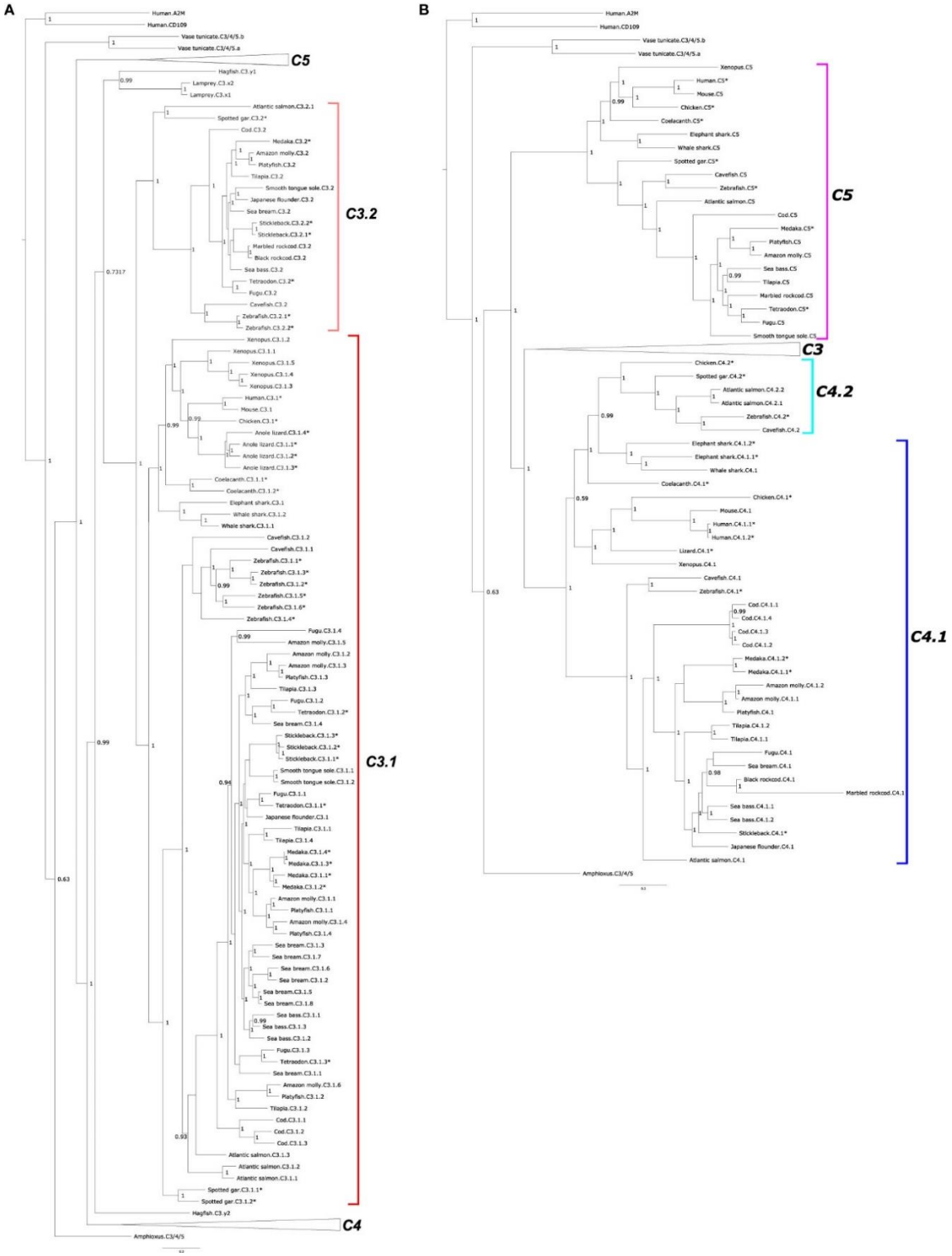


Figure 3.2. Phylogenetic tree of complement factors C3, C4, and C5 in fish and other vertebrates. The tree was built only with full-length (both β- and α-chain) sequences and homologues from a urochordate (*Ciona intestinalis*) and a cephalochordate (*Branchiostoma floridae*) were also included to understand the evolutionary context. The tree was built with the

BI method and branch support values (posterior probability values) are shown. Two subsets of the same phylogenetic tree showing different family members: (A) C3 and (B) C4 and C5 are represented to facilitate interpretation. In (A) the two major C3 clades were named C3.1 (red) and C3.2 (pink). For (B) the two C4 clades were named C4.1 (dark blue) and C4.2 (light blue). The teleost C5 clade is boxed in purple. The C4 sequences of the coelacanth and cartilaginous fishes (elephant shark and whale shark) were named C4.1 based on the similarity of their gene environment with the other vertebrate C4.1 genome regions. Duplicated fish c3 and c4 genes are numbered arbitrarily. The tree was rooted with the deduced protein sequences from human α-macroglobulin (A2M, ENST00000318602.12) and cluster of Differentiation 109 (CD109, ENST00000437994.6). The species for which gene linkage analysis was performed (Figures 3.4–3.6) are indicated by an asterisk (*). Sequence accession numbers are available in **Supplementary table 3.1. A** similar tree built with the Maximum Likelihood (ML) method is presented as **Supplementary figure 3.2**.

3.4.3.2. Cfh family member phylogeny

The phylogenetic tree of the deduced proteins of fish Cfh-members suggested that they shared common ancestry with human CFH/CFHR. Fish Cfh-members grouped in two main teleost sequence clusters named, Cfha and Cfhb, that probably arose from the teleost genome duplication event (**Figure 3.3**). The distribution of the deduced fish Cfh proteins in the tree indicated that evolution of this protein family was complex and the arrangement of the sequences within each cluster suggested they evolved in a lineage and species-specific manner (**Figure 3.3**). All zebrafish sequences grouped in the Cfha cluster and some sequences clustered in close proximity suggesting they represent recent species-specific duplications. The deduced protein of the two medaka *cfh* genes clustered in the Cfhb clade but the stickleback, cavefish and Tetraodon had *cfh* genes in both clades.

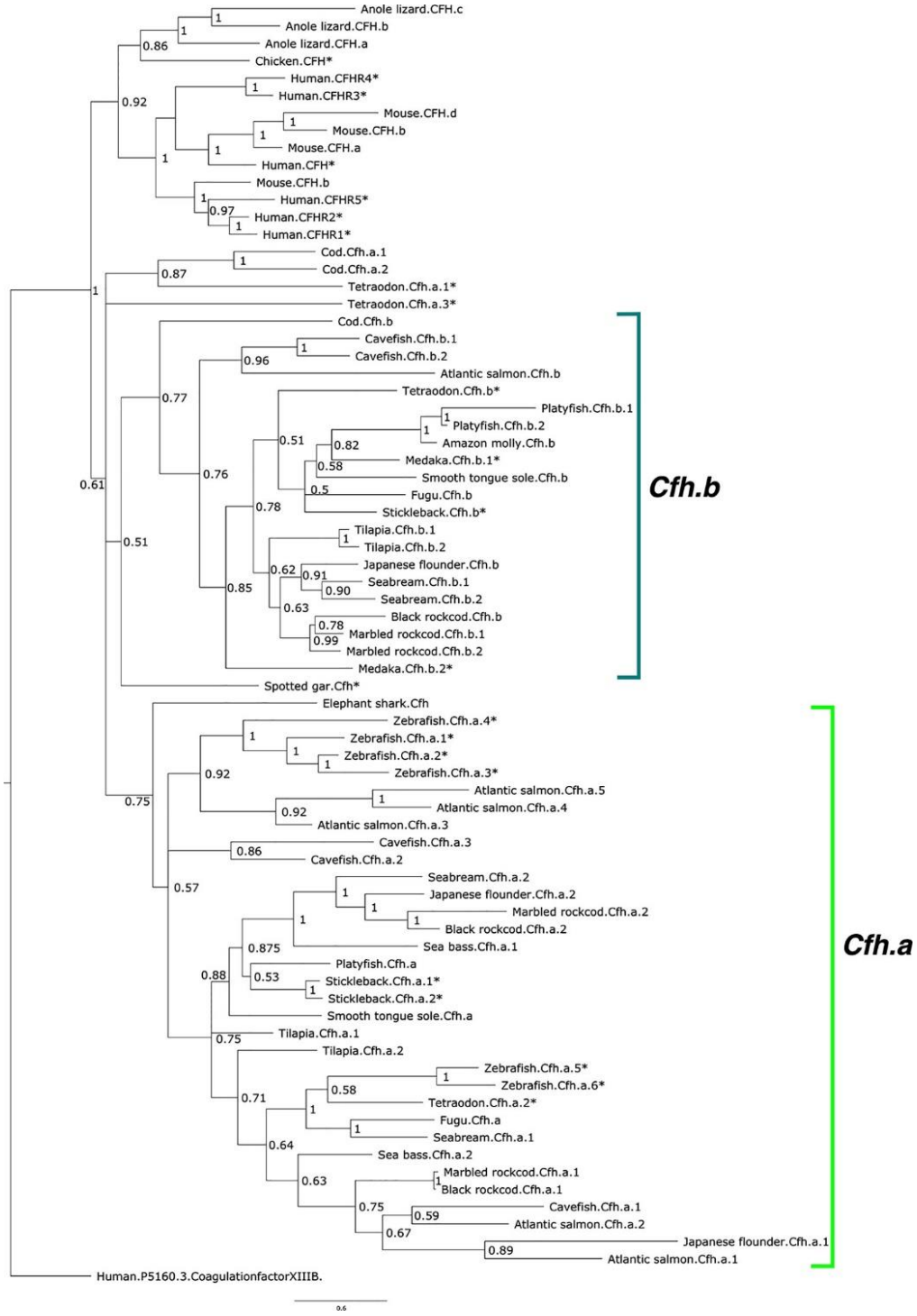


Figure 3.3. Phylogenetic tree of the fish and other vertebrate complement factor H (CFH). The tree was built with the BI method and branch support values (posterior probability values) are shown. The two major CFH clades in teleosts were named Cfha (light green) and Cfhb (green-blue) and arose from the teleost genome duplication event and the multiple genes found in each species were numbered arbitrarily. The tree was rooted with the predicted protein sequence of human coagulation factor XIII B (ENST00000367412.1). The species for which

gene linkage analysis was performed are indicated with an asterisk (*). Accession numbers of the sequences that were used to build the tree are available in **Supplementary table 3.1**. The sequence of the Xenopus Cfh was not included in the tree to simplify the analysis as clustering indicated its evolution was highly divergent from all other species.

3.4.4. Short-range gene linkage

The gene environment of *C3*, *C4*, *C5* and *CFH* genes in human was used as a reference to characterise the homologue genome regions in fish (**Figure 3.4, 3.5, 3.6** and **3.7**).

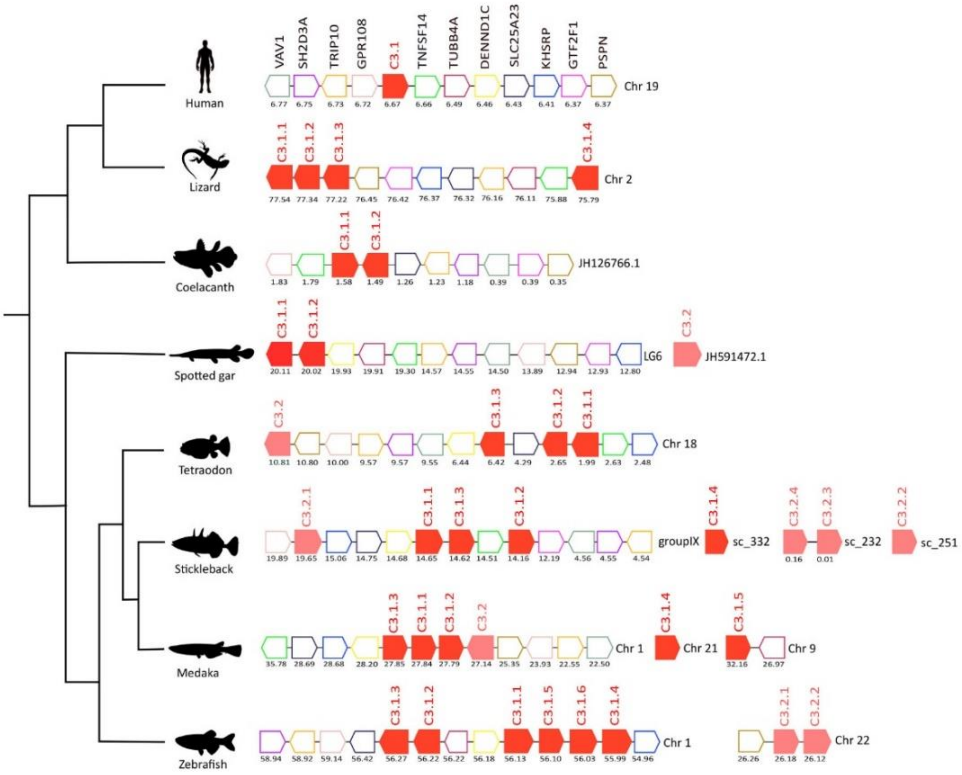


Figure 3.4. Gene synteny analysis of the C3 gene environment in tetrapods and fish. Genes predicted in the genome are represented by arrows and the arrowheads indicate the gene orientation. C3 genes are represented by full-colored arrows: C.3.1 in red and C.3.2 in pink. Neighboring gene families are represented by different colors and gene homologues are indicated in the same color. The genes are mapped on the chromosomes and their actual positions (Mega base pairs, Mbp) is indicated below. Only genes that were conserved in the homologous genome regions in all species compared are represented. Tumor necrosis factor 14 (TNFSF14), Tubulin beta-4A chain (TUBB4A), DENN domain-containing protein 1C

(DENND1C), Solute carrier family 25 member 23 (SLCA25A), Far upstream element-binding protein 2 (KHSRP), General transcription factor IIF subunit 1 (GTF2F1), Persephin (PSPN), Protein GPR108 (GPR108), Cdc42-interacting protein 4 (TRIP10), SH2 domain-containing protein 3A (SH2D3A), Proto-oncogene-vav (VAV1).

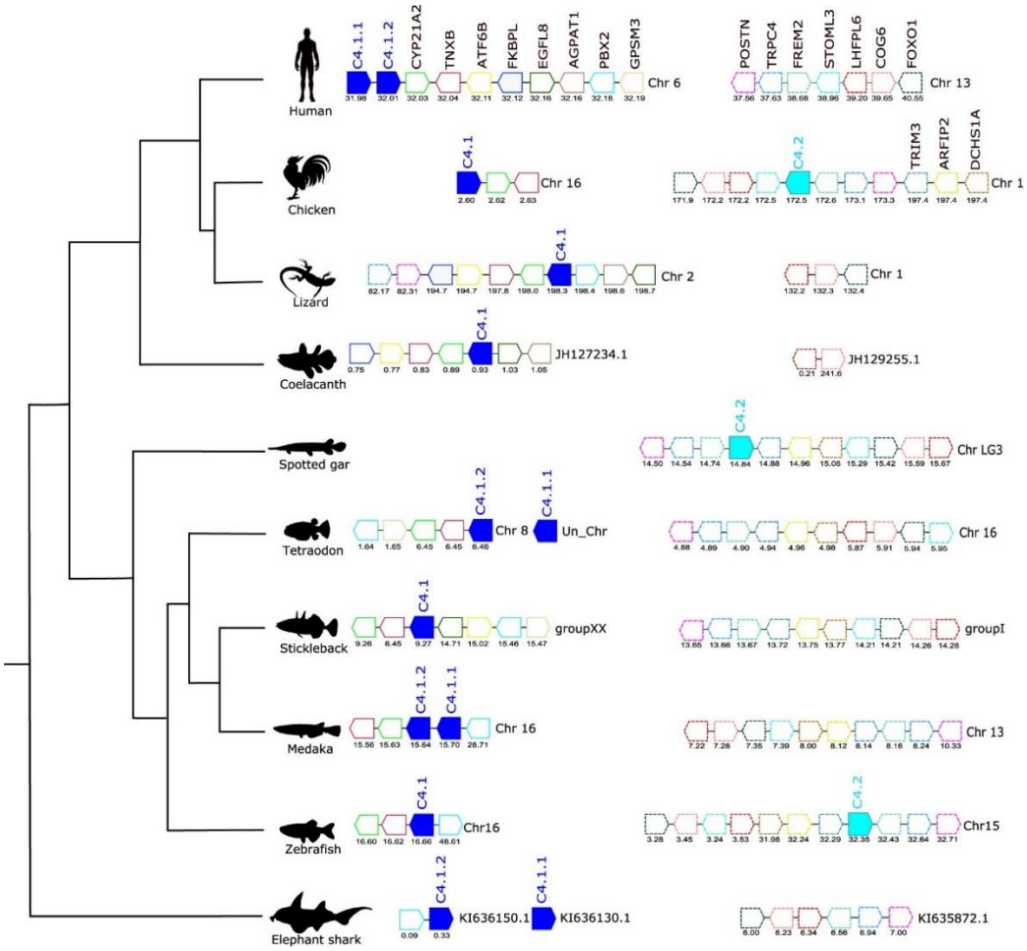


Figure 3.5. Gene synteny analysis of the C4 gene environment in tetrapods and fish. Genes predicted in the genome are represented by arrows and the arrowheads indicate the gene orientation. C4 genes are represented by full- colored arrows: C.4.1 in dark blue and C.4.2 in light blue. Neighboring genes families are represented in different colors and by continuous (C.4.1) and dashed (C.4.2) outlines and homologue genes are indicated in the same color. The genes are mapped on chromosomes based on their actual positions (Mbp) predicted in the genome assemblies. Only genes that were conserved in the homologous genome regions in all species compared are represented. No homologue genome region of human C.4 was found in the spotted gar genome. Steroid 21-hydroxylase (CYP21A2), Tenascin-X (TNXB), FRAS1-related extracellular matrix protein 2 (FREM2), Cyclic AMP-dependent transcription factor

ATF-6 beta (ATF6B), FK506-binding protein-like (FKBPL), Epidermal growth factor-like protein 8 (EGFL8), 1-acyl-sn-glycerol-3-phosphate acyltransferase alpha (AGPAT1), Pre-B-cell leukemia transcription factor 2 (PBX2), G-protein-signaling modulator 3 (GPSM3). FRAS1-related extracellular matrix protein 2 (FREM2), Short transient receptor potential channel 4 (TRPC4), Periostin (POSTN), Dachsous cadherin-related 1a (DCHS1A), Arfaptin-2 (ARFIP2), Tripartite motif-containing protein 3 (TRIM3), Stomatin-like protein 3 (STOML3), LHFPL tetraspan subfamily member 6 protein (LHFPL6), Conserved oligomeric Golgi complex subunit 6 (COG6), and Forkhead box protein O1 (FOXO1).

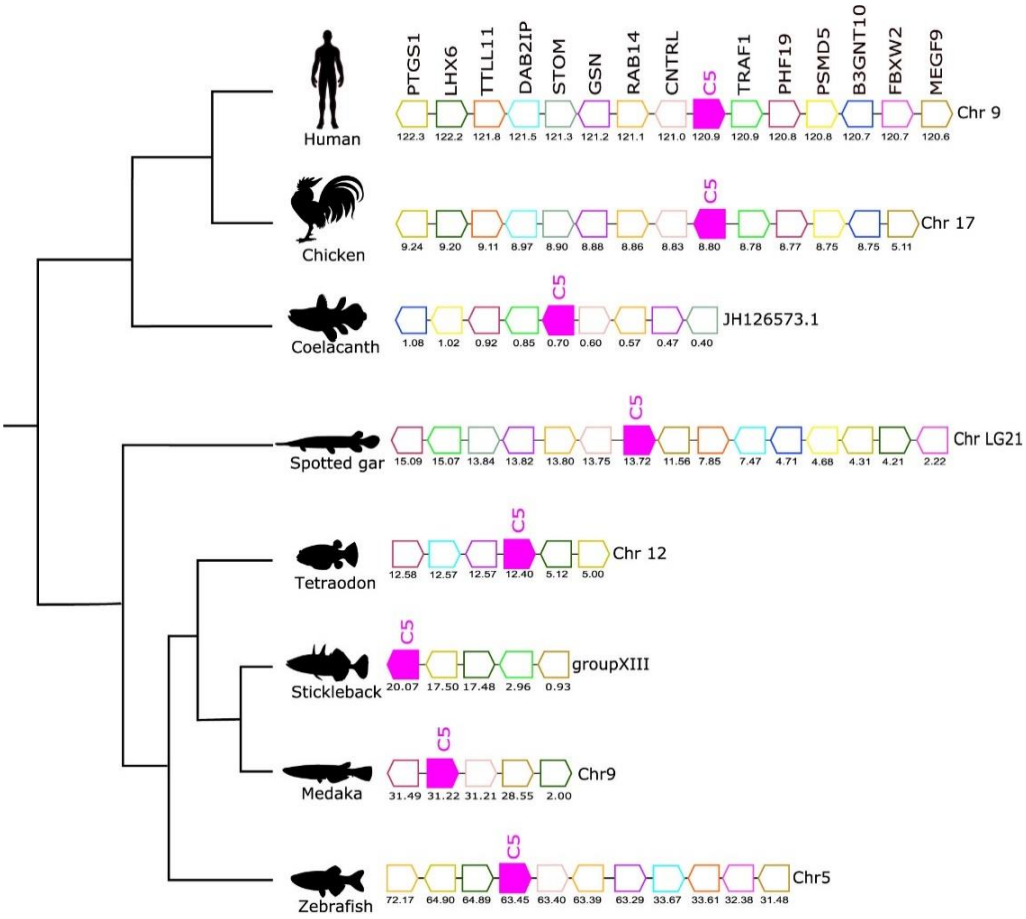


Figure 3.6. Gene synteny analysis of the C5 gene environment in tetrapod and fish. Genes predicted in the genome are represented by arrows and the arrowheads indicate the gene orientation. Neighboring genes families are represented by different colors and homologue genes are represented by the same color. The genes are mapped on the chromosomes according to their predicted positions (Mbp). Only genes that were conserved in the homologous genome regions in all species compared are represented. TNF receptor-associated factor 1 (TRAF1),

PHD finger protein 19 (PHF19), 26S proteasome non-ATPase regulatory subunit 5 (PSMD5), Hexosyltransferase (B3GNT10), F-box/WD repeat-containing protein 2 (FBXW2), Multiple epidermal growth factor-like domains protein 9 (MEGF9), Centriolin (CNTRL), Ras-related protein Rab-14 (RAB14), Gelsolin (GSN), Erythrocyte band 7 integral membrane protein (STOM), Disabled homolog 2-interacting protein (DAB2IP), Tubulin polyglutamylase TTLL11 (TTLL11), LIM/homeobox protein Lhx6 (LHX6), Prostaglandin G/H synthase 1 (PTGS1).

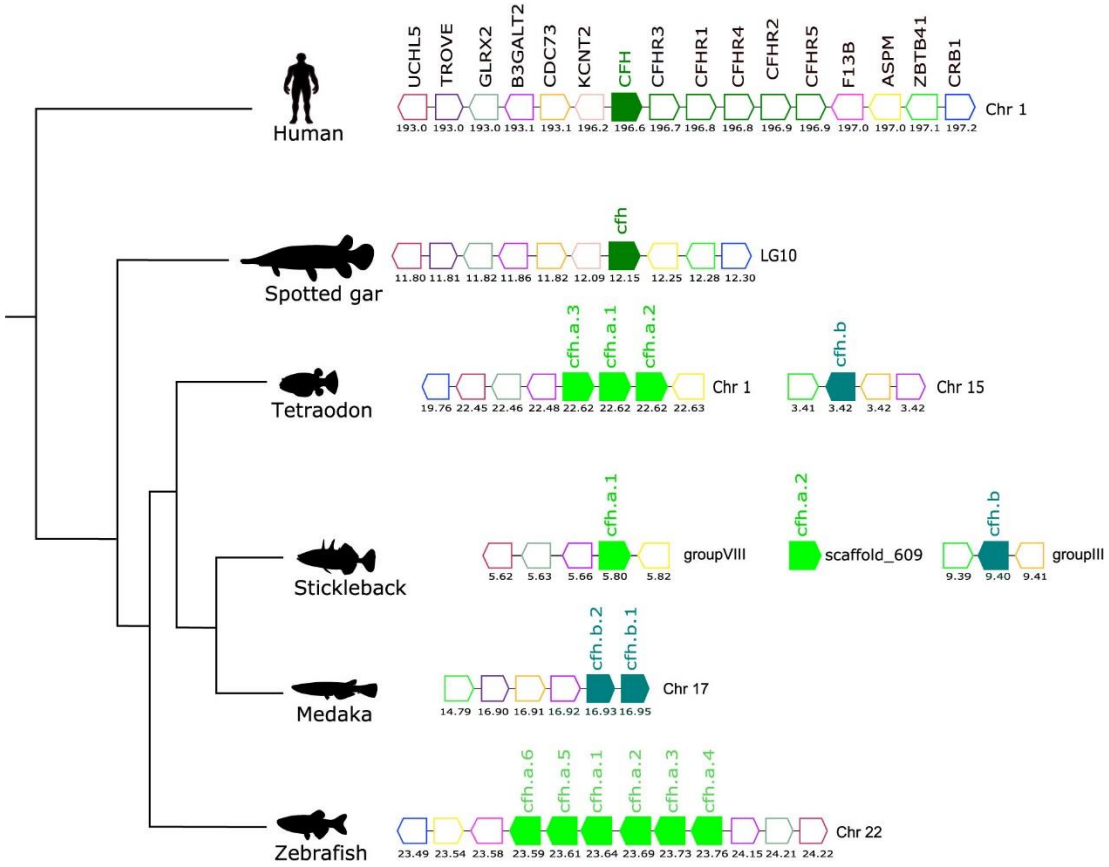


Figure 3.7. **Gene synteny analysis of the CFH/CFHR gene environment in tetrapod and fish.** Arrows represent genes predicted in the genome and arrowheads indicate the gene orientation. The vertebrate CFH genes are represented by full coloured arrows: tetrapod and spotted gar (olive-green) and the teleost duplicates, cfha (light-green), and cfhb (green-blue). Neighboring genes families are represented by different colours and homologue genes between species are represented by the same coloured arrow. The genes are mapped on chromosomes according to their predicted positions (Mbp). Only genes that were conserved in the homologous genome regions in all species are represented. Coagulation factor XIII B chain

(F13B), abnormal spindle microtubule assembly (ASPM), zinc finger and BTB domain containing 41 (ZBTB41), crumbs cell polarity complex component 1 (CRB1), potassium sodium-activated channel subfamily T member 2 (KCNT2), cell division cycle 73 (CDC73), beta-1,3-galactosyltransferase 2 (B3GALT2), Ro60, Y RNA binding protein (TROVE), Ubiquitin C-Terminal Hydrolase L5 (UCHL5).

3.4.5. Complement genes

In humans, *C3* mapped to chromosome 19 and homologues of at least 11 neighbouring genes were found in the genomes of several fish and in the lizard (**Figure 3.4**). In the human genome, a single *C3* gene persisted but in the lizard four *C3* gene copies were found and all genes mapped to chromosome 2, and three were in very close proximity suggesting that they resulted from a recent tandem duplication event. Many of the duplicate teleost *c3* genes were also arranged in tandem suggesting that they were the result of recent species-specific duplication events (**Figure 3.4**). In the zebrafish, *c3* genes were distributed on two chromosomes, the *c3.1* genes mapped to chromosome 1 while the *c3.2* genes mapped to chromosome 22. But in the tetraodon the three *c3.1* genes and the single *c3.2* gene mapped to chromosome 18 in close proximity and in stickleback *c3.2.1* was localized on the same chromosome as the *c3.1* genes (group IX). The results suggest that in the teleosts different genome rearrangements occurred for *c3.1* and *c3.2*. The other stickleback *c3.2* genes and some of the medaka *c3.1* and *c3.2* genes mapped to small genome fragments (scaffolds) that were poorly annotated. The absence of conserved gene linkage for the spotted gar *c3.2* genome region (which maps to a short scaffold JH591472.1) meant gene origin could not be assigned.

The gene environment of *C4* was also conserved across tetrapods and the fishes and *CYP21A2* and *TNXB* genes mapped near to *C4.1* (**Figure 3.5**). In human, the two *C.4* genes (*C4.1.1* and *C4.1.2*, known as *C4A* and *C4B*) mapped in tandem to chromosome 6 but in the chicken the two *C4* genes were localized on different chromosomes and *c4.1* mapped to chromosome 1 and *c4.2*, which mapped to chromosome 16 had a unique gene environment. In coelacanth, tetraodon and stickleback a single *c4* gene copy existed and the neighbouring gene environment was similar to that found in the human *C4* genes (**Figure 3.5**). In the medaka, the duplicate *c4* genes mapped in tandem (*c4.1.1* and *c4.1.2*) but in zebrafish, *c.4.1* and *c4.2*, mapped to different chromosomes. The spotted gar and zebrafish *c4.2* genome regions shared no gene homologues with the flanking region of the human *C4* genes or other teleost *c4.1* genome regions but they were similar to the flanking region of chicken *c4.2* (**Figure 3.5**). In

human, although a homologue region of the vertebrate *c4.2* gene environment was found on chromosome 13 and also in several fish, tetraodon chromosome 16, stickleback group I, medaka chromosome 13 and elephant shark scaffold KI635872.1 the *c4.2* gene was not identified and was presumably deleted from their genomes (**Figure 3.5**).

The gene environment of *C5* was conserved across the different species analysed (**Figure 3.6**). A homologous gene environment and gene order for at least eight human C5 flanking genes on chromosome 9, *TRAF1*, *PHF19*, *PSMD5*, *B3GNT10*, *CNTRL*, *RAB14*, *GSN* and *STOM*, was found in fish. Furthermore, the *C5* gene was linked to the *CNTRL* gene in most of the genomes analysed and this confirmed that in all vertebrates the *C5* gene shared a common origin (**Figure 3.6**).

3.4.6. Complement factor H members

Comparisons of the gene environment of human *CFH* with ray-finned fish *cfh* regions revealed a conserved gene neighbourhood and suggested they shared common ancestry (**Figure 3.7**). In human, the *CFH* gene mapped to chromosome 1 along with the five adjacent *CFHR* genes and formed a gene block. In fish a similar gene block was found on zebrafish chromosome 22 but in other teleost genomes the gene organization differed. In the tetraodon, three tandem *cfh* genes (*cfha*) were found on chromosome 1 but the fourth gene (*cfhb*) mapped to chromosome 15. In the stickleback the three *cfh* genes mapped to different genome regions. In medaka the two *cfh* genes (*cfhb1* and *cfhb2*) mapped in tandem on chromosome 17 (**Figure 3.7**). In addition, the teleost *cfha* genes were in linkage with *aspm*, *b3galt2*, *crb1* and *glrx2* genes and the teleost *cfhb* genes mapped in closed proximity to *trove*, *zbtb41* and *cdc73*, that neighbour human *CFH/CFHR* genes, and this provides further evidence that the extra copies of the *cfh* genes in teleosts resulted from the teleost specific whole genome duplication event (**Figure 3.7**).

3.4.7. Sequence comparisons

Sequence comparison between the deduced complement proteins revealed that the fish proteins shared high sequence similarity and conserved functional and structural domains with human homologues (**Figure 3.8, 3.9, 3.10, 3.11** and **Supplementary figure 3.3, 3.4** and **3.5**). The fish complement proteins encoded by duplicate genes also shared a well conserved structure (**Figure 3.10**, **Supplementary figure 3.6**).

The cartilaginous fish C3, C4 and C5 shared 61-65 %, 52-54 % and 65 % aa

similarity with the human homologues, respectively. The spotted gar C3s shared 64-67% aa sequence similarity with human C3 while C4 and C5 were less conserved (53 % and 58 % aa similarity, respectively). Considering only the teleosts a common pattern of aa sequence similarity was found for complement isoforms and is exemplified by the stickleback. The stickleback C3s shared approximately 54 % aa similarity with the homologue forms of human complement but the duplicate stickleback C3.1 and C3.2 were only 48% similar to each other, although the C3.1.1 and C3.1.2 isoforms were highly conserved (94% aa similar). The teleost C4 shared approximately 59 % aa similarity with human C4 and the duplicate C4.1 and C4.2 isoforms in zebrafish shared 52 % aa similarity. The C5 isoform in teleosts shared approximately 60% aa similarity.

3.4.8. The fish C3

Thirteen major domains (six within the β-chain, six within the α-chain and one shared between both chains) were characterized in human C3 (Janssen et al., 2005) and homologue conserved regions were found in C3 across the fish (Melillo et al., 2015, **Figure 3.8**). The tetra-arginine motif (4ARG, RRRR) is a proteolytic cleavage site that is localized at the end of the β-chain and separates the β-chain from the α-chain during post-translational processing. In human C3, the β-chain consisted of five complete α2-macroglobulin domains (MG1-5) and an α-chain that included an anaphylotoxin-like domain (ANATO), two complete MG domains (MG7 and MG8), a thioester domain region (TED), two parts of a CUB domain (CUBI and CUBII) localized at both sides of the TED domain and the C345C domain which contains the netrin region at the C-terminus of the sequence (**Figure 3.8**). A sixth MG domain (MG6) was shared between the β-chain and the α-chain (MG6I and MG6II). The deduced fish C3.1 and C3.2 proteins had a similar organization and homologue domains of human C3 exist and they are relatively well conserved in terms of sequence. In general, between human and fish C3, the LNK, MG7 and TED domains are the most conserved domains and MG1, ANATO and C345C domains are the most variable (**Figure 3.8**). The 3D structure of human C3 is stabilised by 13 disulphide bonds formed between 26 cysteines, 24 of which are located between MG6I and C345C and have been conserved in fish. The exception is for the cysteines in MG8 that form a disulphide bond in human (**Figure 3.8** annotated in red, **Supplementary figure 3.3**) and, which were only conserved in the coelacanth C3.1.1 and C3.1.2, spotted gar C3.2 and in the lamprey C3.x1 and C3.x2 sequences. A signal peptide was predicted in most of the sequences suggesting these are secreted but in others remain to be identified and they may

have different cellular localizations. The TED domain contained a highly conserved thioester region that is responsible for attaching the complement proteins to the surface of the pathogen (Levine and Dodds, 1990). In human, the thioester region contains a six aa motif, GCGEQN, which is totally conserved in the C3.1 proteins from cartilaginous fish, ray-finned fish and coelacanth (**Figure 3.9**). In C3.2 of the ray-finned fish, the third Gly (G) is mutated to Val (V), suggesting that functional divergence may have occurred between the gene duplicates (**Figure 3.9**). Moreover, zebrafish C3.1.5 contained two aa mutations and stickleback C3.1.2 contained a single aa modification suggesting that functional divergence may also occur between the species-specific duplicates.

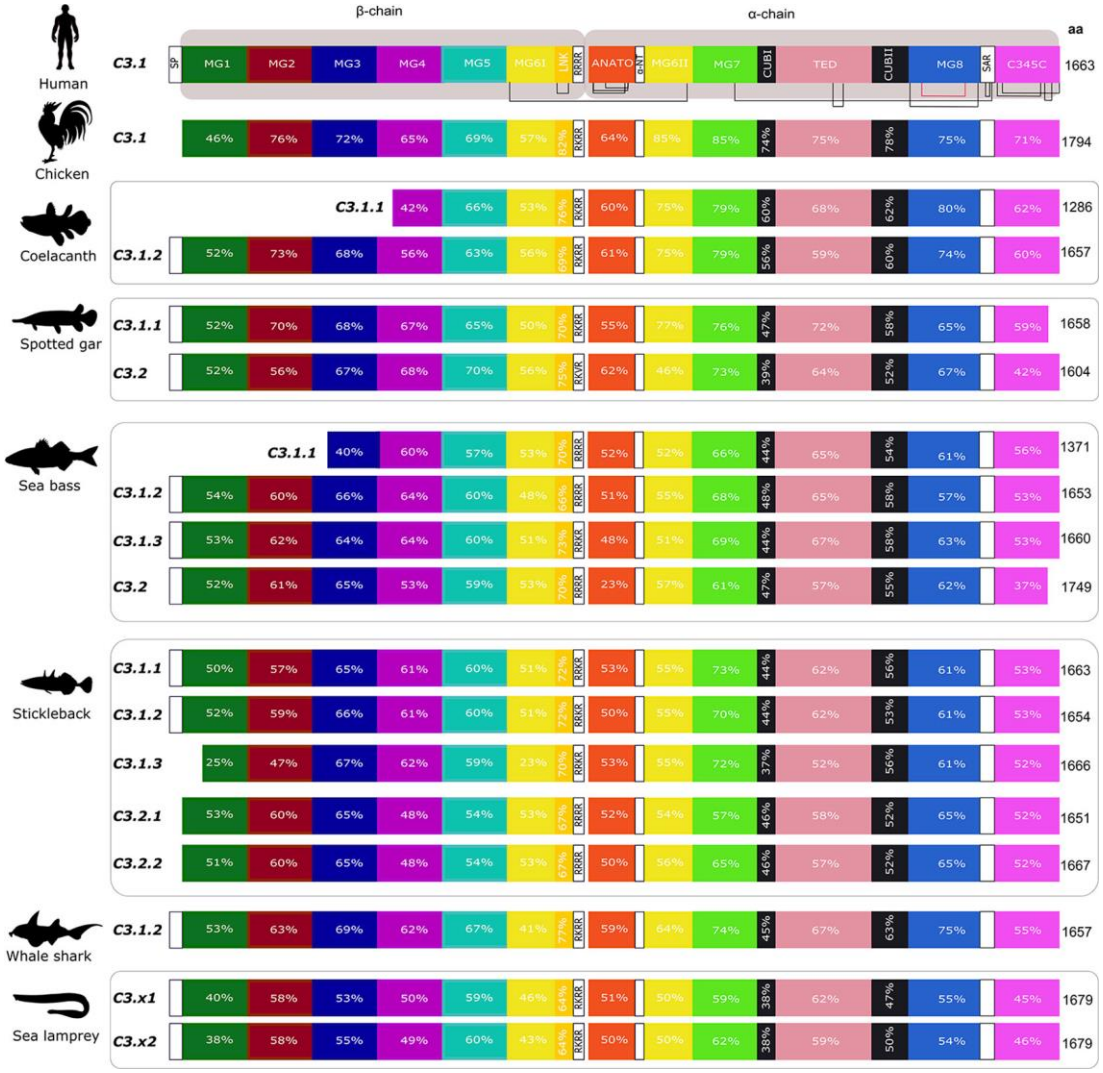


Figure 3.8. Structural comparison of the fish and tetrapod deduced C3 proteins. Thirteen protein domains were annotated according to the crystallographic structure of human C3 (Janssen et al., 2005) and homology modelling of the predicted structure of an Antarctic teleost

C3 (Melillo et al., 2015) and they are represented by blocks with different colors. The two human chains, β- (residues 1–645 aa) and α- (residues 650–1641 aa) are indicated. The homologue regions in fish and chicken were identified using a multiple sequence alignment (Supplementary figure 3.3). The localization of 13 disulphide bridges in the human C3 are indicated and the position of the 24 cysteine residues are generally conserved in other vertebrate C3's with the exception of one disulphide bridge within MG8 (annotated in red) that was only maintained in chicken, the coelacanth C3's, spotted gar C3.2 and in the lamprey C3 sequences (Supplementary figure 3.3). The percent similarity between the human domains and the homologue regions in other vertebrates is indicated as well as the homologue residues of the four arginine motif (4ARG). The low sequence similarity of the sea bass C3.2 ANATO domain with human is due to the larger length of the fish sequence. The predicted total size (aa) of the different C3 proteins is indicated. SP, signal peptide, MG- macroglobulin domain (MG1-8); LNK, link domain; ANATO, anaphylatoxin domain; α-NT, N-terminal region of the cleaved α-chain; CUB, complement C1r/C1s, Uegf, Bmp1 domain; TED, thioester-containing domain; SAR, short anchor region; C345C, carboxy-terminal domain. A SP was not predicted in some of the sequences.

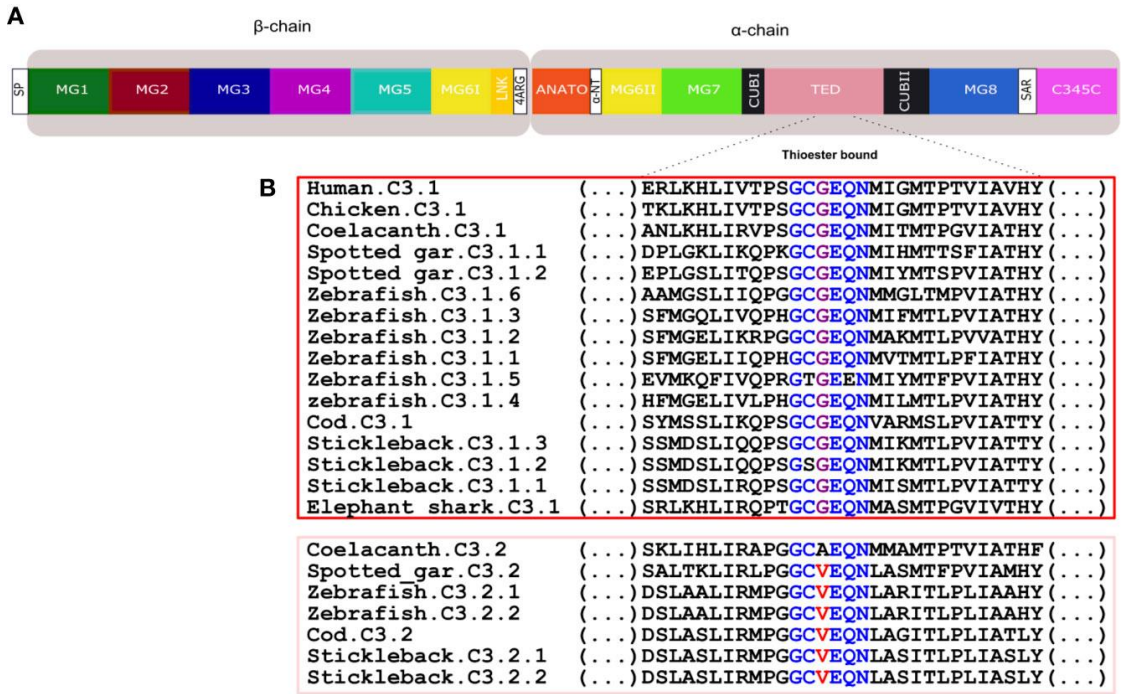


Figure 3.9. Comparison of the fish and human C3 thioester bond. (A) Schematic representation of the thirteen conserved domains of human C3. (B) Sequence comparison of

the fish, human and chicken C3 thioester bond responsible for attaching the complement proteins to the surface of the pathogen. The six aa motif forming the thioester bond are colored and conserved aa are denoted in the same color. The C3.1 (red) and C3.2 (pink) sequences were grouped to better illustrate the differences within this region. SP, signal peptide; MG, macroglobulin domain (1-8); LNK, link domain; ANATO, anaphylatoxin domain; α-NT, N-terminal region of the cleaved α-chain; CUB, complement C1r/C1s, Uegf, Bmp1 domain; TED, thioester-containing domain; SAR, short anchor region; C345C, carboxy-terminal domain.

3.4.9. The gilthead sea bream C3 isoforms

In the gilthead sea bream, the C3.1 forms are highly similar and C3.1.5 and C3.1.8 shared the highest aa similarity (97%). The C3.2 form in gilthead sea bream shared approximately 51.7% aa similarity with the C3.1 isoforms. The shortest C3.1 form was C3.1.6 (1507 aa) but all other C3 isoforms (C3.1 and C3.2) had a similar size to human C3 with the exception of C.3.2 which was longer (1772 aa).

For gilthead sea bream C3.1, five isoforms, C3.1.1, C3.1.2, C3.1.3, C3.1.5, C3.1.6 and C3.1.8 the thioester region identified in human C3 was conserved, but C3.1.4 and C3.1.7 contained aa mutations and C3.2 differed and was identical to C3.2 in other ray-finned fish (**Figure 3.10**, **Supplementary figure 3.6**). Differences within other domains were also identified; C3.1.6 and C3.1.7 lacked a full-length ANATO domain; C3.2 had a longer ANATO domain (81 aa longer) compared to C3.1 in human and other fish; the 4ARG region that precedes the ANATO domain in human was conserved in gilthead sea bream C3.2 but modified to RRKR or RKKR in C3.1; C3.2 had a longer MG1 domain compared to the human and other gilthead sea bream C3s (**Supplementary figure 3.6**). Similar to other teleosts, 24 cysteines located between MG6I and C345C are also conserved in the gilthead sea bream C3s. Additional, cysteine residues only conserved in the gilthead sea bream C3s were within MG1, MG6II and TED (**Figure 3.10**, **Supplementary figure 3.6**). N-glycosylation sites were identified in MG7, CUB2 and MG8 domains and were conserved in other fish but not human (**Supplementary figure 3.2** and **Supplementary figure 3.6**).

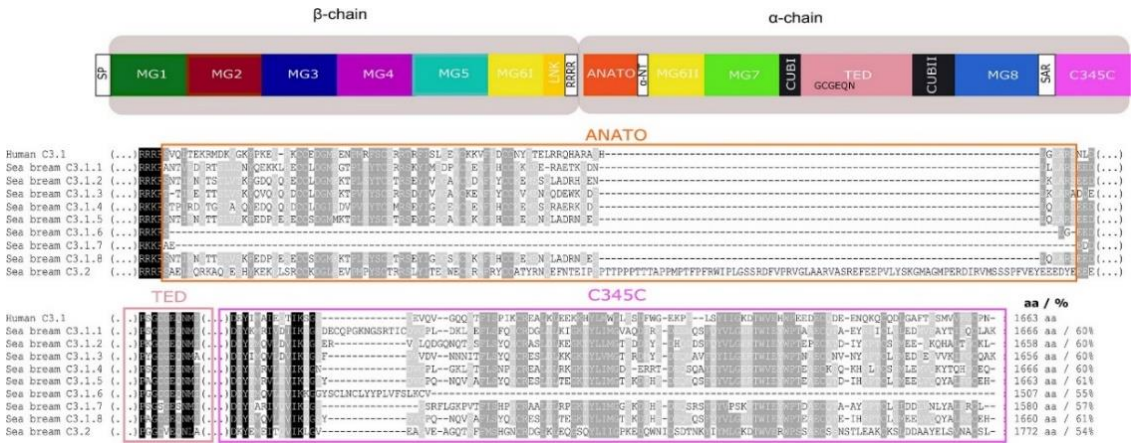


Figure 3.10. Comparison of the gilthead sea bream C3 isoforms with the human C3. The protein domains in which most differences were found ANATO, TED, and C345C and may cause functional differences between the gilthead sea bream and human and within the gilthead sea bream C3 paralogues are represented. Symbol description is available in the legend of **Figure 3.8**. The multiple sequence alignment of the full protein is available as **Supplementary figure 3.6**. Sequence conservation is denotated by background color gradient (black, total conservation; white, no conservation). The percent of similarity between the nine gilthead sea bream C3 isoforms with human C3.1 is indicated.

3.4.10. Fish C4 and C5

Fish C4 had a similar protein organization to C3 and the predicted full-length C4 sequences were similar to human C4 (**Figure 3.11A**, **Supplementary figure 3.4**). The deduced C4 proteins from cartilaginous fish through to tetrapods all possessed the thirteen major functional domains (**Figure 3.11A**, **Supplementary figure 3.4**). The two homologue regions of the human C4 tetra-arginine (RRRR) processing sites, one after LINK domains and the other within the MG8 domains were also present in the fish sequences. In the teleost C4 proteins, a signal peptide sequence was predicted with the exception of the stickleback C4.1 (**Supplementary figure 3.4**). The aa sequence of the C4 thioester domain, GCGEQT, in human was conserved in coelacanth and zebrafish C4.1. In other ray-finned fish and in the elephant shark the second Gly (G) was mutated to Ala (A) but more extensive changes existed in the stickleback C4.1 sequence (**Figure 3.11B**). The aa sequence of chicken C4.1 and C4.2 only differed at the last aa (S and T, respectively) and the C4.2 thioester motif was identical to the spotted gar and zebrafish C4.2. Within the catalytic domain the normally conserved C4.1 histidine (H), involved in the cleavage of the thioester to bind the target cells, was replaced by

an aspartic acid (D) in C4.2. For C5 the deduced protein structure in fish and tetrapods was similar and no thioester domain existed (**Supplementary figure 3.5**).

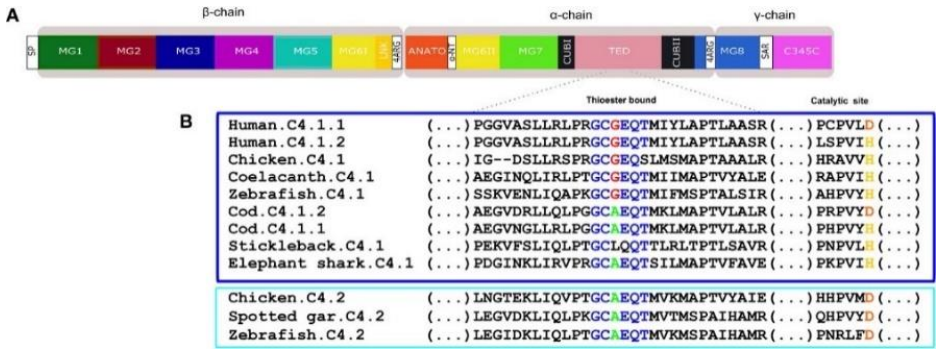


Figure 3.11. (A) Schematic diagram of the thirteen conserved domains including the additional tetra-arginine (RRRR) processing site within MG8 in human C4. (B) sequence alignment of the fish, human and chicken C4 thioester bond that is responsible for attaching the complement proteins to the surface of pathogens. The six aa motif forming the thioester bond are colored and conserved aa are denoted in the same color. The C4.1 (dark blue) and C4.2 (light blue) sequences were grouped to better illustrate the differences within this region. The catalytic site is also represented as duplicate vertebrate C4 genes were previously classified based on the substitution of the normally conserved histidine (H) in C4.1 (named as c4-H-type) by an aspartic acid (D) in C4.2 (named as c4-D type) (Nonaka et al., 2017).

3.4.11. Expression of the complement system in teleost

To assess the specialization of the complement system in fish the expression of *c3*, *c4* and *c5* gene transcripts as well as other members of the complement system and *cfh* was assessed in teleost skin by analysing available transcriptomes.

The sea bass (pelagic, temperate), rockcod (demersal, polar) and Senegalese sole (demersal, temperate) occupy different ecological niche presumably with different microbiota and it was hypothesised that this might be reflected by differences in complement expression in the skin. Analysis of available transcriptome data for sea bass, Senegalese sole and black rockcod and other fish species revealed that fewer components of the complement system were expressed in the sea bass skin relative to the other two species, although this is most likely linked to sequence coverage (**Table 3.3**). All three species expressed complement transcripts

and *c4.1*, *c6* and *c7b* were present in all the explored skin transcriptomes. Transcripts for *c3*, *c5*, *c8* and *cfh* were found in the black rockcod and Senegalese sole. In the sea bass transcripts for *c1q* and *c1s* were identified and in the black rockcod only *c1s* was identified and in the sole only *c1q* was detected. Transcripts for several different *c3* genes were detected in the skin of the rockcod. Different transcripts of *cfh*, *cfh.1* and *cfh.2*, were detected in rockcod skin and a single form, *cfh.1*, was found in Senegalese sole skin and no *cfh* transcripts were found in sea bass skin (**Table 3.3**).

Table 3.3. Expression of complement system in teleost skin.

Complement component	*Sea bass* (*Dicentrarchus labrax*)	Black rockcod (*Notothenia coriiceps*)	Senegalese sole (*Solea senegalensis*)
c1q(a/b/c), *c1s*, *c1r*	*c1qb, c1s*	*c1s*	*c1qa*
c/b2	-	*c/b2*	-
c3 (1/2)	-	*c3.1, c3.2*	*c3.1*
c4 (1/2)	*c4.1*	*c4.1*	*c4.1*
c5	-	-	c5
c6	*c6*	*c6*	*c6*
c7 (a/b)	*c7b*	*c7a, c7b*	*C7a, C7b*
c8 (α/β/γ)	-	*c8γ*	*c8α, c8β, c8γ*
c9	-	-	-
cfh (a/b)	-	*cfha, cfhb*	*Cfhb*
Assembled contigs	31.858	848.752	50.113

RNA-seq transcriptome (Illumina platform) assemblies of the European sea bass skin (GFJW00000000) (Pinto et al., 2017), Senegalese sole skin (PRJEB29449) (Pinto et al., 2019) and the *de novo* transcriptome assemblies of the intestine, skin and head-kidney of two Antarctic fish, the black rockcod and marbled rockcod (*Notothenia rossii*, **Supplementary table 3.1**), were analysed. The number of assembled contigs in each library is given.

To assess if multiple gene isoforms of *c3* were retained due to functional specialization linked to divergent tissue distribution, their expression pattern in immune-related tissue of the gilthead sea bream was characterized. Overall, *c3.1.1*, *c3.1.2*, *c3.1.3*, *c3.1.4 and c3.2* expression was mainly restricted to the liver and they were of very low abundance or undetectable in the spleen, gills or skin of the gilthead sea bream (**Figure 3.12**). In liver, *c3.1.1*

and *c3.2* were most abundant and the least abundant was *c3.1.4*. Low but detectable levels of *c3.1.2* and *c3.1.3* transcripts were detected in the spleen.

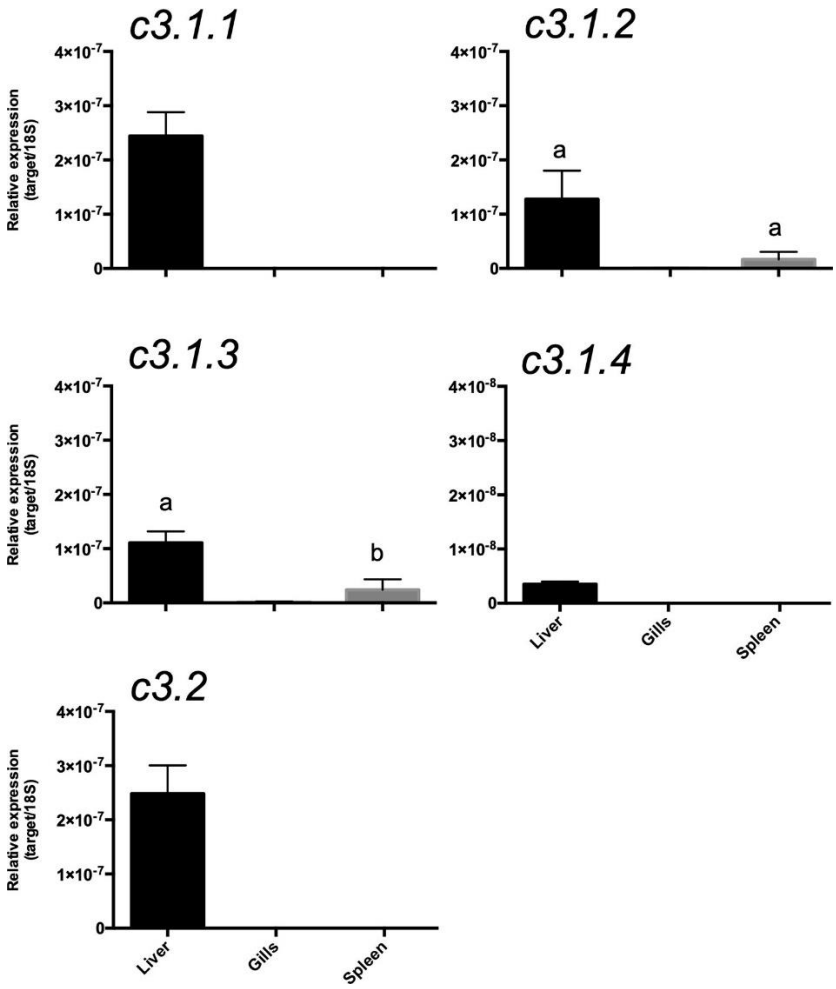


Figure 3.12. Tissue distribution of the *c3* transcripts in the gilthead sea bream (*Sparus aurata*). Gene expression levels were determined by quantitative PCR and normalized using *18s* as the reference. The results are represented as the mean ± SEM of three (n= 3) biological replicates. One-way Anova and Tukey's multiple comparison test was used to identify significant differences in transcript abundance using Prism GraphPad v5 software. Bars with different letters are significantly different ($p < 0.05$).

3.5. Discussion

The immune system in fish is proposed to have contributed to their evolutionary success in the face of innumerable aquatic pathogens (Anderson, D., & Zeeman, 1995). The expansion of *c3* and *cfh* genes in teleosts suggests that the alternative complement pathway was

favoured during evolution. The direct activation of *c3* by pathogens in the environment and the protection this confers may partly explain why teleosts are the most successful of the extant vertebrates (Nelson et al., 2016). Overall, the results suggest that a more complex complement system and innate response evolved in the fish that inhabit a microbial rich aquatic environment compared to terrestrial vertebrates (Kato et al., 2003; Papanastasiou and Zarkadis, 2005) and this potentially increased their capacity to recognize a broader spectrum of pathogens.

The evolution of vertebrates was preceded by two successive major genome tetraploidization events 1R and 2R (Papanastasiou and Zarkadis, 2005; Nakatani et al., 2007) which contributed to increase gene diversity and genome complexity. These genome duplication events occurred very early at the vertebrate origin and most likely before the cyclostome-gnathostome divergence (Cardoso et al., 2020; Putnam et al., 2008; Sacerdot et al., 2018). Our phylogenetic and gene neighbourhood analysis suggests the fish *c3* genes shared a common ancestral origin with the human and other tetrapod genes. The identification in the spotted gar (*c3.2*) of the duplicate family members previously proposed to be teleost specific (Forn-Cuní et al., 2014) suggests that *c3* gene duplication occurred earlier than previously thought potentially at the root of the vertebrate radiation and only persisted in the ray-finned fish (**Figure 3.13**). In teleost genomes both *c3.1* and *c3.2* genes were found and the larger number of *c3.1* genes retrieved in relation to *c3.2* suggests they evolved at different rates within each species. The evolution of the duplicate *c3* genes is currently difficult to resolve due to their species-specific evolution and lack of conserved gene linkage. In lamprey and hagfish duplicate *c3* genes were also found and our phylogenetic analysis revealed that they radiated basal to *c3.1* and *c3.2* in the ray-finned fish. However, gene orthology between cyclostomes and gnathostome genes is uncertain due to the selective pressure under which lamprey and hagfish genomes have been evolving especially within the gene coding regions (Kuraku and Kuratani, 2006; Smith et al., 2013).

The multiple gene copies of the teleost *c3.1* are arranged in tandem in the genomes of some species or in different chromosome region in others, suggesting gene evolution was distinct and probably a consequence of species-specific chromosome rearrangements (Warren et al., 2015). In zebrafish, in which the *c3* duplicates have previously been characterized the two *c3.2* proteins were associated with the regulation of the proinflammatory response, while the six *c3.1* copies had a different response (Forn-Cuní et al., 2014). In general the predicted C3 proteins in teleosts are of a similar size to human C3, as revealed by the *c3* transcripts previously isolated from the rainbow trout and zebrafish and in the present study from gilthead

sea bream (Forn-Cuní et al., 2014; Sunyer et al., 1997; Sunyer et al., 1996). In the common carp (*Cyprinus carpio*) where multiple C3 isoforms have also been described functional divergence of their catalytic activities was reported (Nakao et al., 2000). One C3 variant (C3-Q2) lacks most of the beta-chain and is not functional (Nakao et al., 2000) and this phenomenon may be common in teleosts since many of the species analysed in the present study had incomplete *c3* genes. Further work will be required to better characterise the incomplete *c3* genes (pseudogenes or genome assembly errors) and to better establish the functional diversity of the *c3* gene and protein complex in teleosts.

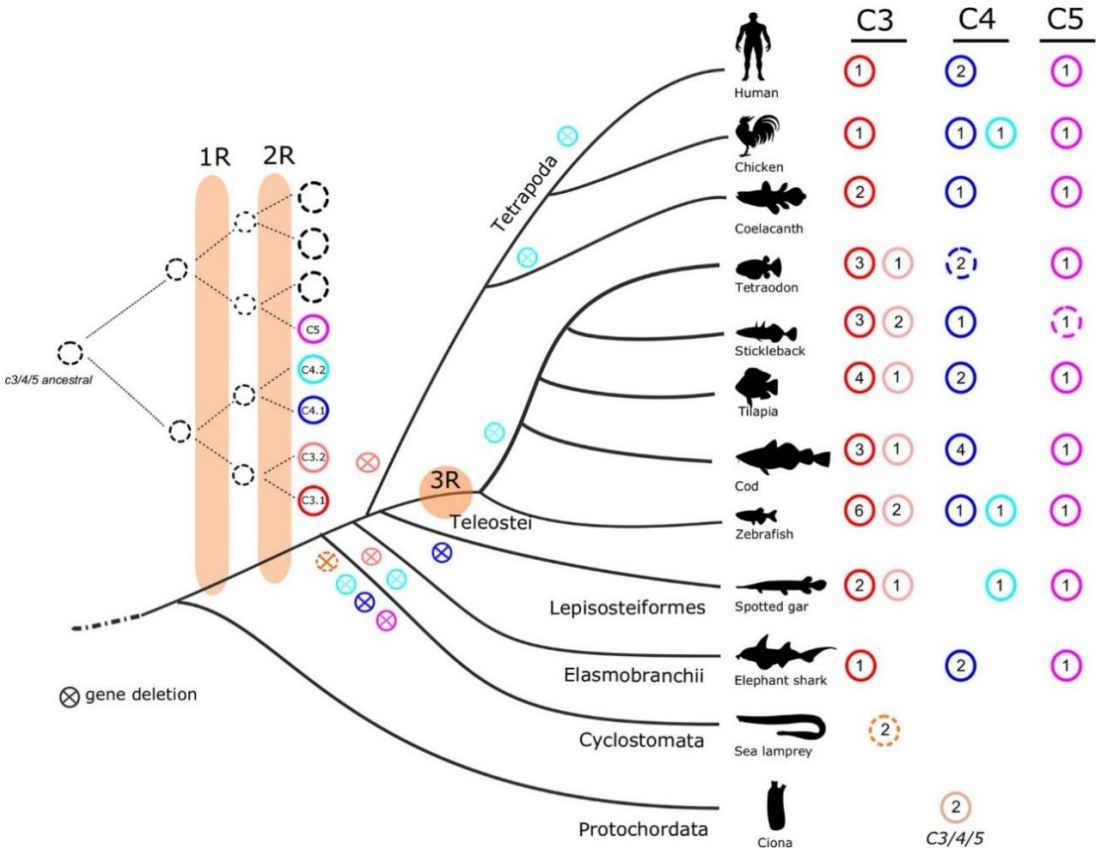


Figure 3.13. Proposed model for the evolution of the vertebrate C3, C4, and C5 genes. The number of full-length c3 (c3.1 in red; c3.2 in pink), c4 (c4.1 in dark-blue; c4.2 in light-blue), and c5 (purple) genes are presented with colored circles. Open circles indicate incomplete sequences. The cross within the circle signifies absence or elimination during evolution. The reconstruction model suggests that vertebrate c3, c4, and c5 evolved from two independent ancestral molecules prior to the vertebrate radiation and that the duplicate vertebrate c3 and c4 genes resulted from the two rounds of genome tetraploidization (1R and 2R). The teleost specific genome duplication is indicated by 3R. The lamprey c3 sequences are represented by

a dashed orange circle as homology for gnathostome c3.1 and c3.2 is uncertain. The existence in tunicate (basal deuterostome) of c3/c4/c5-like molecules suggest that duplication of c3/c4 and c5 ancestral molecules occurred prior to the 1R and 2R.

The gilthead sea bream had nine *c3* gene transcripts highly expressed in liver although as reported in other species, extra-hepatic expression also occurs (Forn-Cuní et al., 2014; Khansari et al., 2018; Qi et al., 2011). Sequence comparison of the gilthead sea bream *c3* isoforms revealed they have a similar length and structure, although C3.1.7 lacked the anaphylatoxin domain and C3.1.6 lacked both the anaphylatoxin domain and had a shorter C345C domain. These characteristics taken together with the different number of conserved cysteine residues and N-glycosylation sites in C3 isotypes emphasises their potentially different structure and function. This idea is reinforced by studies of C3 proteins in gilthead sea bream plasma and five C3 isotypes were reported that were functionally and structurally divergent (Sunyer et al., 1997). Comparison of the gene transcripts identified in the present study with the short N-terminal sequence for the C3 plasma protein previously reported indicated they are likely to correspond to C3.1.1 (C3-3) and C3.1.2 (C3-4) and the remaining three proteins, which included the most abundant, C3-1 and C3-2 proteins, had in general the greatest similarity with our C3.1.3, C3.1.5 and C3.1.8 transcripts and it was not possible to assigned sequence correspondence. Previously published *c3* gene expression studies in immune challenged sea bream (**Table 3.4**) have assessed sea bream *c3.1.5* (HM543456 and CB184637) and *c3.1.8* (CX734936), that were not analysed in our study. The sometimes-contradictory results reported for *c3* gene transcripts may be explained by differences in the *c3* gene amplified. The number of *c3* isoforms with distinct structures and expression in gilthead sea bream as well as in other teleost species highlights the need for further in-depth studies of *c3* in fish.

The variable number of *c3* genes identified in our study and their different organization in teleost genomes suggests that they probably evolved under different pressures. The results of our study indicate that maintenance in the genome of multiple *c3* gene copies in teleosts is unlikely to be due to their divergent tissue distribution. However, variation in the sequence and structure of the deduced teleost C3 proteins and the functional diversity demonstrated in previous studies has presumably favoured their maintenance in the genome. Furthermore, since the bacterial cell wall triggers the hydrolysis of C3 and activation of the alternative complement pathway, we hypothesise that the high diversity of microorganisms in aquatic systems potentially increased the pressure on *c3* isoform evolution in fish (Grossman

and Leive, 1984; Li et al., 2015; Meng et al., 2012; Santoro et al., 1979; Schreiber et al., 1979; Sunyer et al., 1998).

Table 3.4. Expression pattern determined by qPCR of the different gilthead sea bream c3 isoforms obtained in the present study compared with previous studies.

Gilthead sea bream c3 isoforms									
c3.1.1	*c3.1.2*	*c3.1.3*	*c3.1.4*	*c3.1.5*	*c3.1.6*	*c3.1.7*	*c3.1.8*	*c3.2*	*Ref.*
Li	Li, Sp	Li, Sp	Li	nd	nd	nd	nd	Li	Present study
-	-	-	-	-	-	-	Li, In, Hk	-	(Reyes-Becerril et al., 2008)
-	-	-	-	Go	-	-	-	-	(Chaves-Pozo et al., 2008)
-	-	-	-	-	-	-	Li, Sp, In, Hk	-	(Reyes-Becerril et al., 2011)
-	-	-	-	Li, In, Br	-	-	-	-	(Mauri et al., 2011a)
-	-	-	-	-	-	-	Li, In, Hk, Sk	-	(Reyes-Becerril et al., 2012)
-	-	-	-	-	-	-	Hk, In	-	(Guardiola et al., 2012)
-	-	-	-	-	-	-	Go	-	(Valero et al., 2015)[¶]
-	-	-	-	-	-	-	Sp,Sk	-	(Reyes-Becerril et al., 2017)[‡]
-	-	-	-	In,Gi, Sk	-	-	-	-	(Khansari et al., 2018)[ɖ]
-	-	-	-	Li,Sp	-	-	-	-	(Khansari et al., 2019)
				In, Hk			In, Hk		(Leiva‐Rebollo et al., 2020)[?]

nd: Not determined

[¶]: in vitro treatment with poly I:C

‡: injection with lysate and ToxA
ɸ: after biotic (*Vibrio anguillarum* bacterin), abiotic (air exposure), or combination of both stressors
?: Primers may amplify both *c3.1.5* and *c3.1.8*

The existence of duplicate *c4* genes in fish and chicken suggests that gene duplication occurred prior to their divergence but that the duplicate *c4.2* gene did not persist in all fish species (**Figure 3.13**). Duplicate *C4* genes have previously been described in vertebrates including fish (Nakao et al., 2011; Nonaka et al., 2017) and were named C4 H-type and C4 D-type based on the replacement within the catalytic site of the conserved histidine (H) by an aspartic acid (D, Nonaka et al., 2017). Both genes were proposed to result from two independent duplications: one in the cartilaginous fish lineage and a second in the bony vertebrate lineage that was followed by species-specific duplications in mammals (Nonaka et al., 2017). By analysing the C4 gene environment from cartilaginous fish to human, an alternative evolutionary scenario was established in which C4.1 (previously known C4-type H) and C4.2 (previously known C4-type D) arose from a single gene/genome duplication that occurred early during the vertebrate radiation and prior to the gnathostome divergence **(Figure 3.13**). In the present study the persistence of both *C4* genes occurred in the genome of very few species and generally only the *C4.1* gene copy was retained and subsequently duplicated in a species-specific manner.

The insightful study of toll-like receptor (Tlr) evolution in cod (Solbakken et al., 2016), proposed that the acquisition of multiple isoforms was related to palaeoclimatic events. By mapping the Tlr repertoire of 76 teleosts onto a time calibrated species phylogeny it was found that in Gadiformes, selection was towards a higher copy number optima for some, but not all the Tlrs. It was proposed that highly variable palaeoclimatic arctic conditions and variable pathogen loads and pathogen community composition might explain Tlr gene evolution in Gadiformes. In contrast to the Tlr analysis, in our study the *c3* genes from Arctic Atlantic cod and Antarctic marbled and black rockcods, did not undergo large copy number expansion as occurred in other teleosts. Instead *c3* gene copy number tended to be higher in teleosts from warmer waters (tropical and subtropical regions). Based on our results it was difficult to identify a clear association between temperature and complement system evolution in the fish unlike what was shown for TLR evolution in Gadiformes. Nonetheless, in general the ectotherms had multiple *c3* genes (e.g. ray-finned fish, amphibians and reptile) compared to the single *c3* genes retrieved from the chicken and human genomes and further studies will be required to explain the basis of this difference. The other complement system members do

not appear to be under the same evolutionary pressure as *c3* since gene copy number was similar between the different fish species analysed (**Figure 3.1**). The alternative complement pathway is reported to be at least 60 times more active in teleosts than in mammals (Yano et al., 1988) and we hypothesise that this may be explained by the multiple *c3* genes and their protein products present in fish. The high C3 gene family expansion in fish relative to other vertebrates was also accompanied by an expansion of the regulatory protein Cfh and it will be of interest in the future to establish if specific partnerships between C3 forms and regulatory Cfh isoforms have arisen.

The importance of innate immunity in teleosts is not only associated with *c3* duplications but also with the expansion of other innate immunity related genes including the toll-like receptors, NOD-like receptors (NLRs), RIG-I like receptors (RLRs), antimicrobial peptides (AMPs), lectin proteins, and interferons (INFs, Solbakken et al., 2017; Zhang et al., 2015; Zhu et al., 2013). The functional significance of the species-specific *c3* gene duplication events identified in teleosts was not established but may be linked to the dominant role of the innate immune response (Solbakken et al., 2017). This notion seems to be supported by the less abundant *c3* species-specific duplications in cartilaginous fish (only two *c3* duplicates were found in whale shark and a single genes was retrieved from the elephant shark genome) in which alternative adaptive immune strategies have evolved (Buonocore and Gerdol, 2016). In fact, the elasmobranchs have a 10-times higher immunoglobulin repertoire, associated, in part, with the process of affinity maturation, when compared with teleosts, which have a relatively minor secondary antibody response (Bengtén and Wilson, 2015; Voss and Sigel, 1972; Watts et al., 2001). We hypothesise that the large expansion of the *c3* genes in teleosts may have compensated for their weaker secondary antibody response. This also presumably explains why the titre of the alternative complement pathway is several orders of magnitude higher in teleosts than in mammals (Sunyer et al., 1998; Sunyer and Tort, 1995). Furthermore, the broad spectrum of recognition and activation of teleost C3 by red blood cells from a diversity of organisms has been linked to the differing specificity of the multiple C3 isoforms (Sunyer and Tort, 1995). The higher number of *c4* genes in cartilaginous fish relative to teleosts is intriguing. Multiple *C4* genes also exist in human and chicken that only have a single *C3* gene and it may be that in these species the classical or MBP-lectin complement pathways is of greater importance.

Adaptation of fish to multiple environments favours their contact with a diversity of pathogens and fish skin plays an important role in maintaining physiological homeostasis as it is the first physical barrier that protects against pathogens. Many immune factors including

immunoglobulins (Ibarz et al., 2013; Salinas et al., 2011), lectins (Tsutsui et al., 2011), antimicrobial peptides (Bergsson et al., 2005; Cole et al., 1997), C-reactive proteins (Alexander and Ingram, 1992), proteases (Aranishi and Nakane, 1997; Salles et al., 2007), lysozymes (Hikima et al., 2003), transferrin (TF, Ibarz et al., 2013) and also several components of the complement system (Ao et al., 2015; Cordero et al., 2015; Easy and Ross, 2009; Gonzalez et al., 2007; Ibarz et al., 2013; Malachowicz et al., 2017; Provan et al., 2013; Sanahuja and Ibarz, 2015; Valdenegro-Vega et al., 2014) are secreted into the skin mucous and establish a specific chemical barrier (Ángeles Esteban, 2012; Nigam et al., 2012; Raj et al., 2011). The contribution of complement to the protective function of the skin is clear and previously published studies aimed at characterising the overall contribution of fish skin and mucous to immunity and the pathogen response revealed that members of the alternative (C3), classical (C1, C2, C4) and lytic (C5, C6, C7) pathways in skin are modulated by parasites, microorganisms and stressful conditions (**Table 3.5**). The control skin transcriptome data used in the present study revealed differences in complement pathway representation in adult sea bass, rockcod and sole (demersal, mud dwelling). However, the lower complement gene representation (eg. *c3*, *c5*, *c8*, *c9* were missing) in the sea bass skin transcriptome is unlikely to be of biological relevance and is most likely due to the characteristics of the sequencing dataset, particularly since *c3* has previously been reported in the sea bass skin proteome (Cordero et al., 2015). The generalized presence of the three complement cascade pathways in fish skin suggests they may be relevant biomarkers to assess the impact of the environment on teleost immune physiology.

Table 3.5. Complement system members in published transcriptome and proteome data available for teleost skin.

Complement	Teleost species	Study	Conditions	Reference
C4, C5, Bf/C2	*Cyprinus carpio*	Expression analysis	-	(Gonzalez et al., 2007)
C1(q, r, s), C3, C4-B, C6, C7, C8 (α, β, γ), C9	*Larimichthys crocea*	Skin mucus proteome	Air exposure	(Ao et al., 2015)
C3, C1q	*Dicentrarchus labrax*	Skin mucus proteome	-	(Cordero et al., 2015)
C3	*Salmo salar*	Infected skin proteome	Infected with sea lice *(Lepeophtheirus salmonis)*	(Easy and Ross, 2009)
B/C2, C3, C5, C6, Factor B	*Salmo salar*	Epidermal mucus proteome	Infected with sea lice (*Lepeophtheirus salmonis*)	(Provan et al., 2013)
C3, C9	*Salmo salar*	Skin mucus proteome	Infected with amoebic gill disease	(Valdenegro-Vega et al., 2014)
C3	*Oncorhynchus mykiss*	Skin mucus		(Gomez et al., 2013)
C1s, C2, C3, C4, C5, C7, C8(α, β, γ), C9	*Epinephelus coioides*	Skin transcriptome	Infected with *Cryptocaryon irritans*	(Hu et al., 2017)
C1q	*Sparus aurata*	Skin mucus proteome	-	(Sanahuja and Ibarz, 2015)
C1, C3, C4, C5, C6, C7, C8	*Salmo trutta*	Skin transcriptome	-	(Malachowicz et al., 2017)
C3	*Sparus aurata*	Skin mucus proteome	-	(Cordero et al., 2016)

3.6. Conclusion

Herein we provide a comprehensive evolutionary analysis of the complement system in teleosts compared to other vertebrates. Our phylogenetic analysis corroborates and extends previous *in silico* studies in which the early radiation of C5 and the later divergence of C3 and C4 was proposed based on a limited number of vertebrate species and the resemblance of the hagfish sequence to vertebrate C3/C4 (Hughes, 1994). Our study confirmed that *c3* and

c4 shared common ancestry prior to the vertebrate radiation and that the duplicate *c3* and *c4* genes identified potentially arose during the two rounds of vertebrate genome duplications. In basal deuterostomes and other invertebrates putative *c3*/*c4*/*c5* genes exist that diverged from the same ancestral molecule as the vertebrate members (Chen et al., 2018; Marino et al., 2002; Peng et al., 2017). Our results based on sequence clustering confirm the standing hypothesis that the ancestral *c3*/*c4*/*c5* gene duplicated prior to the origin of the vertebrates which led to the emergence of the vertebrate *c5* ancestral gene and *c3/c4* ancestral genes (**Figure 3.13**). Both molecules expanded during the two rounds of genome tetraploidization (1R and 2R) and one originated the vertebrate *c5* and the other the *c3* (*c3.1* and *c3.2*) and *c4* (*c4.1* and *c4.2*) gene members which subsequently underwent distinct lineage and species-specific gene duplication and deletion events. In our analysis the previously identified hagfish and lamprey sequences grouped within the vertebrate C3 cluster suggesting C4 and C5-related forms do not exist (Matsushita, 2018).

A large expansion of the *c3* genes as well as the main regulatory factor of the alternative pathway, *cfh* was identified in all ray-finned fish genomes probably associated with lineage and species-specific duplications suggesting that the complement cascade in teleosts is more complex than in terrestrial vertebrates and that a variety of parallel pathways may exist probably to compensate for the less developed acquired immune response. Tissue expression of complement in the gilthead sea bream indicated that in common with other vertebrates duplicate *c3* genes are primarily expressed in the liver. The common tissue distribution of the *c3* gene transcripts in sea bream suggests that their maintenance in the genome is probably due to functional divergence as revealed by differences in their sequence and structure. The expression of multiple gene transcripts for the *c3* gene in teleosts, highlights the need for new studies to evaluate the response and role of the multiple genes in fish. To conclude, our study is the first comprehensive study that provides a detailed description and comparative evolutionary analysis of C3/C4/C5 and a regulatory factor of the alternative pathway, *cfh*, in deuterostomes with specific attention to fish. Furthermore, the expression and sequence analysis of the multiple C3 isoforms in teleosts highlights the need for functional studies to understand their role and explain why they persisted in the genome.

3.7. See ANNEX III for Supplementary materials

Supplementary table 3.1. Species names, accession numbers and genome localization of the fish and other vertebrate C3, C4, C5, and C1, C2, C6–C9 genes. Genes were named based on phylogenetic clustering (**Supplementary figures 3.1-3.3**) and conservation of the gene environment (**Figures 3.4-3.6** and annotations available from the species genome assemblies). This is only available in digital format in Annex III because the table is very extensive.

Supplementary figure 3.1. Phylogenetic trees of the fish C1, C2, C6-C9 genes. Accession numbers of the sequences are available in **Supplementary table 3.1**. The trees were built with the BI method and posterior probability values are shown. This is only available in digital format in Annex III because the table is very extensive.

Supplementary figure 3.2. Phylogenetic tree of the fish C3, C4 and C5 members with other vertebrates. The tree was built with the ML method and bootstrap branch support values are shown. The two major C3 clades were named C3.1 and C3.2 and duplicated genes for each species were numbered. The two C4 forms were named C4.1 and C4.2. Tree was rooted with the deduced protein sequences from human α-macroglobulin (A2M, ENST00000318602.12) and cluster of Differentiation 109 (CD109, ENST00000437994.6). Sequence accession numbers are available in **Supplementary table 3.1**. This is only available in digital format in Annex III because the table is very extensive.

Supplementary figure 3.3. Multiple sequence alignment of the vertebrate C3 deduced proteins. The thirteen protein domains were annotated according to the crystallographic structure of human C3 (48) and homology modelling of the predicted structure of an Antarctic teleost C3 (32). SP- signal peptide (highlighted in bold), MG- macroglobulin domain (MG1-MG8); LNK-link domain; ANATO- anaphylatoxin domain; α-NT- N-terminal region of the cleaved α-chain, CUB- complement C1r/C1s, Uegf, Bmp1 domain; TED- thioester-containing domain; SAR- short anchor region; C345C- carboxy-terminal domain. The cysteine residues are highlighted in yellow and the small, colored triangles of the same color represent a predicted disulphide bond for human C3.1 available from Uniprot. The conserved position of putative N-glycosylation sites (N-x-T/S, where x represents any amino acid) are highlighted in green.

Sequence conservation is denotated by background color gradient: black- total conserved; white- no conservation. A schematic representation is available in **Figure 3.2**. This is only available in digital format in Annex III because the table is very extensive.

Supplementary figure 3.4. Schematic representation of the predicted structural domains of C4 proteins (full-length and incomplete) from selected fish (spotted gar; three teleost's the stickleback, sea bass, zebrafish; coelacanth; whale shark). The human and chicken sequences were included for comparisons. The thirteen conserved protein domains are annotated based on the crystallographic structure of the human C3 and represented by boxes with different colors (Janssen et al., 2005). The chicken and fish domains were predicted based on a multiple sequence alignment and conserved aa position. The consensus aa sequence of the thioester motif involved in attachment of complement to the target cells (Levine and Dodds, 1990) is indicated. The size (aa) of the sequences represented is indicated as well as the two four arginine residue processing sites at the end of the β- and α-chain. Full-length sequences include β-, α- and γ-chains. The signal peptide (SP) was predicted for some full-length sequences. MG-macroglobulin domain (1–8); LNK- link domain; ANATO- anaphylatoxin domain; α-NT- N-terminal region of the cleaved α-chain; CUB- complement C1r/C1s, Uegf, Bmp1 domain; TED-thioester-containing domain; SAR- short anchor region; C345C- carboxy-terminal domain.

Supplementary figure 3.5. Schematic representation of the predicted structural domains of C5 proteins (full-length and incomplete) from selected fish (spotted gar; three teleost's the stickleback, sea bass and zebrafish; coelacanth; whale shark). The human and chicken sequences were included for comparisons. Twelve conserved protein domains were predicted and are represented by boxes with different colors. The thioester-containing domain is absent from C5. The size (aa) of the sequences represented is indicated as well as the four arginine residue processing sites at the end of the β-chain. Full-length sequences include both β- and α-chains. The signal peptide (SP) was predicted for some full-length sequences. MG-macroglobulin domain (1-8); LNK- link domain; ANATO- anaphylatoxin domain; α-NT- N-terminal region of the cleaved α-chain; CUB- complement C1r/C1s, Uegf, Bmp1 domain; TED-thioester-containing domain; SAR- short anchor region; C345C- carboxy-terminal domain. This is only available in digital format in Annex III because the table is very extensive.

Supplementary figure 3.6. Multiple sequence alignment of the gilthead sea bream deduced C3 isoforms with the human C3. The 13 identified protein domains were annotated according to the crystal structure of human C3 (48) and homology modelling of the predicted structure of Antarctic teleost C3 (32). The two C3 chains β- (residues 1–645 aa) and α- (residues 650–1641) are indicated. SP- signal peptide, MG- macroglobulin domain (MG1-MG8); LNK-link domain; ANATO- anaphylatoxin domain; α-NT- N-terminal region of the cleaved α-chain, CUB- complement C1r/C1s, Uegf, Bmp1 domain; TED- thioester-containing domain; SAR- short anchor region; C345C- carboxy-terminal domain. The cysteine residues are highlighted in yellow and the small, colored triangles of the same color represent a predicted disulphide bond in human C3.1 available from Uniprot. N-glycosylation sites (N-x-T/S, where x represents any amino acid) are highlighted in green. This is only available in digital format in Annex III because the table is very extensive.

3.8. References

Alexander, J.B., Ingram, G.A., 1992. Noncellular nonspecific defence mechanisms of fish. Annu. Rev. Fish Dis. 2, 249–279. https://doi.org/10.1016/0959-8030(92)90066-7

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anastasiou, V., Mikrou, A., Papanastasiou, A.D., Zarkadis, I.K., 2011. The molecular identification of factor H and factor I molecules in rainbow trout provides insights into complement C3 regulation. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2011.06.008

Anderson, D., & Zeeman, M. 1995., 1995. Immunotoxicology in fish.In: Gray M.(ed) Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment. 2nd ed. Taylor Fr. Washington, DC. https://doi.org/10.1177/109158189601500514

Ángeles Esteban, M., 2012. An overview of the immunological defenses in fish skin. isrn immunol. 1–29. https://doi.org/10.5402/2012/853470

Ao, J., Mu, Y., Xiang, L.-X., Fan, D., Feng, M., Zhang, S., Shi, Q., Zhu, L.-Y., Li, T., Ding, Y., Nie, L., Li, Q., Dong, W., Jiang, L., Sun, B., Zhang, X., Li, M., Zhang, H.-Q., Xie, S., Zhu, Y., Jiang, X., Wang, X., Mu, P., Chen, W., Yue, Z., Wang, Z., Wang, J., Shao, J.-Z., Chen, X., 2015. Genome sequencing of the perciform fish larimichthys crocea provides insights into molecular and genetic mechanisms of stress adaptation. PLOS Genet. 11, e1005118. https://doi.org/10.1371/journal.pgen.1005118

Aranishi, F., Nakane, M., 1997. Epidermal proteases of the Japanese eel. Fish Physiol. Biochem. 16, 471–478. https://doi.org/10.1023/A:1007736804243

Bengtén, E., Wilson, M., 2015. Antibody repertoires in fish. Springer, Cham, pp. 193–234. https://doi.org/10.1007/978-3-319-20819-0_9

Bergljot, M., 2006. Innate immunity of fish (overview). Fish Shellfish Immunol. 20, 137–51. https://doi.org/10.1016/j.fsi.2004.09.006

Bergsson, G., Agerberth, B., Jörnvall, H., Gudmundsson, G.H., 2005. Isolation and identification of antimicrobial components from the epidermal mucus of Atlantic cod (*Gadus morhua*). FEBS J. 272, 4960–4969. https://doi.org/10.1111/j.1742-4658.2005.04906.x

Boshra, H., Li, J., Sunyer, J.O., 2006. Recent advances on the complement system of teleost fish. Fish Shellfish Immunol. 20, 239–262. https://doi.org/10.1016/J.FSI.2005.04.004

Buonocore, F., Gerdol, M., 2016. Alternative adaptive immunity strategies: coelacanth, cod and shark immunity. Mol. Immunol. 69, 157–169. https://doi.org/10.1016/j.molimm.2015.09.003

Cardoso, J.C.R., Bergqvist, C.A., Larhammar, D., 2020. Corticotropin-releasing hormone (CRH) gene family duplications in lampreys correlate with two early vertebrate genome doublings. Front. Neurosci. 14, 672. https://doi.org/10.3389/fnins.2020.00672

Chaves-Pozo, E., Liarte, S., Fernández-Alacid, L., Abellán, E., Meseguer, J., Mulero, V., García-Ayala, A., 2008. Pattern of expression of immune-relevant genes in the gonad of a teleost, the gilthead seabream (*Sparus aurata* L.). Mol. Immunol. 45, 2998–3011. https://doi.org/10.1016/j.molimm.2008.01.018

Chen, Y., Xu, K., Li, J., Wang, X., Ye, Y., Qi, P., 2018. Molecular characterization of complement component 3 (C3) in Mytilus coruscus improves our understanding of bivalve complement system. Fish Shellfish Immunol. 76, 41–47. https://doi.org/10.1016/j.fsi.2018.02.044

Cole, A.M., Weis, P., Diamond, G., 1997. Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder. J. Biol. Chem. 272, 12008–13. https://doi.org/10.1074/JBC.272.18.12008

Cordero, H., Brinchmann, M.F., Cuesta, A., Meseguer, J., Esteban, M.A., 2015. Skin mucus proteome map of European sea bass (*Dicentrarchus labrax*). Proteomics 15, 4007–4020. https://doi.org/10.1002/pmic.201500120

Cordero, H., Morcillo, P., Cuesta, A., Brinchmann, M.F., Esteban, M.A., 2016. Differential proteome profile of skin mucus of gilthead seabream (*Sparus aurata*) after probiotic intake and/or overcrowding stress. J. Proteomics 132, 41–50. https://doi.org/10.1016/j.jprot.2015.11.017

Dovezenski, N., Billetta, R., Gigli, I., 1992. Expression and localization of proteins of the complement system in human skin. J. Clin. Invest. 90, 2000–12. https://doi.org/10.1172/JCI116080

Dzik, J.M., 2010. The ancestry and cumulative evolution of immune reactions. Acta Biochim. Pol. 57, 443–66.

Easy, R.H., Ross, N.W., 2009. Changes in Atlantic salmon (*Salmo salar*) epidermal mucus protein composition profiles following infection with sea lice (*Lepeophtheirus salmonis*). Comp. Biochem. Physiol. Part D Genomics Proteomics 4, 159–167. https://doi.org/10.1016/j.cbd.2009.02.001

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Ferreira, V.P., Pangburn, M.K., Cortés, C., 2010. Complement control protein factor H: The good, the bad, and the inadequate. Mol. Immunol. 47, 2187–2197. https://doi.org/10.1016/j.molimm.2010.05.007

Flajnik, M.F., Kasahara, M., 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. 11, 47–59. https://doi.org/10.1038/nrg2703

Forn-Cuní, G., Reis, E.S., Dios, S., Posada, D., Lambris, J.D., Figueras, A., Novoa, B., 2014. The evolution and appearance of c3 duplications in fish originate an exclusive teleost c3 gene form with anti-inflammatory activity. PLoS One 9, e99673. https://doi.org/10.1371/journal.pone.0099673

Gemayel, R., Vinces, M.D., Legendre, M., Verstrepen, K.J., 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. https://doi.org/10.1146/annurev-genet-072610-155046

Gomez, D., Sunyer, J.O., Salinas, I., 2013. The mucosal immune system of fish: The evolution of tolerating commensals while fighting pathogens. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2013.09.032

Gonzalez, S.F., Buchmann, K., Nielsen, M.E., 2007. Complement expression in common carp

(*Cyprinus carpio* L.) during infection with *Ichthyophthirius multifiliis*. Dev. Comp. Immunol. 31, 576–586. https://doi.org/10.1016/j.dci.2006.08.010

Groff, J.M., 2001. Cutaneous biology and diseases of fish. Vet. Clin. North Am. Exot. Anim. Pract. 4, 321–411. https://doi.org/10.1016/S1094-9194(17)30037-3

Grossman, N., Leive, L., 1984. Complement activation via the alternative pathway by purified *Salmonella* lipopolysaccharide is affected by its structure but not its O-antigen length. J. Immunol. 132, 376–85.

Guardiola, F.A., Cerezuela, R., Meseguer, J., Esteban, M.A., 2012. Modulation of the immune parameters and expression of genes of gilthead seabream (*Sparus aurata* L.) by dietary administration of oxytetracycline. Aquaculture 334–337, 51–57. https://doi.org/10.1016/j.aquaculture.2012.01.003

Harboe, M., Garred, P., Karlstrøm, E., Lindstad, J.K., Stahl, G.L., Mollnes, T.E., 2009. The down-stream effects of mannan-induced lectin complement pathway activation depend quantitatively on alternative pathway amplification. Mol. Immunol. 47, 373–380. https://doi.org/10.1016/J.MOLIMM.2009.09.005

Harboe, M., Ulvund, G., Vien, L., Fung, M.&, Mollnes, T.E., 2004. The quantitative role of alternative pathway amplification in classical pathway induced terminal complement activation. Clin. Exp. Immunol. 138, 439–446. https://doi.org/10.1111/j.1365-2249.2004.02627.x

Hawlisch, H., Köhl, J., 2006. Complement and Toll-like receptors: Key regulators of adaptive immune responses. Mol. Immunol. 43, 13–21. https://doi.org/10.1016/j.molimm.2005.06.028

Hikima, J., Hirono, I., Aoki, T., 2003. The lysozyme gene in fish, in: Aquatic Genomics. Springer Japan, Tokyo, pp. 301–309. https://doi.org/10.1007/978-4-431-65938-9_27

Hu, Y., Li, A., Xu, Y., Jiang, B., Lu, G., Luo, X., 2017. Transcriptomic variation of locally-infected skin of *Epinephelus coioides* reveals the mucosal immune mechanism against Cryptocaryon irritans. Fish Shellfish Immunol. 66, 398–410. https://doi.org/10.1016/j.fsi.2017.05.042

Hughes, A.L., 1994. Phylogeny of the C3/C4/C5 complement-component gene family indicates that C5 diverged first. Mol. Biol. Evol. 11, 417–425. https://doi.org/10.1093/oxfordjournals.molbev.a040123

Ibarz, A., Pinto, P.I.S., Power, D.M., 2013. Proteomic approach to skin regeneration in a marine teleost: modulation by oestradiol-17β. Mar. Biotechnol. 15, 629–646. https://doi.org/10.1007/s10126-013-9513-4

Ishiguro, H., Kobayashi, K., Suzuki, M., Titani, K., Tomonaga, S., Kurosawa, Y., 1992. Isolation of a hagfish gene that encodes a complement component. EMBO J. 11, 829–837. https://doi.org/10.1002/j.1460-2075.1992.tb05120.x

Janssen, B.J.C., Huizinga, E.G., Raaijmakers, H.C.A., Roos, A., Daha, M.R., Nilsson-Ekdahl, K., Nilsson, B., Gros, P., 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437, 505–511. https://doi.org/10.1038/nature04005

Kato, Y., Nakao, M., Mutsuro, J., Zarkadis, I.K., Yano, T., 2003. The complement component

C5 of the common carp (*Cyprinus carpio*): cDNA cloning of two distinct isotypes that differ in a functional site. Immunogenetics 54, 807–15. https://doi.org/10.1007/s00251-002-0528-7

Khansari, A.R., Balasch, J.C., Vallejos-Vidal, E., Parra, D., Reyes-López, F.E., Tort, L., 2018. Comparative immune- and stress-related transcript response induced by air exposure and Vibrio anguillarum bacterin in rainbow trout (*Oncorhynchus mykiss*) and gilthead seabream (*Sparus aurata*) mucosal surfaces. Front. Immunol. 9. https://doi.org/10.3389/fimmu.2018.00856

Khansari, A.R., Balasch, J.C., Vallejos-Vidal, E., Teles, M., Fierro-Castro, C., Tort, L., Reyes-López, F.E., 2019. Comparative study of stress and immune-related transcript outcomes triggered by Vibrio anguillarum bacterin and air exposure stress in liver and spleen of gilthead seabream (*Sparus aurata*), zebrafish (*Danio rerio*) and rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol. 86, 436–448. https://doi.org/10.1016/j.fsi.2018.11.063

Kuraku, S., Kuratani, S., 2006. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. Zoolog. Sci. 23, 1053–1064. https://doi.org/10.2108/zsj.23.1053

Kuroda, N., Naruse, K., Shima, A., Nonaka, M., Sasaki, M., Nonaka, M., 2000. Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish. Immunogenetics 51, 117–128. https://doi.org/10.1007/s002510050020

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Leiva-Rebollo, R., Labella, A.M., Borrego, J.J., Castro, D., 2020. Immune gene expression in gilthead seabream (*Sparus aurata*) after Lymphocystis disease virus (LCDV-Sa) challenge resulting in asymptomatic infection. J. Appl. Microbiol. 128, 41–53. https://doi.org/10.1111/jam.14454

Levine, R.P., Dodds, A.W., 1990. The Thioester Bond of C3. Springer, Berlin, Heidelberg, pp. 73–82. https://doi.org/10.1007/978-3-642-74977-3_4

Li, M., Sun, L., Li, J., 2015. Edwardsiella tarda evades serum killing by preventing complement activation via the alternative pathway. Fish Shellfish Immunol. 43, 325–329. https://doi.org/10.1016/J.FSI.2014.12.037

Louro, B., Marques, J.P., Power, D.M., Canário, A.V.M., 2016. Having a BLAST: Searchable transcriptome resources for the gilthead sea bream and the European sea bass. Mar. Genomics 30, 67–71. https://doi.org/10.1016/j.margen.2016.10.004

Malachowicz, M., Wenne, R., Burzynski, A., 2017. De novo assembly of the sea trout (*Salmo trutta m. trutta*) skin transcriptome to identify putative genes involved in the immune response and epidermal mucus secretion. PLoS One 12, e0172282. https://doi.org/10.1371/journal.pone.0172282

Marino, R., Kimura, Y., De Santis, R., Lambris, J.D., Pinto, M., 2002. Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian Ciona intestinalis. Immunogenetics 53, 1055–1064. https://doi.org/10.1007/s00251-001-0421-9

Mateus, A.P., Anjos, L., Cardoso, J.R., Power, D.M., 2017. Chronic stress impairs the local

immune response during cutaneous repair in gilthead sea bream (*Sparus aurata* , L.). Mol. Immunol. 87, 267–283. https://doi.org/10.1016/j.molimm.2017.04.008

Matsushita, M., 2018. The complement system of agnathans. Front. Immunol. 9, 1405. https://doi.org/10.3389/fimmu.2018.01405

Mauri, I., Roher, N., MacKenzie, S., Romero, A., Manchado, M., Balasch, J.C., Béjar, J., Álvarez, M.C., Tort, L., 2011a. Molecular cloning and characterization of European seabass (*Dicentrarchus labrax*) and Gilthead seabream (*Sparus aurata*) complement component C3. Fish Shellfish Immunol. 30, 1310–1322. https://doi.org/10.1016/j.fsi.2011.03.013

Mauri, I., Romero, A., Acerete, L., MacKenzie, S., Roher, N., Callol, A., Cano, I., Alvarez, M.C., Tort, L., 2011b. Changes in complement responses in Gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) under crowding stress, plus viral and bacterial challenges. Fish Shellfish Immunol. 30, 182–188. https://doi.org/10.1016/j.fsi.2010.10.006

Medzhitov, R., Janeway, C., 2000. Innate immunity. N. Engl. J. Med. 343, 338–344. https://doi.org/10.1056/NEJM200008033430506

Melillo, D., Varriale, S., Giacomelli, S., Natale, L., Bargelloni, L., Oreste, U., Pinto, M.R., Coscia, M.R., 2015. Evolution of the complement system C3 gene in Antarctic teleosts. Mol. Immunol. 66, 299–309. https://doi.org/10.1016/j.molimm.2015.03.247

Meng, F., Sun, Y., Liu, X., Wang, J., Xu, T., Wang, R., 2012. Analysis of C3 suggests three periods of positive selection events and different evolutionary patterns between fish and mammals. PLoS One 7, e37489. https://doi.org/10.1371/journal.pone.0037489

Merle, N.S., Church, S.E., Fremeaux-Bacchi, V., Roumenina, L.T., 2015. Complement system part I molecular mechanisms of activation and regulation. Front. Immunol. 6, 262. https://doi.org/10.3389/fimmu.2015.00262

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: Gateway Computing Environments Workshop, GCE 2010. https://doi.org/10.1109/GCE.2010.5676129

Nakao, M., Mutsuro, J., Obo, R., Fujiki, K., Nonaka, M., Yano, T., 2000. Molecular cloning and protein analysis of divergent forms of the complement component C3 from a bony fish, the common carp (*Cyprinus carpio*): presence of variants lacking the catalytic histidine. Eur. J. Immunol. 30, 858–866. https://doi.org/10.1002/1521-4141(200003)30:3<858::AID-IMMU858>3.0.CO;2-M

Nakao, M., Tsujikura, M., Ichiki, S., Vo, T.K., Somamoto, T., 2011. The complement system in teleost fish: Progress of post-homolog-hunting researches. Dev. Comp. Immunol. 35, 1296–1308. https://doi.org/10.1016/j.dci.2011.03.003

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. https://doi.org/10.1101/gr.6316407

Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. Fishes of the world: Fifth Edition. https://doi.org/10.1002/9781119174844

Netea, M.G., Brown, G.D., Kullberg, B.J., Gow, N.A.R., 2008. An integrated model of the

recognition of *Candida albicans* by the innate immune system. Nat. Rev. Microbiol. 6, 67–78. https://doi.org/10.1038/nrmicro1815

Nigam, A.K., Kumari, U., Mittal, S., Mittal, A.K., 2012. Comparative analysis of innate immune parameters of the skin mucous secretions from certain freshwater teleosts, inhabiting different ecological niches. Fish Physiol. Biochem. 38, 1245–1256. https://doi.org/10.1007/s10695-012-9613-5

Nonaka, M.I., Terado, T., Kimura, H., Nonaka, M., 2017. Evolutionary analysis of two complement C4 genes: ancient duplication and conservation during jawed vertebrate evolution. Dev. Comp. Immunol. 68, 1–11. https://doi.org/10.1016/j.dci.2016.11.009

Nydam, M.L., DeTomaso, A.W., 2011. Creation and maintenance of variation in allorecognition loci: Molecular analysis in various model systems. Front. Immunol. https://doi.org/10.3389/fimmu.2011.00079

Papanastasiou, A.D., Zarkadis, I.K., 2005. Gene duplication of the seventh component of complement in rainbow trout. Immunogenetics 57, 703–708. https://doi.org/10.1007/s00251-005-0028-7

Peng, M., Niu, D., Chen, Z., Lan, T., Dong, Z., Tran, T.-N., Li, J., 2017. Expression of a novel complement C3 gene in the razor clam *Sinonovacula constricta* and its role in innate immune response and hemolysis. Dev. Comp. Immunol. 73, 184–192. https://doi.org/10.1016/j.dci.2017.03.027

Petersen, S. V, Thiel, S., Jensen, L., Vorup-Jensen, T., Koch, C., Jensenius, J.C., 2000. Control of the classical and the MBL pathway of complement activation. Mol. Immunol. 37, 803–11.

Pinto, P.I.S., Guerreiro, C.C., Costa, R.A., Martinez-Blanch, J.F., Carballo, C., Codoñer, F.M., Manchado, M., Power, D.M., 2019. Understanding pseudo-albinism in sole (*Solea senegalensis*): a transcriptomics and metagenomics approach. Sci. Rep. 9, 1–16. https://doi.org/10.1038/s41598-019-49501-6

Pinto, P.I.S., Thorne, M.A.S., Power, D.M., 2017. European sea bass (*Dicentrarchus labrax*) skin and scale transcriptomes. Mar. Genomics. https://doi.org/10.1016/j.margen.2017.05.002

Provan, F., Jensen, L.B., Uleberg, K.E., Larssen, E., Rajalahti, T., Mullins, J., Obach, A., 2013. Proteomic analysis of epidermal mucus from sea lice-infected Atlantic salmon, *Salmo salar* L. J. Fish Dis. 36, 311–321. https://doi.org/10.1111/jfd.12064

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, K., et al., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071. https://doi.org/10.1038/nature06967

Qi, P., Wu, B., Guo, B., Zhang, C., Xu, K., 2018. The complement factor H (CFH) and its related protein 2 (CFHR2) mediating immune response in large yellow croaker Larimichthys crocea. Dev. Comp. Immunol. 84, 241–249. https://doi.org/10.1016/j.dci.2018.02.017

Qi, Z.H., Liu, Y.F., Wang, W.N., Wu, X., Xin, Y., Lu, Y.F., Wang, A.L., 2011. Molecular characterization and functional analysis of a complement C3 molecule in the orange-

spotted grouper (*Epinephelus coioides*). Fish Shellfish Immunol. 31, 1284–1290. https://doi.org/10.1016/j.fsi.2011.09.018

Raj, V., Fournier, G., Rakus, K., Ronsmans, M., Ouyang, P., Michel, B., Delforges, C., Costes, B., Farnir, F., Leroy, B., Wattiez, R., Melard, C., Mast, J., Lieffrig, F., Vanderplasschen, A., 2011. Skin mucus of Cyprinus carpio inhibits cyprinid herpesvirus 3 binding to epidermal cells. Vet. Res. 42, 92. https://doi.org/10.1186/1297-9716-42-92

Rebl, A., Goldammer, T., 2018. Under control: The innate immunity of fish from the inhibitors' perspective. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2018.04.016

Reyes-Becerril, M., Ascencio-Valle, F., Meseguer, J., Tapia-Paniagua, S.T., Moriñigo, M.A., Esteban, M.Á., 2012. Debaryomyces hansenii L2-enriched diet enhances the immunity status, gene expression and intestine functionality in gilthead seabream (*Sparus aurata* L.). Aquac. Res. 43, 1107–1118. https://doi.org/10.1111/j.1365-2109.2011.02914.x

Reyes-Becerril, M., Guluarte, C., Ceballos-Francisco, D., Angulo, C., Esteban, M.Á., 2017. Enhancing gilthead seabream immune status and protection against bacterial challenge by means of antigens derived from *Vibrio parahaemolyticus*. Fish Shellfish Immunol. 60, 205–218. https://doi.org/10.1016/j.fsi.2016.11.053

Reyes-Becerril, M., López-Medina, T., Ascencio-Valle, F., Esteban, M.Á., 2011. Immune response of gilthead seabream (*Sparus aurata*) following experimental infection with Aeromonas hydrophila. Fish Shellfish Immunol. 31, 564–570. https://doi.org/10.1016/j.fsi.2011.07.006

Reyes-Becerril, M., Salinas, I., Cuesta, A., Meseguer, J., Tovar-Ramirez, D., Ascencio-Valle, F., Esteban, M.Á., 2008. Oral delivery of live yeast *Debaryomyces hansenii* modulates the main innate immune parameters and the expression of immune-relevant genes in the gilthead seabream (*Sparus aurata* L.). Fish Shellfish Immunol. 25, 731–739. https://doi.org/10.1016/j.fsi.2008.02.010

Ricklin, D., Hajishengallis, G., Yang, K., Lambris, J.D., 2010. Complement: a key system for immune surveillance and homeostasis. Nat. Immunol. 11, 785–797. https://doi.org/10.1038/ni.1923

Riera Romo, M., Pérez-Martínez, D., Castillo Ferrer, C., 2016. Innate immunity in vertebrates: an overview. Immunology 148, 125–39. https://doi.org/10.1111/imm.12597

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model Choice across a Large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., Roest Crollius, H., 2018. Chromosome evolution at the origin of the ancestral vertebrate genome. Genome Biol. 19, 166. https://doi.org/10.1186/s13059-018-1559-1

Sakai, D.K., 1992. Repertoire of complement in immunological defense mechanisms of fish. Annu. Rev. Fish Dis. 2, 223–247. https://doi.org/10.1016/0959-8030(92)90065-6

Salinas, I., Zhang, Y.-A., Sunyer, J.O., 2011. Mucosal immunoglobulins and B cells of teleost fish. Dev. Comp. Immunol. 35, 1346–1365. https://doi.org/10.1016/J.DCI.2011.11.009

Salles, C.M.C., Gagliano, P., Leitão, S.A.T., Salles, J.B., Guedes, H.L.M., Cassano, V.P.F., De-

Simone, S.G., 2007. Identification and characterization of proteases from skin mucus of tambacu, a Neotropical hybrid fish. Fish Physiol. Biochem. 33, 173–179. https://doi.org/10.1007/s10695-007-9128-7

Sanahuja, I., Ibarz, A., 2015. Skin mucus proteome of gilthead sea bream: A non-invasive method to screen for welfare indicators. Fish Shellfish Immunol. 46, 426–435. https://doi.org/10.1016/j.fsi.2015.05.056

Santoro, F., Lachmann, P.J., Capron, A., Capron, M., 1979. Activation of complement by *Schistosoma mansoni* schistosomula: killing of parasites by the alternative pathway and requirement of IgG for classical pathway activation. J. Immunol. 123, 1551–7.

Sarma, J.V., Ward, P.A., 2011. The complement system. Cell Tissue Res. 343, 227–235. https://doi.org/10.1007/s00441-010-1034-0

Schreiber, R.D., Morrison, D.C., Podack, E.R., Müller-Eberhard, H.J., 1979. Bactericidal activity of the alternative complement pathway generated from 11 isolated plasma proteins. J. Exp. Med. 149, 870–82. https://doi.org/10.1084/JEM.149.4.870

Skerka, C., Chen, Q., Fremeaux-Bacchi, V., Roumenina, L.T., 2013. Complement factor H related proteins (CFHRs). Mol. Immunol. 56, 170–180. https://doi.org/10.1016/j.molimm.2013.06.001

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., Morgan, J.R., Buxbaum, J.D., Sachidanandam, R., Sims, C., Garruss, A.S., Cook, M., et al., 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat. Genet. 45, 415–421. https://doi.org/10.1038/ng.2568

Solbakken, M.H., Tørresen, O.K., Nederbragt, A.J., Seppola, M., Gregers, T.F., Jakobsen, K.S., Jentoft, S., 2016. Evolutionary redesign of the Atlantic cod (*Gadus morhua* L.) Toll-like receptor repertoire by gene losses and expansions. Sci. Rep. 6. https://doi.org/10.1038/srep25211

Solbakken, M.H., Voje, K.L., Jakobsen, K.S., Jentoft, S., 2017. Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system. Proc. R. Soc. B Biol. Sci. 284, 20162810. https://doi.org/10.1098/rspb.2016.2810

Sottrup-Jensen, L., Stepanik, T.M., Kristensen, T., Lønblad, P.B., Jones, C.M., Wierzbicki, D.M., Magnusson, S., Domdey, H., Wetsel, R.A., Lundwall, A., 1985. Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. Proc. Natl. Acad. Sci. U. S. A. 82, 9–13. https://doi.org/10.1073/PNAS.82.1.9

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Straub, P.F., Higham, M.L., Tanguy, A., Landau, B.J., Phoel, W.C., Hales, L.S., Thwing, T.K.M., 2004. Suppression subtractive hybridization cDNA libraries to identify differentially expressed genes from contrasting fish habitats. Mar. Biotechnol. https://doi.org/10.1007/s10126-004-3146-6

Sun, G., Li, H., Wang, Y., Zhang, B., Zhang, S., 2010. Zebrafish complement factor H and its related genes: identification, evolution, and expression. Funct. Integr. Genomics 10, 577–

587. https://doi.org/10.1007/s10142-010-0182-3

Sunyer, J.O., Boshra, H., Lorenzo, G., Parra, D., Freedman, B., Bosch, N., 2003. Evolution of complement as an effector system in innate and adaptive immunity. Immunol. Res. 27, 549–564. https://doi.org/10.1385/IR:27:2-3:549

Sunyer, J.O., Tort, L., 1995. Natural hemolytic and bactericidal activities of sea bream *Sparus aurata* serum are effected by the alternative complement pathway. Vet. Immunol. Immunopathol. 45, 333–45.

Sunyer, J O, Tort, L., Lambris, J.D., 1997. Structural C3 diversity in fish: characterization of five forms of C3 in the diploid fish Sparus aurata. J. Immunol. 158, 2813–21.

Sunyer, J. Oriol, Tort, L., Lambris, J.D., 1997. Diversity of the third form of complement, C3, in fish: Functional characterization of five forms of C3 in the diploid fish *Sparus aurata*. Biochem. J. 326, 877–881. https://doi.org/10.1042/bj3260877

Sunyer, J.O., Zarkadis, I.K., Lambris, J.D., 1998. Complement diversity: a mechanism for generating immune diversity? Immunol. Today 19, 519–23.

Sunyer, J.O., Zarkadis, I.K., Sahu, A., Lambris, J.D., 1996. Multiple forms of complement C3 in trout that differ in binding to complement activators. Proc. Natl. Acad. Sci. U. S. A. 93, 8546–51. https://doi.org/10.1073/PNAS.93.16.8546

Tsutsui, S., Komatsu, Y., Sugiura, T., Araki, K., Nakamura, O., 2011. A unique epidermal mucus lectin identified from catfish (*Silurus asotus*): first evidence of intelectin in fish skin slime. J. Biochem. 150, 501–514. https://doi.org/10.1093/jb/mvr085

Valdenegro-Vega, V.A., Crosbie, P., Bridle, A., Leef, M., Wilson, R., Nowak, B.F., 2014. Differentially expressed proteins in gill and skin mucus of Atlantic salmon (*Salmo salar*) affected by amoebic gill disease. Fish Shellfish Immunol. 40, 69–77. https://doi.org/10.1016/j.fsi.2014.06.025

Valero, Y., García-Alcázar, A., Esteban, M.Á., Cuesta, A., Chaves-Pozo, E., 2015. Antimicrobial response is increased in the testis of European sea bass, but not in gilthead seabream, upon nodavirus infection. Fish Shellfish Immunol. 44, 203–213. https://doi.org/10.1016/j.fsi.2015.02.015

Voss, E.W., Sigel, M.M., 1972. Valence and temporal change in affinity of purified 7S and 18S nurse shark anti-2,4 dinitrophenyl antibodies. J. Immunol. 109, 665–73.

Wang, L., Shao, C., Xu, W., Zhou, Q., Wang, N., Chen, S., 2017. Proteome profiling reveals immune responses in Japanese flounder (*Paralichthys olivaceus*) infected with Edwardsiella tarda by iTRAQ analysis. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2017.05.022

Warren, I.A., Naville, M., Chalopin, D., Levin, P., Berger, C.S., Galiana, D., Volff, J.-N., 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome Res. 23, 505–31. https://doi.org/10.1007/s10577-015-9493-5

Watts, M., Munday, B.L., Burke, C.M., 2001. Immune responses of teleost fish. Aust. Vet. J. 79, 570–4.

Whaley, K., Loos, M., Weiler, J.M., 1993. Complement in Health and Disease. Springer Netherlands.

Yano, T., Hatayama, Y., Matsuyama, H., Nakao, M., 1988. Titration of the alternative complement pathway activity of representative cultured fishes. Nippon SUISAN GAKKAISHI 54, 1049–1054. https://doi.org/10.2331/suisan.54.1049

Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., Zhang, G., 2015. Massive expansion and functional divergence of innate immune genes in a protostome. Sci. Rep. 5, 8693. https://doi.org/10.1038/srep08693

Zhang, S., Cui, P., 2014. Complement system in zebrafish. Dev. Comp. Immunol. 46, 3–10. https://doi.org/10.1016/j.dci.2014.01.010

Zhu, L., Nie, L., Zhu, G., Xiang, L., Shao, J., 2013. Advances in research of fish immune-relevant genes: A comparative overview of innate and adaptive immunity in teleosts. Dev. Comp. Immunol. 39, 39–62. https://doi.org/10.1016/J.DCI.2012.04.001.

CHAPTER 4

Microbiome establishment in early developmental stages (egg and larvae)

CHAPTER 4.1

Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream

Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream

Manuscript published in Microorganisms 9 (2021), 2275.
https://doi.org/10.3390/microorganisms9112275

Acknowledgments

The authors acknowledge all the hatcheries for the helpful and productive collaboration.

Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream

Babak Najafpour [1], Patricia I. S. Pinto [1], Katerina A. Moutou [2], Adelino V. M. Canario [1] and Deborah M. Power [1,*]

[1] Centro de Ciências do Mar (CCMAR/CIIMAR), Universidade do Algarve, 8005-139 Faro, Portugal

[2] Department of Biochemistry and Biotechnology, University of Thessaly, Biopolis, 41221 Larissa, Greece

CReDit statement:

Babak Najafpour: Methodology Optimization - Lead, Bioinformatics - Lead, Formal analysis, Validation, Investigation, Visualization, Writing – Original Draft, Review & Editing, Patricia I. S. Pinto: Review & Editing, Data Organization; Katerina A. Moutou: Review & Editing; Adelino V. M. Canario Review & Editing; Deborah M Power: Writing - Review & Editing, Investigation, Supervision, Project administration, Funding acquisition.

4.1.1. Abstract

A comprehensive understanding of how bacterial community abundance changes in fishes during their lifecycle and the role of the microbiota on health and production is still lacking. From this perspective, the egg bacterial communities of two commercially farmed species, the European seabass (*Dicentrarchus labrax*) and the gilthead seabream (*Sparus aurata*), from different aquaculture sites were compared, and the potential effect of broodstock water microbiota and disinfectants on the egg microbiota was evaluated. Moreover, 16S ribosomal RNA gene sequencing was used to profile the bacterial communities of the eggs and broodstock water from three commercial hatcheries. Proteobacteria were the most common and dominant phyla across the samples (49.7% on average). *Vibrio* sp. was the most highly represented genus (7.1%), followed by *Glaciecola* (4.8%), *Pseudoalteromonas* (4.4%), and *Colwellia* (4.2%), in eggs and water across the sites. Routinely used iodine-based disinfectants slightly reduced the eggs' bacterial load but did not significantly change their composition. Site, species, and type of sample (eggs or water) drove the microbial community structure and influenced microbiome functional profiles. The egg and seawater microbiome composition differed in abundance but shared similar functional profiles. The strong impact of site and species on egg bacterial communities indicates that disease management needs to be site-specific and highlights the need for species- and site-specific optimization of disinfection protocols.

Keywords: broodstock water, disinfection, hatchery, microbiome, teleost eggs

4.1.2. Introduction

Knowledge about the symbiotic interdependency between complex multicellular eukaryotes and their microbiota is changing our understanding of animal biology (Gilbert et al., 2012; McFall-Ngai et al., 2013). Studies of the microbiota in humans have highlighted the complexity and unexpected role that the microbiome plays in development and physiology (Pasolli et al., 2019; Proctor et al., 2019). This has transformed our understanding of the importance of microbiota for health and disease and opened up a new research frontier (Hadrich, 2018). In terrestrial animal production systems, the role of the microbiota in traits of interest and as a means to control pathogens and deliver alternatives to conventional pharmaceuticals has stimulated high interest (Brugman et al., 2018). Benefits already accrued are linked to host nutrition, the promotion of epithelial barrier function, stimulation of the immune system, and protection against colonization by pathogens (Pickard et al., 2017). The recent massive increase in knowledge about the composition and function of microbes, including non-culturable bacteria in a wide range of ecological niches, has been made possible by 16S ribosomal RNA (rRNA) gene amplicon sequencing and meta-transcriptomics (Aßhauer et al., 2015; Simon and Daniel, 2011).

The high number and diversity of bacteria and viruses in aquatic systems mean that aquatic organisms such as fish are exposed to a larger number and diversity of microorganisms than terrestrial vertebrates, and this is proposed to have influenced their physiology (Groff, 2001). This suggests that improved knowledge about fish microbiota may create new opportunities for the management of aquaculture production. This notion has accelerated microbiome research, particularly in species of commercial interest for aquaculture (Llewellyn et al., 2014). Most studies so far have focused on the role of the microbiota in specific tissues, mostly the gut of adult fish, due to the recognized importance of this tissue in digestion and the assimilation of essential nutrients and in contributing to good growth and health (Egerton et al., 2018; Legrand et al., 2018, 2020). Furthermore, considerable interest exists in the potential benefits that may be obtained by manipulating the microbiota of the fish gut using probiotics and prebiotics in the diet (Newaj-Fyzul et al., 2014; Ringø et al., 2014; Tellez et al., 2006). Such studies have revealed the potential role of the gut microbiota in resistance to pathogens, improved growth, lipid metabolism, and the immune response (Wang et al., 2018).

Studies investigating fish egg- and embryo-associated microbiota have so far mainly targeted potential pathogens such as *Leucothrix mucor*, *Flexibacter ovolyticus*,

Flavobacterium columnare, and *Pseudoalteromonas piscicida* (Hansen and Olafsen, 1999; Merrifield and Rodiles, 2015; Olafsen, 2001). Although the use of metagenomics to determine the global community of microbes, or microbiota, of fish eggs and embryos is still at an early stage, a positive impact of the intrinsic microbiota on fish egg quality has been recognized (Borges et al., 2021). Dysbiosis has been linked to the colonization of eggs by pathogenic bacteria, leading to high mortality during incubation, and it is now clear that the fish egg microbiota shapes the larval microbiota (Llewellyn et al., 2014; Olafsen, 2001) and may influence subsequent larval performance (Vadstein et al., 2013). Recent studies of eggs from channel catfish (*Ictalurus punctatus*, Abdul Razak et al., 2019) and ballan wrasse (*Labrus bergylta*, Bone et al., 2020) suggest that their microbiota may be influenced by the holding tank, genetics (family effects), or egg disinfection procedures. Clearly, a better understanding of the factors influencing the egg microbiota of aquaculture species can contribute information relevant for improving management and may help to identify whether the microbiota has a role in determining egg quality.

The European seabass (*Dicentrarchus labrax*) and the gilthead seabream (*Sparus aurata*) are the most important farmed fish species in the Mediterranean region, with a combined production of 407,673 tonnes in 2019, with the highest contribution from Turkey (FEAP Annual Report, 2020). The hatchery stage of production is crucial for industry sustainability. Identification of the factors that lead to unpredictable production quality due to high mortality rates or poor-quality eggs and subsequent developmental stages is a priority for gilthead seabream and European seabass hatcheries (Llewellyn et al., 2014; Muniesa et al., 2020).

The main objectives of the present metagenomic study, in the context of the hatchery production of gilthead seabream and European seabass, were therefore to (i) establish the composition and diversity of the bacterial community in eggs collected from three different industrial sites, (ii) assess the impact of disinfection protocols on the egg microbiome, and (iii) evaluate the contribution of the microbiota in the broodstock tank water to the egg bacterial community. This is, to our knowledge, the first metagenomic study of the fish egg microbiome and its modulation by environmental variables in different sites across the Mediterranean coastal waters of Greece.

4.1.3. Materials and Methods

4.1.3.1. Broodstock Culture Conditions and Disinfection Protocol

Eggs from 11 different broodstock (BS) tanks from European seabass and gilthead seabream were obtained from three hatcheries located at different sites in Greek coastal waters (**Figure 4.1.1** and **Table 4.1.1**) in January 2020. The sites were designated 1 to 3 and samples designated according to their site origin, but due to the potential commercial sensitivity of the information, samples and site origin are not linked. The different gilthead seabream and European seabass BS from the same sites were maintained in separate sea water flow through tanks with independent water supplies. The experimental design, experimental tank temperatures and volumes, and sample references are provided in **Figure 4.1.1**.

	Sp.	Broodstock Ta.	Tm.ºC	TaV.m³	Collected sample		Sample code
Site 1	SA	Tank 1	19	13	Egg	B.Dis.	S1.SA.E.BD1
						A.Dis.	S1.SA.E. AD1
					Water	BW1	S1.SA.BW1
	SA	Tank 2	18.7	13	Egg	A.Dis	S1.SA.E.AD2
					Water	BW2	S1.SA.BW2
	SA	Tank 3	18.9	18	Egg	B.Dis.	S1.SA.E.BD3
						A.Dis.	S1.SA.E.AD3
					Water	BW3	S1.SA.BW3
	SA	Tank 4	18.9	18	Egg	B.Dis.	S1.SA.E.BD4
						A.Dis.	S1.SA.E.AD4
					Water	BW4	S1.SA.BW4
Site 2	SA	Tank 1	17.5	50	Egg	B.Dis.	S2.SA.E.BD1
						A.Dis.	S2.SA.E.AD1
	DL	Tank 1	14	50	Egg*	B.Dis.	S2.DL.E.AD1
						A.Dis.	S2.DL.E.AD1
					Water	BW1	S2.DL.BW1
	DL	Tank 2				BW2	S2.DL.BW2
Site 3	SA	Tank 1	19	25	Egg	B.Dis.	S3.SA.E.BD1
						A.Dis.	S3.SA.E.AD1
					Water	BW1	S3.SA.BW1
	SA	Tank 2	17	25	Egg	B.Dis.	S3.SA.E.BD2
						A.Dis.	S3.SA.E.AD2
					Water	BW2	S3.SA.BW2
	DL	Tank 1	14	25	Egg	B.Dis.	S3.DL.E.BD1
						A.Dis.	S3.DL.E.AD1
					Water	BW1	S3.DL.BW1
	DL	Tank 2	14	25	Egg	B.Dis.	S3.DL.E.BD2
						A.Dis.	S3.DL.E.AD2
					Water	BW2Θ	

Figure 4.1.1. A schematic representation of the approximate location of the hatcheries and the experimental design. Egg and water samples were collected from the three different sites/hatcheries (red circles on the map) and different broodstock in January 2020. Egg samples of the gilthead seabream (SA) and the European seabass (DL) were taken before (B.Dis.) and

after (A.Dis.) standard egg disinfection procedures at each site. For easy tracking of samples and their results, sample codes were assigned to each of the collected samples (see last column). Sp. (species); BW (broodstock water); Ta. (tank); Tm. (temperature in °C); TaV. (tank volume in m^3). Note: * the eggs from two broodstock were mixed in this case. Θ—sequencing of this specific water sample did not proceed because insufficient DNA was obtained during extraction.

Table 4.1.1. Broodstock characteristics and conditions (linked to Figure 4.1.1).

Site	BS	Species	Tank	Male (*n*)	Female (*n*)	Density (kg/m^3)	Weight ❁ (kg)
1	BS 1	SA	1	10	31	8.36	2.02
1	BS 2	SA	2	11	20	5.28	2.21
1	BS 3	SA	3	16	30	5.82	2.23
1	BS 4	SA	4	10	27	6.78	3.01
2	BS 1	SA	1	38	105	5.9	1.90
2	BS 1	DL	1	67	68	14.9	5.00
2	BS 2	DL	2	50	75	7.7	2.80
3	BS 1	SA	1	63	20	5.29	1.59
3	BS 2	SA	2	23	25	3.45	1.79
3	BS 1	DL	1	23	28	7.11	3.48
3	BS 2	DL	2	25	31	8.61	3.81

BS = Broodstock; *n* = number of fish; ❁ = weight expressed as the mean/tank.

Different protocols were used for egg disinfection at the different sites; these were the standard operating protocols established for routine use in each of the commercial hatcheries that collaborated in the study. In site 1, gilthead seabream eggs (≅1 kg) were immersed in 10 L of seawater containing 0.25 mL GERM-iod (18 mg iodine/mL, final concentration 0.45 mg iodine/L) for 3 min. Eggs were rinsed after treatment with clean seawater and stocked directly into larval tanks. In site 2, gilthead seabream eggs (≅1 kg) were immersed in 10 L of seawater containing 50 mL of Ovadine (Syndel, 10% polyvinyl pyrrolidine iodine with 1% available iodine) for 7 min. The eggs were not rinsed and were directly stocked into the incubator tanks (500 L). European seabass eggs (≅1 kg) followed a similar protocol to gilthead seabream egg treatments, with the exception that 70 mL of Ovadine was used and the egg incubation tanks were of 200 L. In site 3, gilthead seabream and European seabass eggs were immersed in 10 L of seawater containing 35 mL of Buffodine (Evans Vanodine International plc; iodine-based

and neutral pH disinfectant) for 5 min. After treatment, the eggs were washed in clean seawater and stocked into larval tanks (1000 L).

4.1.3.2. Sample Collection

European seabass and gilthead seabream eggs and broodstock water samples (BW) were collected from each of the three hatchery sites. Floating eggs were collected into a sterile beaker by scooping them from the surface of the water into an egg collector. The eggs were recovered from the seawater by gently straining through a tea strainer, and rinsed in sterile seawater before transfer to sterile 50 mL tubes containing RNA later (Sigma-Aldrich, Madrid, Spain; eggs:RNA later *v*/*v* 1:10). Nineteen egg samples were collected, 9 before and 10 after disinfection. A total of 10 BW samples (400 mL each) were collected by scooping seawater directly from the broodstock tank using a sterile beaker and transferred into 1 L bottles containing 50 mL DESS solution (0.25 M disodium EDTA, pH 8.0, 20% dimethyl sulfoxide, saturated with sodium chloride (Yoder et al., 2006)), mixed well and stored at 4 °C.

4.1.3.3. DNA Extraction

Total DNA from the 29 samples (19 egg samples and 10 water samples) was extracted using a DNeasy Blood & Tissue Kit (Qiagen, Hamburg, Germany) following the manufacturer's instructions, modified to include pre-digestion with lysozyme and RNAse treatment (Pinto et al., 2019). The optimization of the initial mechanical disruption step was established for each sample type, as was the weight of eggs, water volume, lysozyme concentration, and buffers used in the present study.

For most egg samples, 30 mgs of eggs suspended in RNA later was used for disruption and yielded sufficient DNA for 16S rRNA library construction. For some samples (S1.SA.E.AD3, S1.SA.E.AD4, S2.SA.E.BD1, S2.SA.E.AD1, S3.DL.E.AD2, S3.DL.E.BD1, and S3.DL.E.AD1), 90 mg of eggs was extracted to ensure sufficient DNA yield for subsequent analysis. Lysis mix (200 µL of lysis buffer 20 Mm Tris-HCl, pH 8; 2 mM sodium EDTA; 1.2% Triton X-100; 40 mg/mL lysozyme mixed with 200 µL of AL buffer from the Qiagen kit) was added to each egg sample with two iron beads (Qiagen stainless steel beads of 5 mm) per sample. Initial mechanical disruption of the eggs was carried out using 3 cycles of 30 s at 30 Hz in a Tissue Lyser (Qiagen). The iron beads were then removed and 400 mg of 0.1 mm zirconia/silica beads per tube was added and a second step of mechanical disruption (3 cycles of 5 min at 25 Hz) targeting the bacterial cells was performed.

For BW (400 mL), the particulate matter containing the microorganisms was concentrated into a single 50 mL sterile tube by centrifugation of the water at 16,100 *g* for 20 min at 4 °C. The resulting pellet was suspended at room temperature in 2 mL sterile extraction tubes (Sarstedt, Nümbrecht, Germany) by adding 400 µL of the lysis mix and approximately 400 mg of 0.1 mm zirconia/silica beads (Biospec). Mechanical disruption of the pellet was carried out at room temperature in the Tissue Lyser using 3 cycles of 5 min at 25 Hz.

4.1.3.4. 16 S rRNA Library Construction and Sequencing

The extracted DNA was shipped on ice to Stab Vida, Lda (Lisbon, Portugal), where the integrity and quantity of the DNA was confirmed using 1.5% agarose gel electrophoresis and a Qubit 2 fluorometer (ThermoFisher Scientific, Lisbon, Portugal). Metagenomic amplicon library construction was carried out using an Illumina 16S Metagenomic Sequencing Library preparation protocol, with 12.5 ng DNA per sample, and primers targeting the V3 and V4 hypervariable regions of the 16S rRNA gene for amplification (Klindworth et al., 2013). From the 29 samples, 28 libraries were generated and successfully sequenced by Stab Vida, Lda using a MiSeq Reagent Kit v3 and generated 300 bp paired-end sequencing reads in an Illumina MiSeq instrument.

4.1.3.5. Sequence Processing and Bioinformatics

The quality of the raw sequencing reads was evaluated using FastQC (Andrews, 2010), and the reads were denoised using the DADA2 plugin of QIIME 2 v2020.2 (Callahan et al., 2016; Caporaso et al., 2010) and included read filtering, dereplication, and chimera filtering. After preliminary analysis to evaluate the presence of host DNA in the data generated, an additional filtering step was introduced. Specifically, the operational taxonomic units (OTUs) generated were queried using BLAST against an in-house sequence database, created using the two host DNA genomes, gilthead seabream *Sparus aurata* (Genbank assembly accession GCA_900880675) and European seabass *Dicentrarchus labrax* (GCA_000689215). The parameter settings for host-specific DNA filtering were: word size 11, match point 2, mismatch point-3, gap existence-5/gap extension-2. OTUs corresponding to contaminating host DNA were removed using the QIIME 2 filtering options, and the remaining OTUs were analyzed.

The rarefaction curves for the samples of each site were plotted using the *rarecurve* function in the R package vegan v 2.5-6. QIIME 2 v2020.2 was used for the identification and classification of OTUs using the scikit-learn classifier against the SILVA

(release 132 QIIME) database (Quast et al., 2013), with a cut-off threshold set at 97% similarity. For classification purposes, only OTUs in the dataset containing at least 10 sequence reads were considered. QIIME 2 was also used to calculate commonly used alpha- and beta-diversity metrics (Shannon's diversity index and Bray–Curtis distance) and the output was imported into R to produce principal coordinate analysis (PCoA) and for visualization of the data using the packages qiime2R v 0.99.6 and ggplot2 v 3.3.5.

4.1.3 6 Functional Analysis

Functional predictions based on the 16S rRNA metagenomics profiles were run on the web-based platform Microbiome Analyst (Chong et al., 2020) using the Tax4Fun method (Aßhauer et al., 2015). The associations between functional categories (Kyoto Encyclopedia of Genes and Genomes, KEGG pathways) and the experimental factors (site, disinfection, species, and sample type) were tested using the global test algorithm (Goeman et al., 2004).

4.1 3 7. Quantitative Analysis of 16S rRNA Gene

The 16S rRNA gene was quantified in genomic extracts, in duplicate reactions, by quantitative polymerase chain reaction (qPCR), run in a Bio-Rad CFX96 qPCR Instrument (Bio-Rad Laboratories, Hercules, CA, USA). The effects of aquaculture site, species, and disinfection on the bacterial load of the eggs were evaluated using the quantified 16S rRNA gene across samples. The primers used for estimation of bacterial loads were those recommended by the Earth Microbiome Project (http://www.earthmicrobiome.org/protocols-and-standards/16s/ accessed on 15th June 2020) and targeted a fragment of approx. 300 bp between positions 515 and 806 of the 16S rRNA gene. The sense primer 16S-515fbY or 515F(Parada) sequence is 5'-GTGYCAGCMGCCGCGGTAA-3′ and the antisense primer 16S-806rbN or 806R(Apprill) is 5′-GGACTACNVGGGTWTCTAAT-3′ (Apprill et al., 2015; Caporaso et al., 2012; Parada et al., 2016), using standard codes for degenerate bases (Y = C or T; N = A, C, T, or G; M = A or C; W = A or T; V = A, C, or G).

The final qPCR reaction volume was 10 µL and contained 200 nM of each primer, 2 µL of the template cDNA (10 ng), and 5 µL of 2X Forget-Me-Not™ EvaGreen® qPCR Master Mix (Biotium). Thermocycling conditions were 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s, 50 °C for 10 s, and 72 °C for 10 s, with a final melting curve generated by increasing the temperature from 60 °C to 95 °C, with increments of 0.5 °C each 10 s. The absence of non-specific amplification and primer dimers was verified by analysis of melting curves and running

representative amplification products on 2% agarose gels, which confirmed single peaks and products of the expected size. Standard curves were included in all qPCR plates and were prepared from serial dilutions of a plasmid containing a 1013 bp fragment of the 16S rRNA gene from *Mycoplasma* cloned and sequenced from a *Solea senegalensis* gut sample (Pinto et al., 2019). Control reactions were added to all qPCR plates, including a no template control to confirm the absence of reagent contamination. The reaction efficiency of the qPCR and coefficient of determination (r^2) were 95.6% and 0.986, respectively.

4.1.3.8. Statistics

Overall, four variables were considered in the statistical tests to analyze the bacterial communities: (i) disinfection (eggs before versus eggs after disinfection), (ii) sample type (broodstock water vs. eggs), (iii) species (gilthead seabream vs. European seabass), and (iv) geographical location of the aquaculture site (site 1 vs. site 2 vs. site 3). Statistical significance in all tests was set at $p < 0.05$. Data normality and homogeneity were tested using the Shapiro–Wilk normality test. Statistical analyses were performed in the R environment.To compare alpha-diversity by Shannon indexes for each of the 4 variables, one-way ANOVA (analysis of variance) was applied. To evaluate beta-diversity, data homogeneity was controlled using the *betadisper* function (evaluating beta-dispersion) and the *permutest* function. Based on the principal coordinate analysis (PcoA) of beta-diversity data, permutational analysis of variance (PERMANOVA) using the *adonis* function and Bray–Curtis distances were applied to test whether the overall microbial community differed with each variable under analysis. Differential abundance analyses were run using the R package ALDEx2 (1.20.0) to find features and pathways that had different abundance across variables. The centered log-ratio-transformed values were analyzed using a general linear model (glm) and Kruskal–Wallis tests. To specify features/OTUs with a significantly different relative abundance for the defined variables, the output was filtered using glm.ep based on the confidence interval of 95% ($p < 0.05$). The paired samples Wilcoxon test was used to compare qPCR 16S rRNA abundance measurements before and after disinfection and the Kruskal–Wallis test to compare abundance between species and site.

4.1.4. Results

4.1.4.1. Sequencing and rarefaction outcome

A total of 11.7 million paired-end reads (with an average read number of 417,809 ± SD of 158,260) were produced from the 28 metagenomic 16S libraries. Sequence assembly yielded 3 million paired-reads, with a mean of 106,893 reads per library, which, after quality control and trimming of low-quality sequences, resulted in 2 million reads, with a mean of 72,072 sequences per library (**Supplementary table 4.1.1**). Taxonomic classification of these reads identified a total of 1,819 unique features (OTUs).

The alpha rarefaction curves confirmed that the sequencing depth was sufficient to cover the microbial community diversity across samples, as they reached a plateau in all 28 libraries (**Figure 4.1.2**). Overall, less variation in OTUs between libraries was observed for site 3 compared to the other sites, and no obvious pattern was observed in the total number of OTUs in eggs before and after disinfection in any of the sites. However, in sites 1 and 3, the highest numbers of OTUs were in two specific egg samples after disinfection. In site 2, the highest number of OTUs was detected in one egg sample before disinfection.

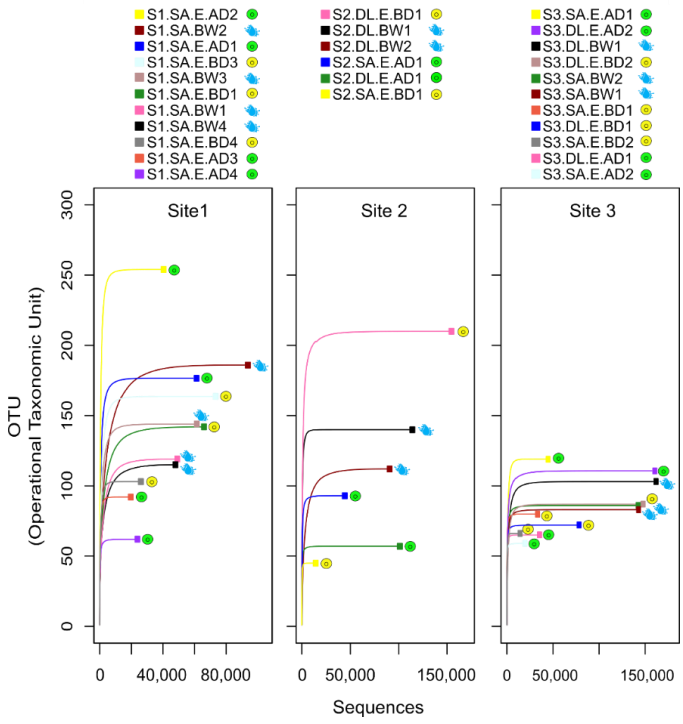


Figure 4.1.2. Rarefaction plots of the sequencing data from the 16S rRNA microbiome libraries. The rarefaction curves related to the eggs and water samples of each site (Site 1, Site

2, Site 3) are plotted separately. Symbols show different types of samples: eggs before disinfection = yellow circle; eggs after disinfection = green circle; BS water = blue drop. The sequences generated from each of the 28 libraries reached a plateau for all the samples, revealing that all bacterial diversity was covered. The rarefaction curves were plotted using the *rarecurve* function in the R package vegan (v 2.5-6). Information about the sample labels is presented in the legend of **Figure 4.1**.

4.1.4.2. Bacterial Community Taxonomic Composition

The bacterial composition and their relative abundance were determined in each sample at different taxonomic levels and the data are presented for all taxa detected at > 1% in abundance across all samples (**Supplementary table 4.1.2**). Proteobacteria were the dominant phylum and were highly represented in all the samples (mean = 49.7%, range = 10.7–70.7%) (Figure 3a). Bacteroidetes was the second most dominant phylum across samples (mean = 15.9, range = 0.11–45.2%, **Figure 4.1.3a**). The highest relative abundance and representation of Cyanobacteria was observed in site 3 (mean = 28.9%, range = 9.3–42.0%), followed by site 2 (mean = 12, range = 0.01–39.4%) and site 1 (mean = 2.4, range = 0–8.2%, **Figure 4.1.3a**).

The two most abundant families that were represented in almost all samples irrespective of site or fish species were *Flavobacteriaceae* (mean = 10.3%, range = 0–38.7%) and Vibrionaceae (mean = 8.9%, range = 0.12–40.3%). *Colwelliaceae*, *Rhodobacteraceae*, *Pseudoalteromonadaceae*, and *Alteromonadaceae* (mean ≅ 5%) were the next most abundant families across all egg and water samples (**Figure 4.1.3b**). All families, except for *Pseudoalteromonadaceae*, were more abundant in water than in eggs (**Figure 4.1.3b**). *Shewanellaceae* (mean = 3.3%, range = 0–31.1%) showed high relative abundance in gilthead seabream egg samples from site 1 (**Figure 4.1.3b**).

The relative abundance (%) of microbial genera for each sample is described in **Supplementary table 4.1.3**. The 15 most abundant genera per aquaculture site are shown in **Figure 4.1.4a** and per sample type (eggs before and after disinfection, and broodstock water) in **Figure 4.1.4b**. Some bacterial genera were not specified (mean = 22.1%, range = 1.1–83.6) because they were not represented in the SILVA database (NA in **Supplementary table 4.1.2**). Cyanobacteria (mean = 13.2%, range = 0.02–82.6%) were among the highest proportion of the 22.1% unidentified bacterial genera (**Figure 4.1.4**). *Vibrio* was the most represented genus in eggs and broodstock water at the three sites (mean = 7.1%, range = 0–35.6%). *Glaciecola* (mean = 4.8%, range = 0–31.0%), *Pseudoalteromonas* (mean = 4.4%, range = 0–16.8%), and

Colwellia (mean = 4.2%, 0.07–15.8%) were the genera with the next highest relative abundance across the samples (**Figure 4.1.4**).

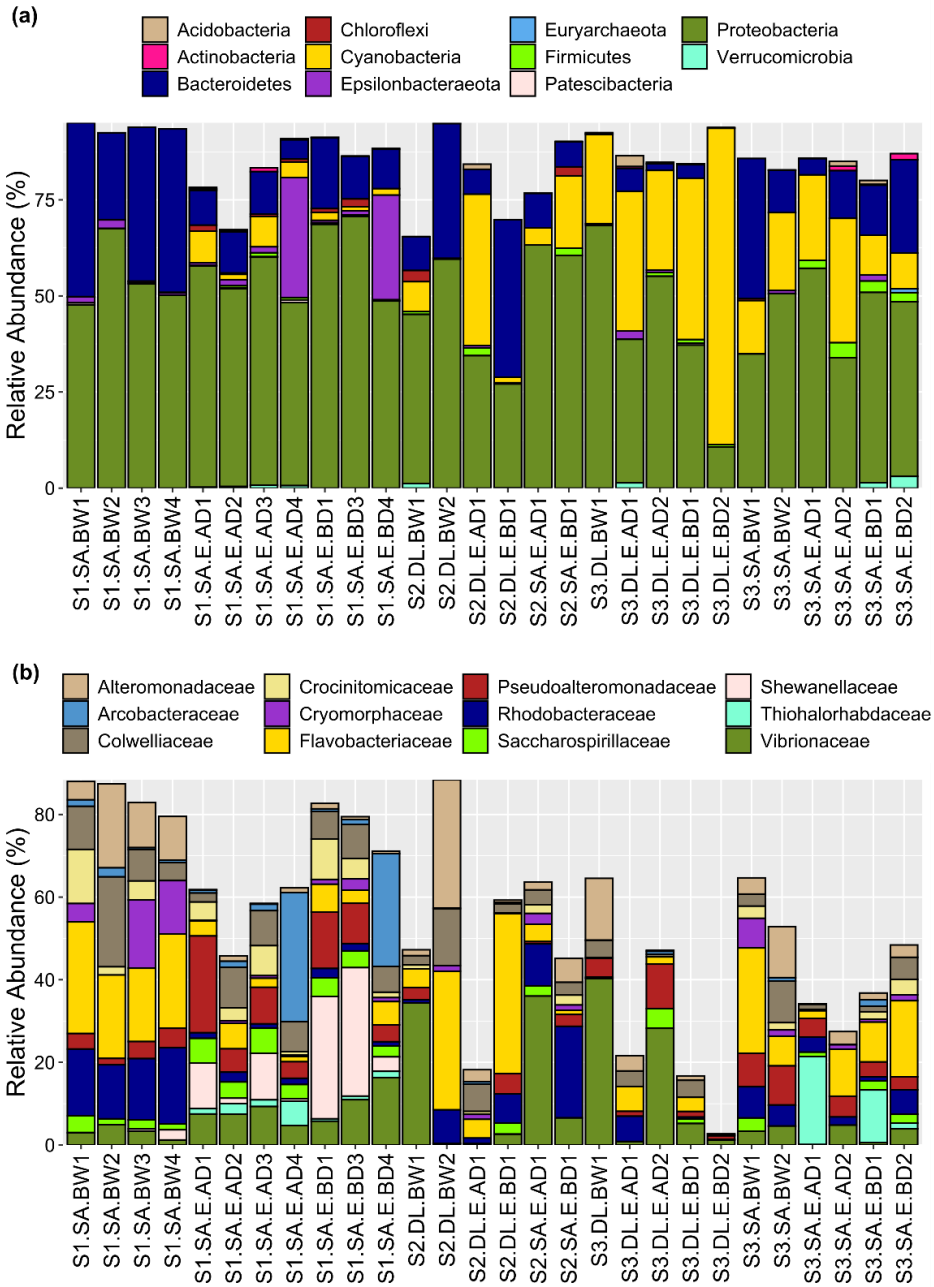


Figure 4.1.3. The relative abundance of the 11 phyla (a) and 12 families (b) across egg and water samples. The relative proportions of phyla and families (in % in relation to the total profiled microbiome) were plotted using the ggplot2 R package. Information about the sample labels is presented in the legend of **Figure 4.1.1**.

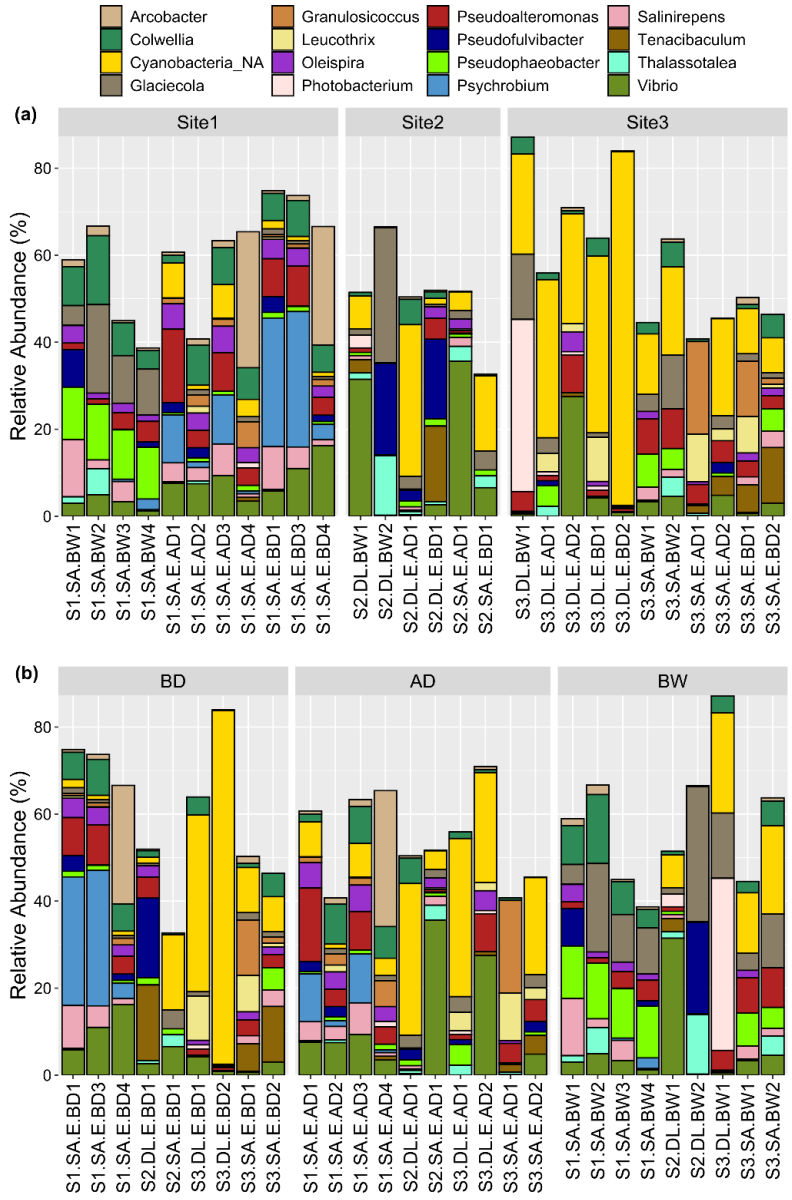


Figure 4.1.4. The relative percentage of the most abundant bacterial genera (top 15) across all egg and water samples. The bacterial genera present at the highest relative percentage, based on the sum of the percentage detected in all samples, were selected for presentation. The same data are rearranged and presented (using the ggplot2 R package) in relation to (**a**) the aquaculture site or (**b**) the type of sample including eggs (before disinfection, BD, and after disinfection, AD) and water samples (broodstock water, BW).

Consideration of site-specific relative abundance of bacterial genus revealed that *Psychrobium* was very abundant in some egg samples from site 1 (mean = 3.32%, range = 0–31.1% across all samples). Cyanobacteria were abundant in most of the egg and in all water

samples from site 3 and in some egg samples from site 2 (**Figure 4.1.4a**). *Glaciecola* and *Pseudophaeobacter* (mean = 3%, range = 0–12.8% across all samples) contributed at a higher relative proportion to the microbiota in water samples compared to egg samples (**Figure 4.1.4b**). Comparison of the bacterial community profiles before disinfection (BD) and after disinfection (AD) suggested that disinfection had a relatively mild effect on the microbiota overall, although some genera appeared to be more affected by disinfection. For example, *Psychrobium* abundance was substantially reduced in two gilthead seabream egg samples (S1.SA.E.AD1 and S1.SA.E.AD3) and *Pseudofulvibacter* was reduced in one European seabass egg sample (S2.DL.E.AD1, **Figure 4.1.4b**).

4.1.4.3. Bacterial Community Diversity Alpha-Diversity

Shannon's diversity indexes (reflecting the microbial community alpha-diversity) were not significantly modified by disinfection (egg before vs. egg after disinfectant usage), sample type (broodstock water vs. egg), species (gilthead seabream eggs vs. European seabass eggs), or site (site 1, site 2, or site 3, **Figure 4.1.5**).

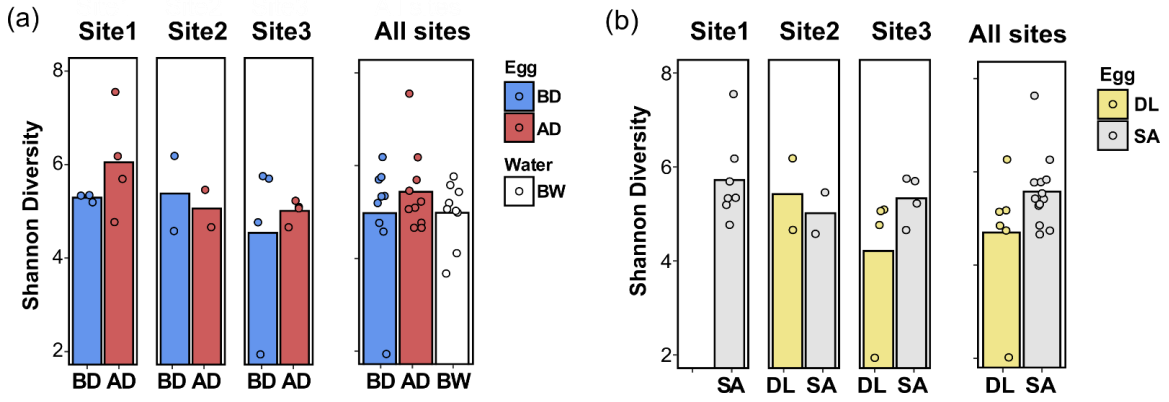


Figure 4.1.5. The Shannon index of alpha diversity. The Shannon index is presented according to (**a**) the sample type, including eggs (before disinfection, BD, and after disinfection, AD) and water samples (broodstock water, BW) across the three aquaculture sites (Site 1, Site 2, and Site 3), and (**b**) the species (gilthead seabream, SA, and European seabass, DL) across the three aquaculture sites (Site 1, Site 2, and Site 3), or in the representative plot on the right of each panel, compiling the average Shannon index for all three sites. No significant difference was observed in the Shannon index, by sample type, species, or site.

4.1.4.4. Beta-Diversity

To visualize the differences between bacterial community composition and distance across all collected samples (eggs and water), PcoA analysis was performed and the plot displayed in a two-dimensional space (**Figure 4.1.6**). The distinction in the microbial communities between sample type (eggs vs. water) was most obvious in sites 1 and 3 compared to site 2 (Figure 6). No clear separation was observed in the bacterial composition of the eggs before and after disinfection (p = 0.96, **Table 4.1.3**). Site and species had a significant effect on the egg bacterial community, and the bacterial composition of the eggs and broodstock water was significantly different (Bray–Curtis distance analysis in PERMANOVA, $p < 0.001$, **Tables 4.1.2 and 4.1.3**).

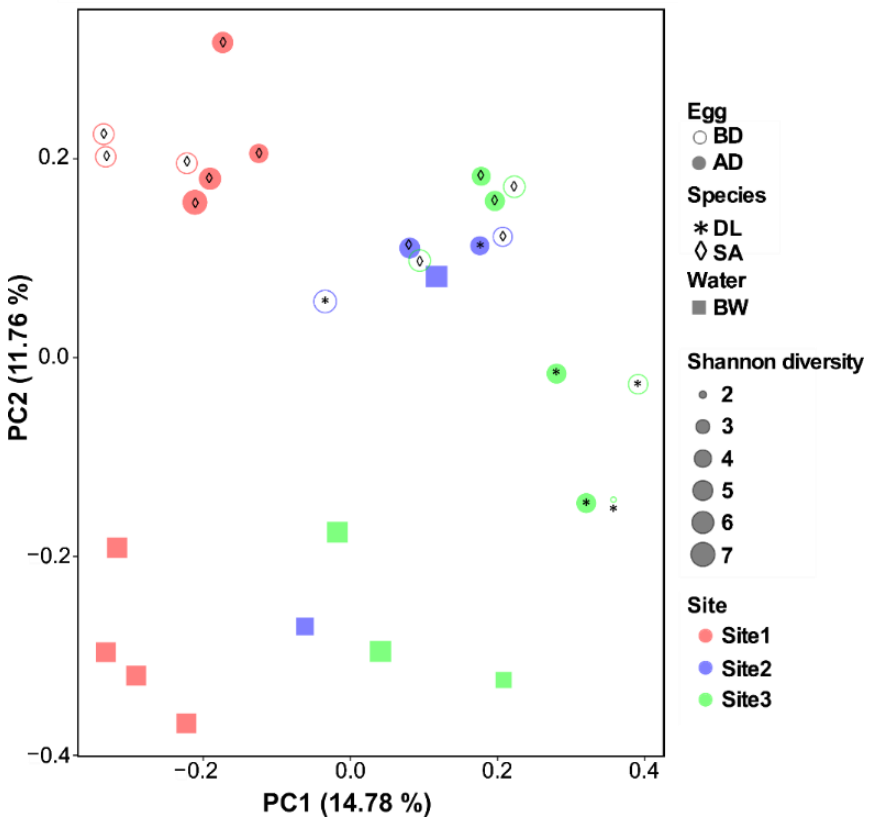


Figure 4.1.6. Visual representation of differences in the microbiota composition of eggs and water samples (beta diversity) using principal coordinates analysis (PcoA). Egg samples include eggs before (BD) and after (AD) disinfection from European seabass (DL, marked with * in the plot) and gilthead seabream (SA, marked with ◊ in the plot). Water samples were collected from broodstock tanks (BW) across three hatcheries/sites and also analyzed. PcoA analysis (Bray–Curtis distances) was run in the R environment using qiime2R and ggplot2 packages.

Table 4.1.2. PERMANOVA analysis (nr. of permutations = 1000) across all egg and water samples based on two factors: aquaculture site (Site 1, Site 2, and Site 3) and type of sample (egg vs. water).

	Df	Sum Sq	Mean Sq	F. Model	R2	Pr (>F)
Site	1	1.23	1.23	3.93	0.12	0.0009 ***
Sample type (eggs vs. water)	1	1.09	1.09	3.49	0.10	0.0009 ***
Residuals	25	7.85	0.31		0.77	
Total	27	10.19			1.00	

Df: degrees of freedom; Sq: square; Significant code: *** $p < 0.001$.

Table 4.1.3. PERMANOVA analysis (permutation = 1000) across egg samples based on three factors: aquaculture site (Site 1, Site 2, and Site 3), species (European seabass vs. gilthead seabream eggs), and disinfection (egg before vs. after disinfection).

	Df	Sum Sq	Mean Sq	F. Model	R2	Pr (>F)
Site	1	1.07	1.07	3.64	0.16	0.0009***
Species	1	0.70	0.70	2.37	0.10	0.0009 ***
Disinfection	1	0.17	0.17	0.57	0.02	0.9620
Site: Species	1	0.56	0.56	1.90	0.08	0.0059 **
Residuals	14	4.12	0.29		0.62	
Total	18	6.63			1.00	

Df: degrees of freedom; Sq: square; Significant code: *** $p < 0.001$, ** $p < 0.01$.

4.1.4.5. Relative Abundance

There were significant differences in OTU relative abundance ($p < 0.05$) according to site, sample type, and species (**Supplementary table 4.1.3**). Site comparisons of the bacterial abundance specified 73 OTUs with significantly different relative abundance ($p < 0.05$) and included 36 bacterial genera, such as *Oleispira*, *Colwellia*, *Psychrobium*, *Vibrio*, *Pseudoalteromonas*, *Psychromonas*, *Arcobacter* (listed in **Supplementary table 4.1.3**, Site). Comparison of the egg microbial community with broodstock water identified 36 OTUs with significantly different relative abundance ($p < 0.05$) and included 14 genera, such as *Pseudomonas*, *Salinirepens*, *Colwellia*, *Psychrobium*, *Leucothrix*, *Pseudophaeobacter* (listed in **Supplementary table 4.1.4**, Type). The relative abundance of 14 OTUs was significantly different in the comparison of gilthead seabream with European seabass eggs across the three

sites ($p < 0.05$, listed in **Supplementary table 4.1.4**, Species). *Pseudomonas* and *Photobacterium* were among the genera with higher relative abundance in European seabass eggs ($p < 0.05$), while *Vibrio* was more abundant in gilthead seabream eggs ($p < 0.05$). The genus of some OTUs with significant changes in their relative abundance was not identified in databases (represented by NA in **Supplementary table 4.1.4**). Egg disinfection did not cause statistically significant changes in OTU abundance in any of the sites ($p > 0.05$).

4.1.4.6. Functional Prediction

In general, Tax4Fun functional predictions identified 6311 KEGG orthologous and 133 KEGG pathways (**Supplementary table 4.1.5**). There was a significant association of 88 pathways with site, 7 pathways with disinfection, and 49 pathways with fish species ($p < 0.05$). No pathway was significantly associated with the type of sample (water vs. egg). The most abundant pathways were determined by mean of the relative abundance of the pathway across all samples (**Table 4.1.4**). Analysis of the "site" variable with the Kruskal–Wallis H test identified 10 significant pathways, based on the Benjamini–Hochberg-corrected *p*-value (BH): neomycin, kanamycin, and gentamicin biosynthesis (ko00524), fructose and mannose metabolism (ko00051), ascorbate and aldarate metabolism (ko00053), pentose and glucuronate interconversions (ko00040), tetracycline biosynthesis (ko00253), steroid hormone biosynthesis (ko00140), D-alanine metabolism (ko00473), ether lipid metabolism (ko00565), glycerolipid metabolism (ko00561), glycolysis/gluconeogenesis (ko00010). Ascorbate and aldarate metabolism (ko00053) and isoquinoline alkaloid biosynthesis (ko00950) were significantly changed for the "species" variable (BH, $p < 0.05$).

Table 4.1.4. The means of the relative abundance (%) of the top 10 pathways across the different variables.

Pathway		Site			Species		Type		Disinfection		All
KO	Name	S1	S2	S3	Sa	Dl	E	W	B	A	
ko01230	Biosynthesis of amino acids	6.7	7.4	7.2	6.9	7.4	7.0	7.1	7.0	7.1	7.0
ko01200	Carbon metabolism	6.9	6.7	6.6	6.7	6.6	6.7	6.8	6.7	6.7	6.7
ko00230	Purine metabolism	4.3	4.7	4.5	4.3	4.5	4.4	4.6	4.5	4.3	4.5
ko00240	Pyrimidine metabolism	3.1	3.3	3.3	3.1	3.4	3.2	3.4	3.3	3.1	3.2
ko00620	Pyruvate metabolism	2.9	2.9	2.9	2.9	2.9	2.9	3.0	2.9	2.9	2.9
ko00330	Arginine and proline metabolism	2.4	2.5	2.5	2.4	2.5	2.4	2.4	2.4	2.4	2.4
ko00260	Glycine, serine, and threonine metabolism	2.3	2.3	2.4	2.3	2.3	2.3	2.5	2.3	2.2	2.3
ko00250	Alanine, aspartate, and glutamate metabolism	2.3	2.3	2.3	2.3	2.2	2.3	2.3	2.3	2.3	2.3
ko00010	Glycolysis/Gluconeogenesis	2.1	2.2	2.4	2.1	2.4	2.2	2.3	2.2	2.2	2.2
ko00720	Carbon fixation pathways in prokaryotes	2.2	2.3	2.2	2.2	2.3	2.2	2.3	2.2	2.2	2.2

S1 = site 1, S2 = site 2, S3 = site 3; Sa = gilthead seabream, Dl = European seabass; E = egg, W = water; B = before, A = after.

4.1.4.7. Quantitative Analysis of 16S rRNA Gene

The total bacterial load of the eggs before and after disinfection, based on the quantification of the 16S rRNA gene, did not change significantly (**Figure 4.1.7**). However, the average load of total bacteria decreased after eggs' disinfection. No significant difference was detected in the bacterial load of eggs from different species, while the bacterial load of the seabass eggs was lower than that of the seabream eggs. A significant difference was observed in the total bacterial load of the eggs collected from site 1 compared to site 3 (**Figure 4.1.7**).

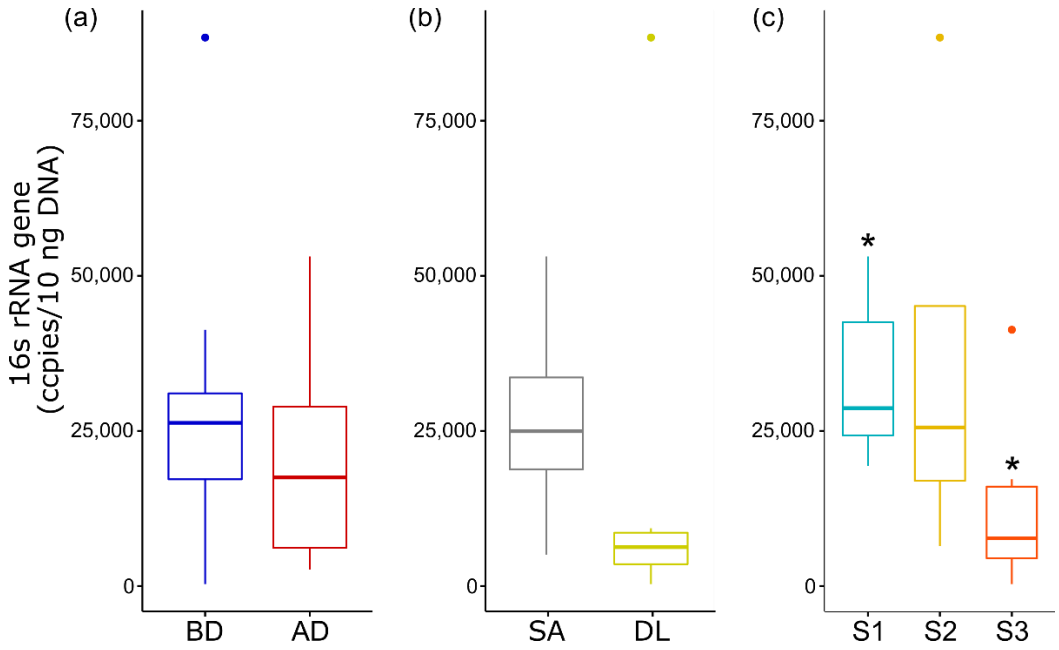


Figure 4.1.7. The total bacterial load of eggs from the three aquaculture sites quantified by qPCR of 16S rRNA. The box plots represent the different variables: (**a**) disinfection—before disinfection (BD) and after disinfection (AD); (**b**) species—gilthead seabream (SA) and European seabass (DL); (**c**) site—site 1 – 3. * Significance code: at $p < 0.05$.

4.1.5. Discussion

Metagenomic studies of European seabass and gilthead seabream egg microbiomes using next-generation sequencing have not been reported, despite the economic value of the species and the risks to production of diseases during the hatchery stage (Muniesa et al., 2020). The present study carried out metagenomic profiling to identify the bacterial communities associated with eggs and water from commercial hatcheries of European seabass and gilthead seabream.

The results of our study and previous studies on channel catfish eggs (Abdul Razak et al., 2019) and gilthead seabream larvae at 2 and 34 days post-hatch (dph, Califano et al., 2017) suggest that Proteobacteria and Bacteroidetes are the most abundant and probably most common phyla colonizing fish eggs and larval fish stages. However, our data revealed that the relative abundance of the main bacterial phyla varied with site, as shown by the high relative abundance of Cyanobacteria and relatively less abundant Firmicutes in most of the samples from site 3 and in some of the samples from site 2, compared to site 1. High relative abundance of Cyanobacteria was previously reported in the microbiota of eggs and larvae of the channel catfish and in juveniles of grass carp (*Ctenopharyngodon Idella*, Zeng et al., 2020),

both freshwater species. Since the broodstock water in sites 2 and 3 had a lower temperature (mean ≅ 16 °C) and salinity than site 1, environmental conditions including these two factors may explain the higher abundance of Cyanobacteria in these sites. Studies characterizing the growth and physiology of Cyanobacteria have demonstrated that temperature and salinity directly influence their growth (Silveira and Odebrecht, 2019). Corroborating evidence for the importance of salinity on fish-associated microbiota also comes from studies of the skin-associated microbiota in Atlantic salmon (*Salmo salar*) transitioning between fresh and saltwater. Firmicutes, Actinobacteria, Verrucomicrobia, and Cyanobacteria were more abundant in Atlantic salmon skin microbiota in freshwater compared to the skin microbiota in seawater, indicating that their abundance was highly affected by salinity (Lokesh and Kiron, 2016).

Most of the bacterial families with high relative abundance in the present study (*Vibrionaceae*, *Colwelliaceae*, *Pseudoalteromonadaceae*, *Alteromonadaceae*, *Shewanellaceae*, *Saccharospirillaceae*, and *Thiotrichaceae*) were also found to be among the top egg bacterial families in ballan wrasse at a commercial marine hatchery (Bone et al., 2020). Another abundant family, *Flavobacteriaceae*, with 17.6% average abundance in water and 6.8% in the egg samples, was also one of the most abundant families of bacteria in fertilized brown trout eggs but not in the ballan wrasse egg microbiota (Bone et al., 2020; Wilkins et al., 2015). This suggests that *Flavobacteriaceae* (Flavobacteriales order) may be an example of a site-specific bacterial family. *Cryomorphaceae* (another family of the Flavobacteriales order) and *Rhodobacteraceae* were among the most abundant families detected in the broodstock water of gilthead seabream from site 1. Interestingly, *Cryomorphaceae* was also highly abundant in the tank water of 34 dph gilthead seabream larvae compared to the tank water of 2 dph larvae, and *Rhodobacteraceae* was abundant at all stages in the tank water and in the food source (*Artemia nauplii*, Califano et al., 2017). *Rhodobacteraceae* were among the most abundant bacterial families in lumpfish (*Cyclopterus lumpus* L.) rearing water and eggs (Roalkvam et al., 2019). Taken together, the results of our study and previous studies (Califano et al., 2017; Roalkvam et al., 2019) highlight the likely importance of *Rhodobacteraceae* in marine fish egg microbiota and the relative and possibly significant contribution of microbiota present in water and food.

In our study, *Thiohalorhabdaceae* and *Arcobacteraceae* showed higher relative abundance on eggs collected from some of the broodstock than in the corresponding broodstock water. Comparing the results of the fish egg microbiota in the present and previous studies

(Bone et al., 2020; Roalkvam et al., 2019; Wilkins et al., 2015, 2016) reveals that *Thiohalorhabdaceae* and *Arcobacteraceae* were uncommon and not always present in fish egg microbiota. Geographical location/local conditions may account for the presence of these bacterial families in eggs from gilthead seabream and European seabass. Support for this idea comes from studies showing that algal samples collected from the French Tamaris coast situated in the Mediterranean sea were highly abundant in *Thiohalorhabdaceae* (Paix et al., 2020). Broodstock genetics and health status have been proposed to influence the egg microbial community of brown trout (Wilkins et al., 2016) and the gut and skin microbiota of the Atlantic Salmon (*Salmo salar*, Webster et al., 2018). It remains to be established if the higher relative abundance of some bacterial families (e.g., *Thiohalorhabdaceae*) in the egg microbiota in our study using broodstock from different companies and species was not only driven by the surrounding water but also broodstock genetics.

Direct and indirect factors that can cause pathogenic bacteria to spread in hatcheries include broodstock health, diet, water, and vertical transmission from gonadal fluids (Migaud et al., 2013; Pradeep et al., 2016; Vliet et al., 2015). Therefore, providing an extremely clean environment is one of the key challenges in hatcheries and, for this reason, egg disinfection postfertilization is a common practice, with the aim of eliminating opportunistic diseases (Swaef et al., 2016). A range of disinfectants and protocols based on hydrogen peroxide, glutaraldehyde, ozone, and iodophors are available for the treatment of eggs to decrease bacterial/fungal loads, and their effectiveness depends on factors such as pH and temperature (Swaef et al., 2016). Analysis of the impact of the iodine-based disinfectant protocols on the egg microbiota in the present study suggested that they were largely ineffective and did not significantly impact the alpha- or beta-diversity, the relative abundance, or the total bacterial load of the bacterial community of seabass and seabream eggs. The results of previous studies using conventional microbiological approaches are contradictory in relation to the effectiveness of iodine-based disinfectants on specific pathogenic bacteria of the egg surface, e.g., ranging from total removal to the absence of an effect of iodophors on *Flavobacterium psychrophilum* (Swaef et al., 2016), and total bacterial loads in cultures of egg bacterial communities, including *Vibrio* spp. (CAN et al., 2010). Two factors are proposed to explain the poor disinfection capacity of iodine-based treatments: (i) the majority of bacteria are inside the eggs or strongly attached to the egg chorion and so superficial disinfection protocols have no effect, and/or (ii) the current protocols for iodine-based disinfectants are inadequate and optimization of disinfection protocols is required. An important caveat of most microbiome studies, including the present study, is that 16S rRNA can be amplified from both viable and

dead bacteria. Disinfectants such as iodine rapidly penetrate microorganisms and attack key groups of proteins (McDonnell and Russell, 1999) and it is possible that, because the egg samples were taken immediately after treatment, DNA from dead bacteria still had not broken down. Further studies are needed that couple metagenomics and the assessment of bacterial viability to provide a better understanding of the efficacy of iodine-based disinfectants.

The egg and water microbial communities had significantly different beta-diversity and abundance, and included 36 out of 2,444 significantly different OTUs between the two sample types. This suggests that a relatively small number of bacteria may explain the separation between the microbiota of broodstock water and eggs observed in the PcoA analysis. However, with the higher relative abundance of some potentially pathogenic bacteria in gilthead seabream and European seabass eggs, such as *Pseudomonas*, *Pseudoalteromonas*, *Leucothrix*, and *Arcobacter*, more studies are needed to understand their growth dynamics and pathogenicity to establish risk. The presence in production systems of pathogenic bacteria such as some *Vibrio* and *Photobacterium* species was reported to cause mass mortality of gilthead seabream and European seabass larvae and juveniles (Abdel-Aziz et al., 2013). Bacterial pathogens of adult fish such as *Pseudomonas* and *Flavobacterium psychrophilum*, which cause bacterial cold-water disease (BCWD), have previously been detected at the egg stage of brown trout (*Salmo trutta*, Wilkins et al., 2015). Furthermore, mass mortalities of cod (*Gadus morhua*) eggs identified, as the causal factor, a pathogenic bacterial species, *Leucothrix mucor* (Johnson et al., 1971). Although we did not detect large variation in the microbial community abundance between European seabass and gilthead seabream eggs, the relatively higher abundance of two *Pseudomonas* OTUs and one *Photobacterium* OTU in the European seabass eggs were examples of species-specific OTUs. Therefore, the results of the present study support the idea proposed from observations of cod and halibut eggs in the 1980s that the ability of bacteria to colonize eggs may depend on the nature of the egg chorion or presence of bactericidal enzymes (e.g., lysozyme and lectins, Hansen and Olafsen, 1989). It will be important in the future to study the functional role of the egg microbiota and to establish which are beneficial and which are pathogenic members of the microbial community. The use of metagenomics and complimentary approaches such as PCR, quantitative PCR, and bacterial culture will extend our understanding of the contribution of the microbial community to fish egg physiology and quality (Fusco and Quero, 2012; O et al., 2007).

The 16S rRNA gene qPCR analysis and the functional inference analysis further supported the site/geographical location's effect on the egg microbial community and suggested

that “site” was the main factor determining the egg microbial community compared to all the other factors analyzed. The association of some pathways such as caprolactam degradation, geraniol degradation, and benzoate degradation with disinfection suggests that, in addition to the elimination of bacteria, they may favor the maintenance of some bacterial genera or species. The reported antibacterial effects on some bacterial genera of benzoate (Metwally and Mohamed, 2020), geraniol (KIM et al., 1995; Pontes et al., 2019)**,** and caprolactam and the degradation of these compounds by others further support the idea of alternative secondary effects for disinfectants (Baxi, 2013; Höschle and Jendrossek, 2005; Uhlik et al., 2012).

The most abundant bacterial genera were highlighted across all eggs and water samples (**Figure 4.1.8**). In general, an association of the microbiome with the site, species, and type of sample (water/egg) was observed. The contribution of the broodstock water to the egg microbiome was evident from the detection of a similar bacterial composition in eggs and water. The bacterial genera with significant changes in their relative abundance were also identified based on variables such as site, species, and type of sample (eggs and water). Optimized methods for sample collection and processing were developed, and the usefulness of current disinfection protocols for eggs collected from broodstock water was determined. The similarity and differences found between gilthead seabream and European seabass egg microbiomes were identified. The contribution of different factors (species, site, and disinfection) to the bacterial community of eggs and broodstock water in commercial hatcheries was established and this knowledge will contribute to the development of future strategies for hatchery management.

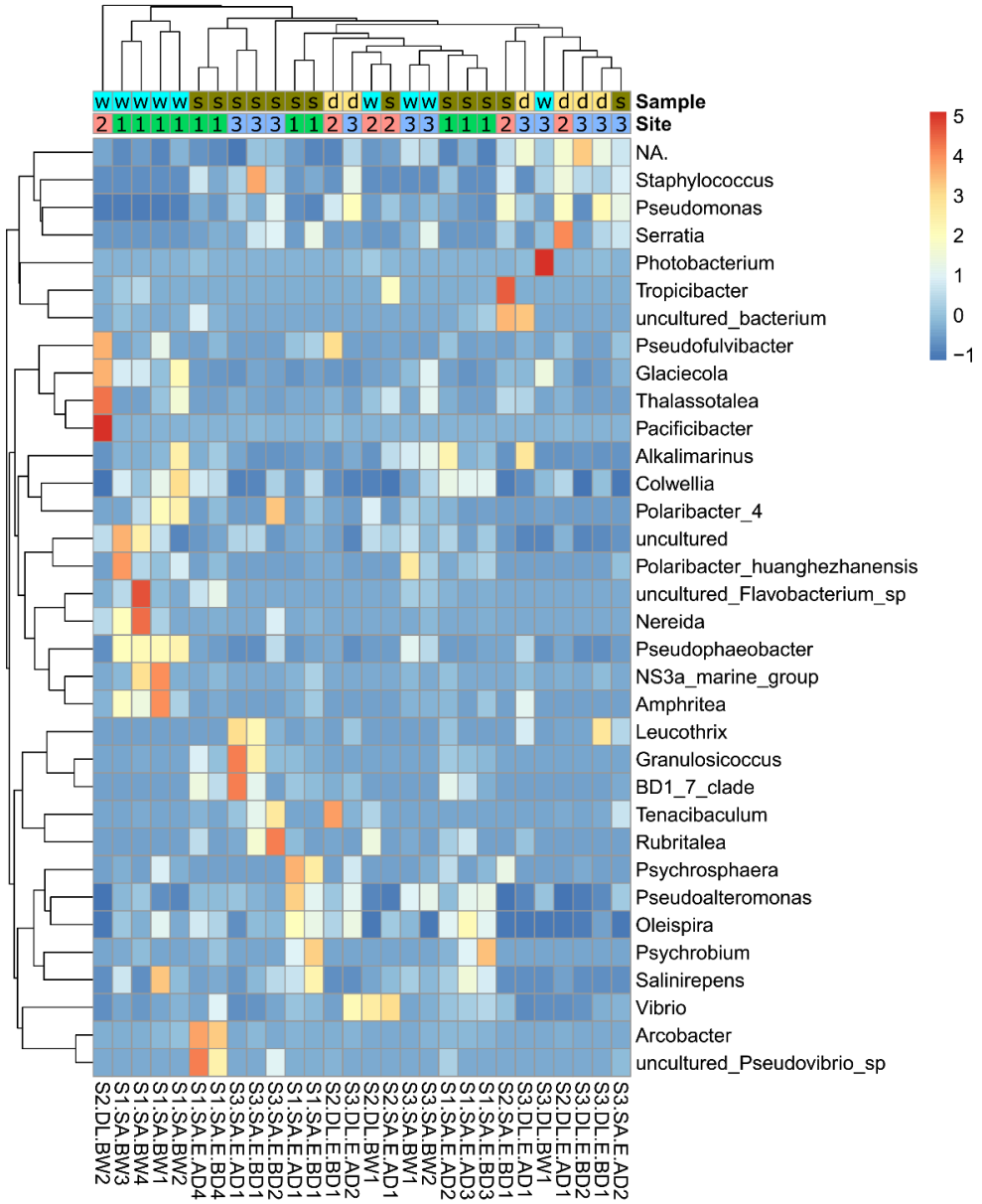


Figure 4.1.8. The heatmap of top bacterial genera across eggs and broodstock holding water. Sample: gilthead seabream (s), European seabass (d), water (w); site: site 1–3.

In summary, the results show that Proteobacteria are the preponderant phyla of the bacterial community found in both European seabass and gilthead seabream eggs (before or after disinfection) and also in broodstock water across different hatcheries. *Vibrio* spp. was the most highly represented genus (7.1% on average), followed by *Glaciecola* (4.8%), *Pseudoalteromonas* (4.4%), and *Colwellia* (4.2%), in eggs and water across the sites. Site, species, and type of sample (eggs or water) drove the microbial community structure and influenced microbiome functional profiles. The overall composition and the relative abundance

of bacterial genera of the eggs and water microbiome were similar, but differences were found in the relative abundance of some bacterial genera/OTUs. This indicates that the water microbiome makes a high contribution to the eggs' bacterial communities. The results of our study highlight the need for further investigation into the egg microbiome and the importance for hatcheries of optimized disinfection protocols that take into consideration the initial bacterial composition, disinfectant composition, and the species. The risk of some potential pathogenic species related to the *Pseudomonas*, *Pseudoalteromonas*, *Leucothrix*, and *Arcobacter* genera, with high relative abundance in egg samples compared to water, needs to be assessed in future studies.

4.1.6. See ANNEX IV for Supplementary materials

Supplementary table 4.1.1. Sequencing statistics of the 28 libraries prepared from European seabass and gilthead seabream eggs before and after disinfection and from broodstock water. The number of raw, merged and filtered sequence reads obtained for each sample are presented.

Supplementary table 4.1.2. Bacterial composition and their relative abundance in each of the samples. A total of 294 OTUs were selected, using as the criteria their presence at >1% relative abundance in the 28 samples studied. NA means that the OTU at the taxonomic level was not identified through comparison with the SILVA database. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.1.3. The top 15 genera detected in each egg and water sample. The relative percentage of each genus was calculated in relation to all detected genera in each sample. Information on sample labels is provided in **Figure 4.1.1**. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.1.4. Differential abundance analysis of microbiota of eggs and water using three factors: Site (Site 1, Site 2, Site 3), type of sample (egg vs. water), and species (European seabass vs. gilthead seabream). OTUs with significantly differential (relative) abundance are presented in different Excel sheets in the table for site, broodstock water, egg, and species. The R package ALDEx2 (1.20.0) was run to identify the OTUs that were differentially abundant across the conditions tested. The centered log-ratio-transformed values were analyzed with a general linear model (glm) and a Kruskal–Wallis test applied. The ALDEx2 output included the expected p-value of a Kruskal–Wallis test (kw.ep), expected Benjamini–Hochberg-corrected p-value of a Kruskal–Wallis test (kw.eBH), expected p-value of a glm test (glm.ep), and expected Benjamini–Hochberg-corrected p-value of a glm test (glm.eBH). This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.1.5. KEGG ontology terms and association of the pathways with the site (Site 1, Site 2, Site 3), disinfection usage, and species (European seabass and gilthead seabream). Functional prediction was performed using Tax4Fun and identified 6311 KEGG

Orthology terms and 133 pathways. Functional predictions of the metagenomics profiles were performed using the web-based platform Microbiome Analyst and Tax4Fun. The associations between functional categories and the experimental factor (site, species, disinfection) were tested based on the global test algorithm and are presented in different Excel sheets in the folder. Statistic.Q and Expected.Q = test for homogeneity; Pval = p-value; Holm.p = p-value for Holm–Bonferroni method; FDR = false discovery rate. This is only available in on digital format in Annex IV because the table is very extensive.

4.1.7. References

Abdel-Aziz, M., Eissa, A.E., Hanna, M., Okada, M.A., 2013. Identifying some pathogenic *Vibrio*/*Photobacterium* species during mass mortalities of cultured Gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) from some Egyptian coastal provinces. Int. J. Vet. Sci. Med. 1, 87–95. https://doi.org/10.1016/J.IJVSM.2013.10.004

Abdul Razak, S., Griffin, M.J., Mischke, C.C., Bosworth, B.G., Waldbieser, G.C., Wise, D.J., Marsh, T.L., Scribner, K.T., 2019. Biotic and abiotic factors influencing channel catfish egg and gut microbiome dynamics during early life stages. Aquaculture 498, 556–567. https://doi.org/10.1016/j.aquaculture.2018.08.073

Andrews, S., 2010. Fastqc a quality control tool for high throughput sequence data. Babraham Bioinformatics

Apprill, A., McNally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75, 129–137. https://doi.org/10.3354/ame01753

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Baxi, N.N., 2013. Influence of ε-caprolactam on growth and physiology of environmental bacteria. Ann. Microbiol. 63, 1471–1476. https://doi.org/10.1007/s13213-013-0610-4

Bone, A., Bekaert, M., Papadopoulou, A., McMillan, S., Adams, A., Davie, A., Desbois, A.P., 2020. Bacterial communities of ballan wrasse (*labrus bergylta*) eggs at a commercial marine hatchery. Curr. Microbiol. 1, 3. https://doi.org/10.1007/s00284-020-02286-8

Borges, N., Keller-Costa, T., Sanches-Fernandes, G.M.M., Louvado, A., Gomes, N.C.M., Costa, R., 2021. Bacteriome structure, function, and probiotics in fish larviculture: the good, the bad, and the gaps. Annu. Rev. Anim. Biosci. 9. https://doi.org/10.1146/annurev-animal-062920-113114

Brugman, S., Ikeda-Ohtsubo, W., Braber, S., Folkerts, G., Pieterse, C.M.J., Bakker, P.A.H.M., 2018. A comparative review on microbiota manipulation: lessons from fish, plants, livestock, and human research. Front. Nutr. 5, 80. https://doi.org/10.3389/FNUT.2018.00080/BIBTEX

Califano, G., Castanho, S., Soares, F., Ribeiro, L., Cox, C.J., Mata, L., Costa, R., 2017. Molecular taxonomic profiling of bacterial communities in a gilthead seabream (*Sparus aurata*) hatchery. Front. Microbiol. 8, 204. https://doi.org/10.3389/FMICB.2017.00204/BIBTEX

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Can, E., Saka, Ş., Firat, M.K., 2010. Disinfection of gilthead sea bream (*Sparus aurata*), red porgy (*Pagrus pagrus*), and common dentex (*Dentex dentex*) eggs from sparidae with different disinfectants. Kafkas Üniversitesi Vet. Fakültesi Derg. 16, 299–306.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pẽa, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., et al. 2010. QIIME

allows analysis of high-throughput community sequencing data. Nat. Methods. https://doi.org/10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Chong, J., Liu, P., Zhou, G., Xia, J., 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat. Protoc. 15, 799–821. https://doi.org/10.1038/s41596-019-0264-1

Swaef, E.D., Van den Broeck, W., Dierckens, K., Decostere, A., 2016. Disinfection of teleost eggs: a review. Rev. Aquac. 8, 321–341. https://doi.org/10.1111/raq.12096

Egerton, S., Culloty, S., Whooley, J., Stanton, C., Ross, R.P., 2018. The gut microbiota of marine fish. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.00873

FEAP Annual Report, 2020.

Fusco, V., Quero, G.M., 2012. Nucleic acid-based methods to identify, detect and type pathogenic bacteria occurring in milk and dairy products. Struct. Funct. Food Eng. https://doi.org/10.5772/49937

Gilbert, S.F., Sapp, J., Tauber, A.I., 2012. A symbiotic view of life: we have never been individuals. Q. Rev. Biol. 87, 325–341. https://doi.org/10.1086/668166

Goeman, J.J., Van de Geer, S., De Kort, F., van Houwellingen, H.C., 2004. A global test for groups fo genes: testing association with a clinical outcome. Bioinformatics 20, 93–99. https://doi.org/10.1093/bioinformatics/btg382

Groff, J.M., 2001. Cutaneous biology and diseases of fish. Vet. Clin. North Am. Exot. Anim. Pract. 4, 321–411. https://doi.org/10.1016/S1094-9194(17)30037-3

Hadrich, D., 2018. Microbiome research is becoming the key to better understanding health and nutrition. Front. Genet. 9, 212. https://doi.org/10.3389/fgene.2018.00212

Hansen, G.H., Olafsen, J.A., 1999. Bacterial interactions in early life stages of marine cold water fish. Microb. Ecol. https://doi.org/10.1007/s002489900158

Hansen, G.H., Olafsen, J.A., 1989. Bacterial colonization of cod (*Gadus morhua* L.) and halibut (*Hippoglossus hippoglossus*) eggs in marine aquaculture. Appl. Environ. Microbiol. 55, 1435–1446. https://doi.org/10.1128/aem.55.6.1435-1446.1989

Höschle, B., Jendrossek, D., 2005. Utilization of geraniol is dependent on molybdenum in *Pseudomonas aeruginosa*: evidence for different metabolic routes for oxidation of geraniol and citronellol. Microbiology 151, 2277–2283. https://doi.org/10.1099/mic.0.27957-0

Johnson, P.W., Sieburth, J.M., Sastry, A., Arnold, C.R., Doty, M.S., 1971. *Leucothrix mucor* infestation of benthic crustacea, fish eggs, and tropical algae1. Limnol. Oceanogr. 16, 962–969. https://doi.org/10.4319/lo.1971.16.6.0962

Kim, J.M., Marshall, Cornell, J.A., Iii, J.F.P., Wei, C.I., 1995. Antibacterial activity of carvacrol, citral, and geraniol against *Salmonella typhimurium* in culture medium and on fish cubes. J. Food Sci. 60, 1364–1368. https://doi.org/10.1111/j.1365-

2621.1995.tb04592.x

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1. https://doi.org/10.1093/nar/gks808

Legrand, T.P.R.A., Catalano, S.R., Wos-Oxley, M.L., Stephens, F., Landos, M., Bansemer, M.S., Stone, D.A.J., Qin, J.G., Oxley, A.P.A., 2018. The inner workings of the outer surface: skin and gill microbiota as indicators of changing gut health in yellowtail kingfish. Front. Microbiol. 8, 2664. https://doi.org/10.3389/fmicb.2017.02664

Legrand, T.P.R.A., Wynne, J.W., Weyrich, L.S., Oxley, A.P.A., 2020. A microbial sea of possibilities: current knowledge and prospects for an improved understanding of the fish microbiome. Rev. Aquac. 12, 1101–1134. https://doi.org/10.1111/raq.12375

Llewellyn, M.S., Boutin, S., Hoseinifar, S.H., Derome, N., 2014. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front. Microbiol. 5, 1–1. https://doi.org/10.3389/FMICB.2014.00207

Lokesh, J., Kiron, V., 2016. Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. Sci. Rep. 6, 1–10. https://doi.org/10.1038/srep19707

McDonnell, G., Russell, A.D., 1999. Antiseptics and disinfectants: activity, action, and resistance. Clin. Microbiol. Rev. 12, 147.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H. V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., et al., 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1218525110

Merrifield, D.L., Rodiles, A., 2015. The fish microbiome and its interactions with mucosal tissues, in: Mucosal Health in Aquaculture. Elsevier Inc., pp. 273–295. https://doi.org/10.1016/B978-0-12-417186-2.00010-8

Metwally, N.H., Mohamed, M.S., 2020. New imidazolone derivatives comprising a benzoate or sulfonamide moiety as anti-inflammatory and antibacterial inhibitors: Design, synthesis, selective COX-2, DHFR and molecular-modeling study. Bioorg. Chem. 99, 103438. https://doi.org/10.1016/j.bioorg.2019.103438

Migaud, H., Bell, G., Cabrita, E., Mcandrew, B., Davie, A., Bobe, J., Herráez, M.P., Carrillo, M., 2013. Gamete quality and broodstock management in temperate fish. Rev. Aquac. https://doi.org/10.1111/raq.12025

Muniesa, A., Basurco, B., Aguilera, C., Furones, D., Reverté, C., Sanjuan-Vilaplana, A., Jansen, M.D., Brun, E., Tavornpanich, S., 2020. Mapping the knowledge of the main diseases affecting sea bass and sea bream in Mediterranean. Transbound. Emerg. Dis. 67, 1089–1100. https://doi.org/10.1111/tbed.13482

Newaj-Fyzul, A., Al-Harbi, A.H., Austin, B., 2014. Review: developments in the use of probiotics for disease control in aquaculture. Aquaculture. https://doi.org/10.1016/j.aquaculture.2013.08.026

O, L., FJ, D.C., FX, M., 2007. Pathogen detection: a perspective of traditional methods and biosensors. Biosens. Bioelectron. 22, 1205–1217.

https://doi.org/10.1016/J.BIOS.2006.06.036

Olafsen, J.A., 2001. Interactions between fish larvae and bacteria in marine aquaculture, in: Aquaculture. Elsevier, pp. 223–247. https://doi.org/10.1016/S0044-8486(01)00702-5

Paix, B., Carriot, N., Barry-Martinet, R., Greff, S., Misson, B., Briand, J.F., Culioli, G., 2020. A multi-omics analysis suggests links between the differentiated surface metabolome and epiphytic microbiota along the thallus of a mediterranean seaweed holobiont. front. microbiol. 11. https://doi.org/10.3389/fmicb.2020.00494

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N., 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 176, 649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Pickard, J.M., Zeng, M.Y., Caruso, R., Núñez, G., 2017. Gut microbiota: role in pathogen colonization, immune responses, and inflammatory disease. Immunol. Rev. https://doi.org/10.1111/imr.12567

Pinto, P.I.S., Guerreiro, C.C., Costa, R.A., Martinez-Blanch, J.F., Carballo, C., Codoñer, F.M., Manchado, M., Power, D.M., 2019. Understanding pseudo-albinism in sole (*Solea senegalensis*): a transcriptomics and metagenomics approach. Sci. Rep. 9, 1–16. https://doi.org/10.1038/s41598-019-49501-6

Pontes, E.K.U., Melo, H.M., Nogueira, J.W.A., Firmino, N.C.S., de Carvalho, M.G., Catunda Júnior, F.E.A., Cavalcante, T.T.A., 2019. Antibiofilm activity of the essential oil of citronella (*Cymbopogon nardus*) and its major component, geraniol, on the bacterial biofilms of *Staphylococcus aureus*. Food Sci. Biotechnol. 28, 633–639. https://doi.org/10.1007/s10068-018-0502-2

Pradeep, P.J., Suebsing, R., Sirthammajak, S., Kampeera, J., Jitrakorn, S., Saksmerprome, V., Turner, W., Palang, I., Vanichviriyakit, R., Senapin, S., Jeffs, A., Kiatpathomchai, W., Withyachumanarnkul, B., 2016. Evidence of vertical transmission and tissue tropism of Streptococcosis from naturally infected red tilapia (*Oreochromis* spp.). Aquac. Reports 3, 58–66. https://doi.org/10.1016/j.aqrep.2015.12.002

Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., Buck, G.A., Snyder, M.P., Strauss, J.F., Weinstock, G.M., White, O., Huttenhower, C., 2019. The integrative human microbiome project. Nature 569, 641–648. https://doi.org/10.1038/s41586-019-1238-8

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590. https://doi.org/10.1093/NAR/GKS1219

Ringø, E., Dimitroglou, A., Hoseinifar, S.H., Davies, S.J., 2014. Prebiotics in finfish: an update, in: Aquaculture Nutrition. John Wiley & Sons, Ltd, Chichester, UK, pp. 360–400. https://doi.org/10.1002/9781118897263.ch14

Roalkvam, I., Drønen, K., Dahle, H., Wergeland, H.I., 2019. Microbial communities in a flow-through fish farm for lumpfish (*Cyclopterus lumpus* l.) during healthy rearing conditions. Front. Microbiol. 10, 1594. https://doi.org/10.3389/fmicb.2019.01594

Silveira, S.B., Odebrecht, C., 2019. Effects of salinity and temperature on the growth, toxin production, and akinete germination of the *Cyanobacterium nodularia* spumigena. Front. Mar. Sci. 6, 339. https://doi.org/10.3389/fmars.2019.00339

Simon, C., Daniel, R., 2011. Metagenomic analyses: Past and future trends. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.02345-10

Tellez, G., Higgins, S.E., Donoghue, A.M., Hargis, B.M., 2006. Digestive physiology and the role of microorganisms, in: Journal of Applied Poultry Research. Poultry Science Association, pp. 136–144. https://doi.org/10.1093/japr/15.1.136

Uhlik, O., Wald, J., Strejcek, M., Musilova, L., Ridl, J., Hroudova, M., Vlcek, C., Cardenas, E., Mackova, M., Macek, T., 2012. Identification of bacteria utilizing biphenyl, benzoate, and naphthalene in long-term contaminated soil. PLoS One 7, e40653. https://doi.org/10.1371/journal.pone.0040653

Webster, T.M.U., Consuegra, S., Hitchings, M., Garcia de Leaniz, C., 2018. Interpopulation variation in the atlantic salmon microbiome reflects environmental and genetic diversity. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/aem.00691-18

Vadstein, O., Bergh, Ø., Gatesoupe, F.-J., Galindo-Villegas, J., Mulero, V., Picchietti, S., Scapigliati, G., Makridis, P., Olsen, Y., Dierckens, K., Defoirdt, T., Boon, N., De Schryver, P., Bossier, P., 2013. Microbiology and immunology of fish larvae. Rev. Aquac. 5, S1–S25. https://doi.org/10.1111/j.1753-5131.2012.01082.x

Vliet, D.V., Loch, T.P., Faisal, M., 2015. *Flavobacterium psychrophilum* Infections in salmonid broodstock and hatchery-propagated stocks of the great lakes basin. J. Aquat. Anim. Health 27, 192–202. https://doi.org/10.1080/08997659.2015.1088488

Wang, A.R., Ran, C., Ringø, E., Zhou, Z.G., 2018. Progress in fish gastrointestinal microbiota research. Rev. Aquac. 10, 626–640. https://doi.org/10.1111/RAQ.12191

Wilkins, L.G.E., Fumagalli, L., Wedekind, C., 2016. Effects of host genetics and environment on egg-associated microbiotas in brown trout (*Salmo trutta*). Mol. Ecol. 25, 4930–4945. https://doi.org/10.1111/mec.13798

Wilkins, L.G.E., Rogivue, A., Schütz, F., Fumagalli, L., Wedekind, C., 2015. Increased diversity of egg-associated bacteria on brown trout (*Salmo trutta*) at elevated temperatures. Sci. Rep. 5, 17084. https://doi.org/10.1038/srep17084

Yoder, M., Ley, I.T. De, King, I.W., Mundo-Ocampo, M., Mann, J., Blaxter, M., Poiras, L., Ley, P. De, 2006. DESS: a versatile solution for preserving morphology and extractable DNA of nematodes. Nematology 8, 367–376. https://doi.org/10.1163/156854106778493448

Zeng, A., Tan, K., Gong, P., Lei, P., Guo, Z., Wang, S., Gao, S., Zhou, Y., Shu, Y., Zhou, X., Miao, D., Zeng, F., Liu, H., 2020. Correlation of microbiota in the gut of fish species and

water. 3 Biotech 10, 472. https://doi.org/10.1007/s13205-020-02461-5

CHAPTER 4.2

Quantifying dominant bacterial genera detected in fish eggs and larvae metagenomic 16S rRNA gene data using genus-specific primers

Quantifying dominant bacterial genera detected in fish eggs and larvae metagenomic 16S rRNA gene data using genus-specific primers

Manuscript published in MicrobiologyOpen, 11 (3) 2022, e1274.
https://doi.org/10.1002/mbo3.1274

Acknowledgments: The authors acknowledge all the aquaculture hatcheries for supplying the experiment samples. The authors also thank João C. R. Cardoso of the Comparative Endocrinology and Integrative Biology group, CCMAR, for his helpful comments. European Union Horizon 2020 Programme (PerformFISH, Grant nº 727610; AQUAEXCEL3.0, Grant nº871108) funded this study and the PhD grant to B.N. This output reflects only the authors' views and the European Union cannot be held responsible for any use that may be made of the information contained herein. Additional funding was obtained from the Portuguese Foundation for Science and Technology (FCT) through project UID/Multi/04326/2020 and researcher contract DL57/2016/CP1361/CT001 to P. I. S. P. and from CRESC Algarve 2020 and COMPETE 2020 through the project EMBRC. PT ALG-01-0145-FEDER-022121.

Quantifying dominant bacterial genera detected in fish eggs and larvae metagenomic 16S rRNA gene data using genus-specific primers

Babak Najafpour, Patricia I.S. Pinto, A.V.M. Canario and Deborah M. Power*

Centro de Ciências do Mar (CCMAR/CIMAR), Universidade do Algarve, 8005-139 Faro, Portugal

CReDit statement:

Babak Najafpour: Methodology Optimization - Lead, Bioinformatics - Lead, Formal analysis, Validation - Lead, Investigation - Lead, Visualization, Writing – Original Draft, Review & Editing; Patricia Pinto: Methodology-Supporting, Review & Editing; Adelino V. M. Canario, Analysis-Supporting, Resources-Supporting, Supervision-Supporting, Review & Editing-Supporting; Deborah M. Power: Review & Editing, Supervision, Project Administration, Funding Acquisition.

4.2.1. Abstract

The goal of this study was to design genus-specific primers for rapid evaluation of the most abundant bacterial genera identified using amplicon-based sequencing of the 16S rRNA gene in fish-related samples and surrounding water. Efficient genus-specific primers were designed for eleven bacterial genera including *Alkalimarinus*, *Colwellia*, *Enterovibrio*, *Marinomonas*, *Massilia*, *Oleispira*, *Phaeobacter*, *Photobacterium*, *Polaribacter*, *Pseudomonas* and *Psychrobium*. The specificity of the primers was confirmed by the phylogeny of the sequenced polymerase chain reaction (PCR) amplicons that indicated primers were genus specific except in the case of *Colwellia* and *Phaeobacter*. Copy number of the 16S rRNA gene obtained by quantitative PCR using genus-specific primers and the relative abundance obtained by 16S rRNA gene sequencing using universal primers were well correlated for the five analysed abundant bacterial genera. Low correlations between quantitative PCR and 16S rRNA gene sequencing for *Pseudomonas* was explained by the higher coverage of known Pseudomonas species by the designed genus-specific primers than the universal primers used in 16S rRNA gene sequencing. The designed genus-specific primers are proposed as rapid and cost-effective tools to evaluate the most abundant bacterial genera in fish-related or potentially other metagenomics samples.

Keywords: abundant microbiota, aquaculture site, eggs, genus-specific primers, larvae, qPCR, 16S rRNA gene

4.2.2. Introduction

The influence of symbiotic and pathogenic interactions of bacterial microbiota on terrestrial and aquatic vertebrates is of high interest (Sharpton, 2018). The advent of cost-effective next generation sequencing (NGS) has opened-up new avenues of culture-free and high throughput analysis of the entire microbiota in any ecosystem and has significantly modified understanding of their role in animal health and disease (Cao et al., 2017). Since 16S rRNA gene is present in all bacteria, it is the most common reference gene for studies of bacterial phylogeny and taxonomy, and also for studies of the composition and the relative proportion of microorganisms in a given habitat (Janda and Abbott, 2007; Simon and Daniel, 2011). The structure of the 16S rRNA gene explains its versatility for metagenomics, since universal primers can be designed in regions that are highly conserved across species and the intervening hypervariable regions can be used to assign operational taxonomic units (OTU) to the genus taxon (Baker et al., 2003; Wang and Qian, 2009). Furthermore, the hypervariable regions offer the opportunity for the development of genus and even species targeted quantitative PCR (qPCR).

Aquatic organisms are exposed to a ubiquitous and abundant microbiota and intensification of aquaculture has increased interest in characterising the microbiota of fish. Initially traditional microbiological approaches based on in vitro culture were used, but more recently metagenomics approaches have been deployed (Martínez-Porchas and Vargas-Albores, 2017). Aquaculture differs from terrestrial farming systems as there is a much larger number of species exploited and this is coupled to a wide variety of environmental conditions (e.g. temperature, salinity) and geographical locations (Casa-Resino et al., 2021). The variability of fish microbiota has been linked to geography, species and environmental conditions, and indicates that fish microbiomes have a degree of farm site-specificity (Najafpour et al., 2021). Metagenomics studies targeting fish, indicate that *Vibrio* and *Pseudomonas* are the dominant bacterial genera (reviewed by Egerton et al., 2018). However, the exclusive dependence on relative abundance data from NGS can lead to misinterpretation of microbial community structure (Jian et al., 2020). For this reason, it has been proposed that the use of genus-specific primers can provide complimentary, quantitative data, to corroborate NGS results and contribute to better understand microbial diversity and population structure (Zhou et al., 2014). A general literature review of microbiome studies in fish reveals that genus-specific primers for the most representative bacterial genera with high relative abundance are unavailable. Although genus-specific primers exist for the detection of pathogen-containing

genera such as *Aeromonas*, *Vibrio*, *Edwardsiella* and *Streptococcus*, their use has not been correlated with 16S rRNA metagenomic profiles in fish (Zhang et al., 2014).

We previously generated 16S rRNA metagenomics datasets for gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) eggs from several commercial production sites in Europe and identified the profile of the main bacterial genera (Najafpour et al., 2021). The objective of the present study was to develop a quick, cost effective and practical approach for large scale screening of the core microbiome during aquaculture production cycles. We report the design of genus-specific primers, exploiting the hypervariable characteristics of the 16S rRNA gene, for the dominant bacterial genera represented in our in-house metagenomic 16S rRNA gene datasets from eggs, larvae, live feed, and tank water samples from seabream and seabass aquaculture sites in Europe (Najafpour et al., 2021). The efficiency and specificity of the genus-specific primers were confirmed by quantitative PCR (qPCR) and sequencing of the PCR amplicons, and coverage of each genus was validated by comparison of qPCR and metagenomics data (Najafpour et al., 2021).

4.2.3. Experimental procedures

4.2.3.1. Selection of target genera

The summarized workflow for genus selection and primer design is presented in **Figure 4.2.1**. The most represented bacterial genera in seabream and seabass hatcheries in Europe were identified using in-house metagenomic datasets of 16S rRNA gene sequences obtained from eggs, larvae, live feed and hatchery water (sequenced by Lifesequencing S.L.-ADM, Spain and Stab Vida, Lda, Portugal). The corresponding 16S rRNA gene sequences of target genera were obtained from the LPSN and Silva (SSU r138.1) databases (Parte et al., 2020; Quast et al., 2013).

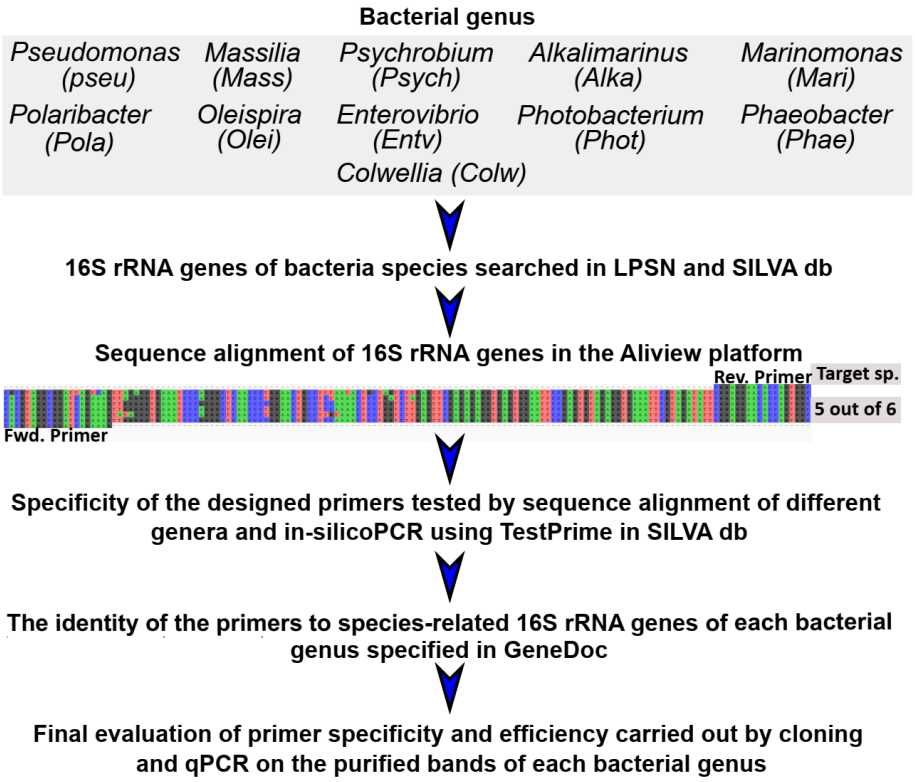


Figure 4.2.1. The workflow followed for the design of the bacterial 16S rRNA genus-specific primers for the most abundant genera in the microbiome of fish larvae, fish eggs, zooplankton, phytoplankton, and water. The most abundant genus detected in each data set was ranked using an in-house 16S rRNA gene database. The egg microbiome profile obtained from 16S rRNA gene sequencing has previously been reported (Chapter 4.1, Najafpour et al., 2021).

4.2.3.2. Sequence alignment and primer design

Multiple sequence alignments of the retrieved 16S rRNA gene sequences from the LPSN and Silva databases were performed using the MUSCLE algorithm (Edgar, 2004) in the Aliview platform v 1.27 (Larsson, 2014). Genus-specific forward (Fw) and reverse (Rv) primers were designed manually with the objective of obtaining primers that amplified the maximum number of species in each of the target bacterial genus. The size of the amplicon was set at between 85 to 250 bp. Criteria used for PCR primer selection included the melting temperature (Tm), percentage of GC content, GC clamp, secondary structure, and the tendency to form primer-dimers. Primers were analysed and optimized using the oligonucleotide sequence calculator, OligoEvaluator™ (Sigma-Aldrich, Germany,

http://www.oligoevaluator.com) and OligoAnalyzer™ Tool (https://www.idtdna.com/calc/analyzer). In general, the threshold for primer selection included $\Delta G > -9$ to minimise the likelihood of self-dimers, hairpins, and heterodimers and 50 - 55 % GC content for both the Fw and Rv primers to favour specific annealing to the targeted templates (**Supplementary table 4.2.1**). Primers that did not comply with the selection criteria were rejected, apart from the primers for *Enterovibrio* for which it was not possible to meet all the criteria (**Supplementary table 4.2.1**).

To further confirm that the designed genus-specific primer pairs would anneal to the maximal number of species in a given genus, in-silico PCR simulations using TestPrime 1.0, available in the SILVA platform, was used (Klindworth et al., 2013). The primer pairs with the maximum specificity for each taxonomic group and with the best match to the selection criteria were selected and synthesized (**Supplementary table 4.2.2**, Specanalitica, Carcavelos, Portugal). On arrival, primers were resuspended in sterile, nuclease free water to prepare 100 µM stocks (**Supplementary table 4.2.2**).

4.2.3.3. Evaluation of primer specificity by PCR amplification

The performance of the genus-specific primers was initially evaluated by running conventional PCR (Bio-Rad, T100 Thermal Cycler). Genomic DNA was extracted from seabream and seabass eggs, whole larvae and rotifers (feed) using a DNeasy Blood & Tissue Kit (Qiagen, Germany) as previously described (Najafpour et al., 2021). The PCR was carried out using 2 µl (approximately 40 - 80 ng) of genomic DNA, 2.5 µl of 10X Dream Taq Green Buffer containing 20 mM MgCl2 (Thermo Scientific, Lithuania), 0.5 µl of dNTPs (10 µM stock), 1.25 µl of the Fw primer (10 µM), 1.25 µl of the Rv primer (10 µM), 0.2 µl of Dream Taq DNA polymerase (5 U/µl, Thermo Scientific, Lithuania), and 17.3 µl of sterile, nuclease free water to give a final reaction volume of 25 µl. The PCR thermocycle consisted of 1 cycle of 95°C for 3 min followed by 34 cycles of 95°C for 10 s, a gradient of melting temperatures tested for each primer pair (57 - 64°C) for 10 s and 72°C for 10 s and a final cycle at 72 °C for 5 min. The genus-specific primers were tested in PCR amplifications of genomic DNA extracted from larval intestine, whole larvae, rotifers, and egg samples of seabream and seabass that had a high relative abundance of a given bacterial genus in 16S rRNA metagenomics.

The 16S rRNA gene PCR reaction products were run on a 2% agarose gel in TAE buffer and the specific amplicons generated by each primer pair were purified using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, UK) following the

manufacturer's instructions. The purified PCR products were ligated into the pGEM-T Easy cloning vector (Promega, Madison, USA) which permits colony selection based on white/blue colour and resistance to the antibiotic ampicillin. The ligation reactions were performed overnight at 4 °C in a 10 µl reaction mix containing 5 ul 2X Rapid Ligation Buffer, 50 ng/µl pGEM®-T Easy Vector, 1.5 U of T4 DNA Ligase (Promega, Madrid, Spain), the optimal concentration of each of the purified PCR amplicons (between 19.8 - 80.5 ng) and sterile water, to give a final reaction volume of 10 µl. DH5α competent cells (*E. coli*) were transformed with 5 µl of the ligation reaction, plated on LB agar containing 75 µg/ml ampicillin, 0.5 mM IPTG and 80 µg/ml X-Gal, and incubated at 37 °C overnight. Minipreps of plasmid DNA were prepared from white colonies and at least two colonies per genus were sequenced using the Sanger method.

The primer specificity and performance were evaluated by assigning taxonomy to each amplicon sequence using the SILVA 16S rRNA gene database as reference and phylogenetic tree generation, based on the SILVA workflow (De novo including neighbours). For the phylogenetic tree the sequences of the selected colonies were aligned using the Randomized Axelerated Maximum Likelihood (RAxML) tool with a GTR model and Gamma rate model for likelihoods (Stamatakis, 2014).

4.2.3.4. Real time qPCR optimization and primer efficiency

After confirming primer specificity by amplicon sequencing, quantitative PCR (qPCR) reactions were optimized using a Bio-Rad CFX96 qPCR Instrument (Bio-Rad Laboratories, Hercules, CA). The primer performance was initially evaluated using a gradient of melting temperatures (Tm) in a reaction volume of 10 µl containing 200 nM of each primer, 2 µl of DNA (80 ng DNA/2 µl) for the samples, or 2 µl of serial dilutions (corresponding to 102 to 107 template copies in the reaction) for the standard curve, 5 µl of 2X Forget-Me-Not™ EvaGreen® qPCR Master Mix (Biotium) and 2.4 µl of sterile nuclease free water. Thermocycling conditions were 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s, the optimized melting temperature for each primer pair for 10 s (between 58 - 61 °C) and 72°C for 10 s. A final melting curve was generated by increasing the temperature up to 95 °C in increments of 0.5 ºC every 10 s to confirm single reaction products were obtained. Control reactions included substitution of genomic DNA by water to confirm the absence of contamination.

4.2.3.5. Bacterial genus quantification and correlation with 16S rRNA microbiome profiling

To compare qPCR and 16S rRNA gene abundance estimates, Spearman correlations were used together with scatter plots generated using the R package ggplot2 v 3.3.5 (Wickham, 2016). *Colwellia*, *Oleispira*, *Phaeobacter*, *Pseudomonas*, and *Psychrobium* were quantified in nine egg samples (Najafpour et al., 2021b, **Supplementary table 4.2.3**). *Massilia*, *Phaeobacter* and *Pseudomonas*, were quantified in seabream (n = 9; 3 larvae samples in the age range of 5-15 dph and 6 samples in the age range of 43-58 dph) and seabass (n = 15; 7 larvae samples in the age range of 5-7 dph and 8 samples in the age range of 42-46 dph) larvae (**Supplementary table 4.2.3**). In the case of *Pseudomonas* where the correlation between 16S rRNA gene sequencing and copy number determined by qPCR was low, further in silico analysis was done to assess genus-specific primer performance in comparison to the universal primers used for 16S rRNA metagenomics studies. In-silico PCR v 0.5.1 implemented in Ubuntu (20.04.2 LTS) and unique *Pseudomonas* 16S rRNA gene sequences from the Silva database with a minimum length of 900 bp were used.

4.2.4. Results

4.2.4.1. Primer specificity

In total, 11 bacterial genera were targeted based on their high relative abundance in fish-related samples as determined by 16S rRNA gene sequencing (**Table 4.2.1**). The species for which the 16S rRNA genes were readily amplified by each primer pair are indicated in **Supplementary table 4.2.4 and Supplementary figure 4.2.1**.

Primer specificity was confirmed by the presence of a single band in agarose gel electrophoresis (**Supplementary figure 4.2.2**) and the sequences of the PCR amplicon. In the phylogenetic tree most of the 16S rRNA amplicon sequences generated clustered with the corresponding taxa in the SILVA 16S rRNA gene database, confirming the specificity of the genus-specific primers (**Figure 4.2.2**). In the *Pseudomonas* clade, the sequence of the16S rRNA amplicon from two samples clustered with several *Pseudomonas* species and included an uncultured *Pseudomonas* sp. (SILVA ID KP398534) and *Pseudomonas azotifigens* (AB189452; **Figure 4.2.2a**). In the *Massilia* clade, the sequence of the 16S rRNA amplicons from 2 samples clustered with several *Massilia* species including *Massilia niastensis* (EU808005), and an uncultured *Massilia* sp. (AB636963). In the *Psychrobium* clade, the sequence of the 16S rRNA gene amplicons from 2 samples clustered with *Psychrobium* species,

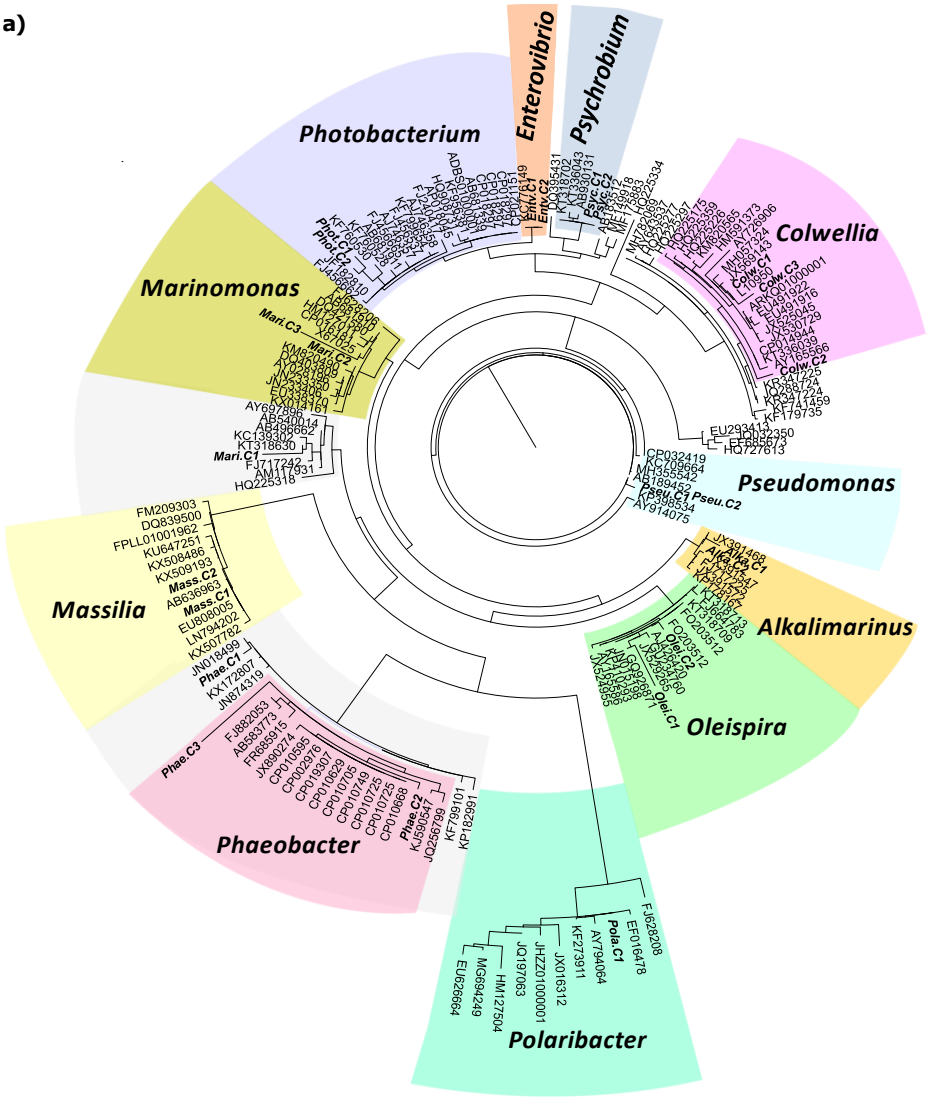
including an uncultured *Psychrobium* sp. (KT318702), *Psychrobium conchae* (AB930131; Figure 2a). In the *Phaeobacter* clade, the sequenced 16S rRNA gene amplicons clustered with *Phaeobacter* species, including *Phaeobacter inhibens* (CP010668), and *Sedimentitalea* sp. (JN018499; **Figure 4.2.2a**). In the *Colwellia* cluster, one of the 16S rRNA gene amplicons clustered with unidentified *Colwellia* spp. (L10950 and JX569143; **Figure 4.2.2a**) and the other amplicon clustered with *Thalassotalea* sp. (**Figure 4.2.2b**).

Table 4.2.1. Most abundant bacterial genera detected by 16S rRNA metagenomics across different sample types.

Bacterial genus	Maximum relative abundance of each bacterial genus in different samples (%)֍						
	Fish egg	Fish larvae	Fish intestine	rotifer	artemia	algae	water
Pseudomonas	3.8	5.2	5.4	2.1	14.8	1.3	-
Massilia	-	36.1	1.2	-	-	2.2	89.6
Psychrobium	31.3	8.6	3.7	-	-	-	33.4
Phaeobacter	-	3.4	-	-	-	-	-
Marinomonas	1.4	9	2.7	24.2	6.4	-	14.6
Polaribacter	11.8	13.6	94.7	-	-	4.7	19.2
Alkalimarinus	1.9	3.2	29.4	-	-	-	-
Enterovibrio	-	-	94.4	-	-	-	-
Photobacterium	1.5	3.2	21.4	-	-	-	40.8
Oleispira	6.6	2.2	1.7	-	0.3	-	8.3
Colwellia	9.5	4.8	1	-	-	-	20.6

֍ = Identified in the 10 top bacterial genera in each sample type, (-) signifies not detected within the most abundant bacterial genera of the specific sample type.

a)



b)

Sequence	Identity	Quality	SILVA taxa
Phae.C1	98.0	97	Rhodobacteraceae
Phae.C2	98.7	98	*Phaeobacter*
Phae.C3	73.6	57	Unclassified
Pseu.C1	99.6	99	Pseudomonadaceae
Pseu.C2	99.6	99	Pseudomonadaceae
Psyc.C1	99.2	98	Gammaproteobacteria
Psyc.C2	99.2	98	Gammaproteobacteria
Mari.C1	96.4	96	Pseudomonadales
Mari.C2	99.3	99	*Marinomonas*
Mari.C3	72.3	58	Unclassified
Pola.C1	100.0	100	*Polaribacter*
Colw.C1	100.0	100	*Colwellia*
Colw.C2	96.8	99	*Thalassotalea*
Colw.C3	96.9	93	*Colwellia*
Entv.C1	85.3	81	*Enterovibrio*
Entv.C2	84.7	81	*Enterovibrio*
Mass.C1	100.0	100	*Massilia*
Mass.C2	100.0	100	*Massilia*
Olei.C1	98.7	95	*Oleispira*
Olei.C2	99.4	98	*Oleispira*
Phot.C1	97.6	96	*Photobacterium*
Phot.C2	98.8	98	*Photobacterium*
Alka.C1	98.2	97	*Alkalimarinus*
Alka.C2	99.4	99	Pseudomonadales

Figure 4.2.2. Analysis of the specificity of the 11 designed genus-specific primer pairs by

building a phylogenic tree using the 16S rRNA gene amplicon sequences and sequences retrieved from the SILVA database. (a) Phylogenetic clustering of the qPCR amplified and sequenced 16S rRNA genes, Phao.C1, Phao.C2, Phao.C3, Pseu.C1, Pseu.C2, Psyc.C1, Psyc.C2, Mari.C1, Mari.C2, Mari.C3, Pola.C1, Colw.C1, Colw.C2, Colw.C3, Entv.C1, Entv.C2, Mass.C1, Mass.C2, Olei.C1, Olei.C2, Phot.C1, Phot.C2, Alka.C1, Alka.C2 (NCBI accession number: OM685062-OM685080) and the 16S rRNA gene sequences retrieved from the SILVA database. The phylogeny was performed using the randomized axelerated maximum likelihood (RAxML) tool with a GTR model and Gamma rate model for likelihoods (Stamatakis, 2014). (b) Least-common-ancestor (LCA) sequence classification, which classified each amplicon based on the last level of classified taxa using the available sequences in the SILVA database. The alignment quality score and identity (%) for each amplicon alignment using the LCA method are presented. The number of neighbors per query sequence was set at 10. GTR, generalised time reversible.

4.2.4.2. Primer efficiency and correlation between qPCR and 16S RNA gene abundance

The efficiency of bacterial genus-specific primers was evaluated using qPCR with the optimized annealing temperature of each primer pair (**Table 4.2.2**). All the genus-specific primers had an acceptable efficiency within the range 92 - 105.5 %.

The five most abundant bacterial genera in seabream and seabass eggs were quantified. In general, the relative abundance profiles of the 16S rRNA gene sequencing and qPCR amplification of different bacterial genera were matched with some exceptions (**Figure 4.2.3**).

Table 4.2.2. List of bacterial genus-specific primers designed and optimized for amplification of 16S rRNA genes in fish.

Bacterial genus	Primer		Size (bp)	T_m (°C)	Eff. (%)	R^2
Pseudomonas	Pseu-F	ACCGCATACGTCCTACGG	250	61	99.9	0.99
	Pseu-R	CGAAGACCTTCTTCACACACG				
Massilia	Mass-F	GCGTAGAGATGTGGAGGAAC	142	61	96.4	1.0
	Mass-R	RACCCRACAACTAGTAGACATCG				
Psychrobium	Psyc-F	GGAGGAAACTCTGATGCAGC	128	61	97.7	0.99
	Psyc-R	GTCCTTCTTCTGCGAGTAACG				
Phaeobacter	Phae-F	CACGTAGGCGGATCAGAAAG	149	61	97.6	0.99
	Phae-R	GCCACTGGTGTTCCTCCG				
Marinomonas	Mari-F	GAAGCACCGGCTAACTCTG	140	58.6	95.8	0.99
	Mari-R	GTGCMATTCCAAGGTTGAG				
Polaribacter	Pola-F	CTGGTTGACTTGAGTCATATGG	85	58	98.1	1.0
	Pola-R	CGCAATCGGTATTCTGTG				
Alkalimarinus	Alka-F	CGTAGGTGGTTTGTTAAGCGAG	165	61	95.9	1.0
	Alka-R	GTCCAGTAAGTCGCCTTCG				
Enterovibrio	Entv-F	GTGAGTAATGGCTGGGAACC	163	58	105.5	0.97
	Entv-R	CTTGGTGAGCCATTACCTCAC				
Photobacterium	Phot-F	TRGCCCAGGTGRGATTAG	167	58	92	1
	Phot-R	GGCTGCATCAGGGTTTCC				
Oleispira	Olei-F	CGGCTAATTTAGTGCCAG	171	58.2	95.2	1.0
	Olei-R	CCACTAACCTCTCTCGTACTC				
Colwellia	Colw-F	ATACGAGGGGTGCAAGCG	188	61	98.3	1.0
	Colw-R	GATGTTCCTTCCAATCTCTACGC				

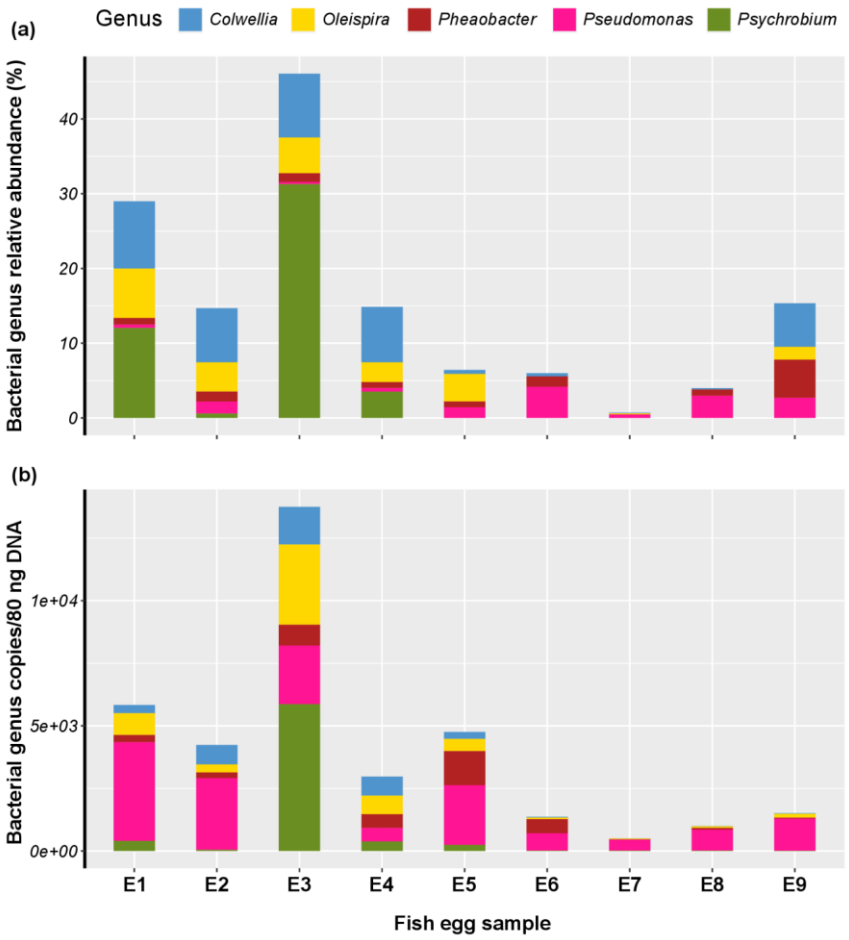


Figure 4.2.3. Analysis of the efficiency of five genus-specific primer sets by comparing the relative abundance of the genus detected in seabream and seabass eggs using 16S rRNA gene sequencing and the copy number determined by qPCR. (a) The relative abundance (%) of five bacterial genera determined by 16S rRNA gene sequencing in nine egg samples from seabream (SA, n = 6) and seabass (DL, n = 3) from three different aquaculture sites (S1, S2, and S3). (b) Quantitative analysis of five bacterial genera in nine egg samples from seabream (n = 6) and seabass (n = 3) using qPCR with the designed genus-specific primers. Full details of the experimental design, DNA extraction method, and the microbiome profile obtained using 16S rRNA gene sequencing of eggs samples are available in chapter 4.1 (Najafpour et al., 2021). The graphical plots were generated in the R package ggplot2 v 3.3.5. qPCR, quantitative polymerase chain.

A significant positive correlation was obtained for the relative abundance (%) of bacterial genera detected by both methods (**Figure 4.2.4**). In egg samples (n = 9) highly significant positive correlations were found for *Colwellia* (r = 0.82), *Oleispira* (r = 0.86), and

Psychrobium (r = 0.86, **Figure 4.2.4**). In seabream (n = 9) and seabass (n = 15) larvae, *Massilia* (r = 0.32) and *Phaeobacter* (r = 0.83) abundance also had significant positive correlations (Figure 4). For *Pseudomonas*, no correlation was found between the qPCR results for larvae (r = - 0.063, n = 24) and egg (r = -0.22, n = 9) samples and the 16S rRNA gene sequencing (**Figure 4.2.4**).

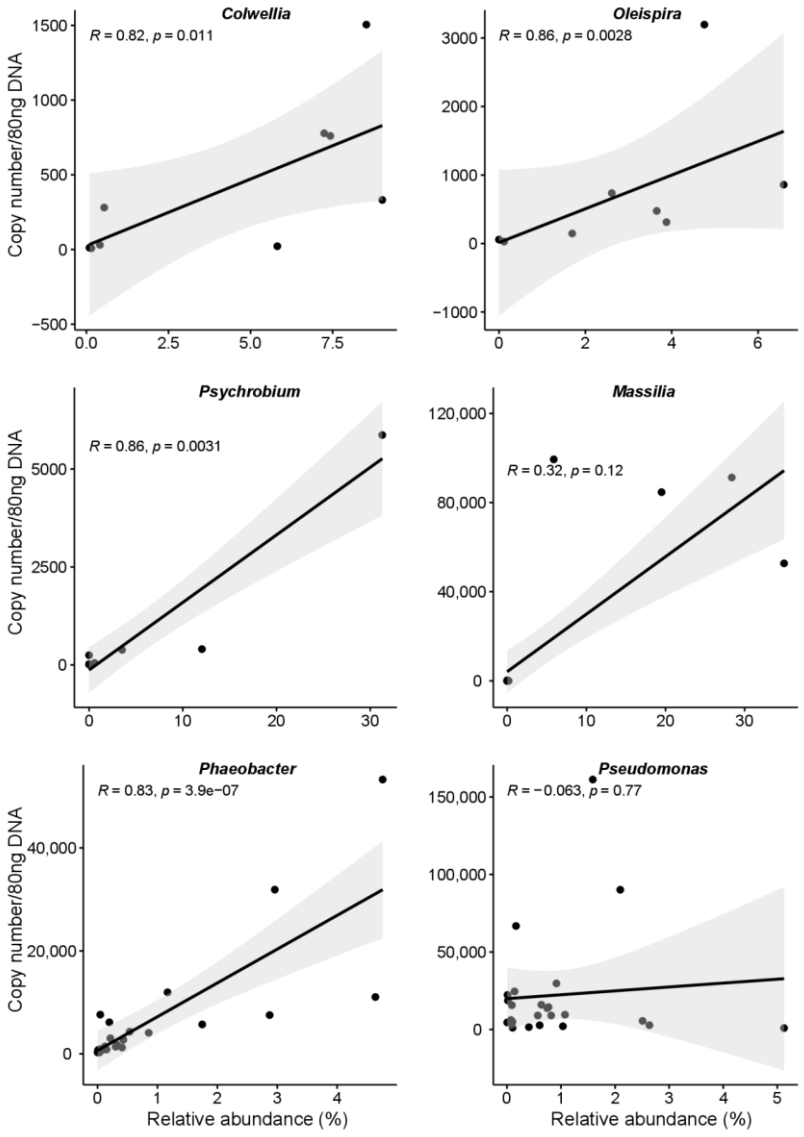


Figure 4.2.4. The correlation between genus-specific copy number established by qPCR quantification and relative abundance of six bacterial genera determined by 16S rRNA gene sequencing of seabream and seabass egg samples. The scatter plots were generated using the "Spearman" method in an R environment. Two sets of samples and databases were used for the correlation analysis. (1) For the genera, *Colwellia*, *Oleispira*, and

Psychrobium, the correlation analysis was established by comparing the relative abundance of the genera in nine egg samples of seabass (n = 3) and seabream (n = 6) obtained from 16S rRNA gene sequencing (Najafpour et al., 2021) and the copy number determined using genus-specific primers and qPCR. (2) For the genera, *Massilia*, *Phaeobacter*, and *Pseudomonas*, the correlation analysis was established by comparing the relative abundance of the genera in 24 larval samples obtained from 16S rRNA gene sequencing (in-house database) and copy number determined using genus-specific primers and qPCR. The list of samples used in the analysis is available in Table A3. qPCR, quantitative polymerase chain.

In-silico PCR analysis comparing the *Pseudomonas*-specific primers and the universal primers generally used for metagenomics (Klindworth et al., 2013) identified 27,734 and 26,780 unique sequences of the *Pseudomonas* 16S rRNA gene (total number in Silva = 60869, filtered sequences number with the minimum size 900 bp = 41607), respectively.

4.2.5. Discussion

Genus-specific primers for rapid and cost-effective high throughput monitoring of core microbial genera of seabream and seabass aquaculture were successfully developed. The potential functional importance of the abundant bacterial genera selected makes further studies about their turnover important due to their high potential impact on aquaculture production.

A number of studies have focused on *Pseudomonas* spp. due to their widespread distribution and the presence of species that are human, animal, and plant pathogens (Palleroni, 2015). For example, *P. bactica* is a pathogen of marine fish and primers targeting the gyrB gene (c390-F1 and c390-R1) and rpoD gene have been developed for rapid diagnosis (López et al., 2016). A multiplex PCR based on oprI and oprL genes was developed for the detection of Pseudomonas strains from a bacterial collection isolated from water (Matthijs et al., 2013). The qPCR comparisons, in the present study, of the abundance of five bacterial genera (*Colwellia*, *Oleispira*, *Pseudomonas*, *Psychrobium*, *Phaeobacter*) in the egg samples revealed higher than expected copy number of the Pseudomonas genus relative to the results of the 16S rRNA metagenomic sequencing. The Pseudomonas genus-specific primer pair designed in the present study were of broad scope and had the potential to anneal to 226 out of the 254 16S rRNA gene sequences represented in the LPSN and SILVA databases. Insight into why a higher than expected copy number of *Peudomonas* was detected by qPCR came from in-silico PCR comparisons of the *Pseudomonas* genus-specific and the 16S rRNA gene universal primers

(Silva database, SSU r138.1) since it revealed the universal primers (Klindworth et al., 2013) had lower efficiency and annealed to fewer of the Pseudomonas sequences represented in the database. Furthermore, the relatively short amplification product generated by the *Pseudomonas* genus-specific primers favours high qPCR efficiencies compared to previous primers designed to amplify a larger DNA sequence encompassing the ITS1 region (Locatelli et al., 2002). Bergmark et al. (2012) reported *Pseudomonas* and *Burkholderia* genus-specific primers for their detection in soil and proposed the use of genus-specific primers for both qPCR and 16S rRNA gene sequencing approaches to overcome issues related to the resolution of the 16S rRNA gene databases, although the exponential increase in available bacterial 16S rRNA gene sequences has now ameliorated this problem (Glöckner, 2019).

A high relative abundance of *Oleispira* and *Colwellia* genera was observed in our in-house database of 16S rRNA gene sequences obtained for egg and water samples from aquaculture installations. The copy number of *Oleispira* and *Colwellia* in the same samples determined by qPCR using genus-specific primers and their relative abundance using 16S rRNA gene sequencing gave a strong positive correlation. However, amplicon sequencing of products generated by the *Colwellia* primers yielded a *Colwellia*-specific amplicon and another amplicon that matched the *Thalassotalea* genera of the *Colwelliaceae* family. The similarity between the sequence of the *Colwellia* genus-specific primers and the 16S rRNA gene sequence of *Thalassotalea* (LPSN database) should be considered when using the primers. *Oleispira* and *Colwellia* are marine hydrocarbon-degrading bacteria (Mason et al., 2014) and obligate hydrocarbonoclastic bacteria (OHCB) and are usually present in the environment in very low numbers (Golyshin et al., 2010). Pollution or the addition of hydrocarbons to water induces a rapid bloom of OHCB (Kasai et al., 2002) and indicates that variations in *Oleispira* and *Colwellia* abundance may be a useful indicator of excess hydrocarbons in aquaculture systems due, for example, to addition of lipid enriched feeds.

In seabream and seabass eggs, *Pseudophaeobacter* was relatively more abundant than *Phaeobacter*, which was not detected by 16S rRNA gene sequencing in eight out of nine egg samples (Najafpour et al., 2021). In contrast, *Phaeobacter* was more common in 16S rRNA gene sequences of seabream and seabass larvae and the *Pseudophaeobacter* genus was not detected (in-house database). *Phaeobacter* spp. are common in marine organisms and the environment (Martens et al., 2006; Yoon et al., 2007; Zhang et al., 2008) and have been proposed as a probiotic against pathogenic *Vibrio* spp. in seabass and cod, *Gadus morhua* (D'Alvise et al., 2013; Grotkjær et al., 2016), although in juvenile squid they was associated

with mortality (Won, 2009). The abundance-dependent detection of *Phaeobacter* or *Pseudophaeobacter* in seabream and seabass samples, highlights one of the challenges when designing primers for short amplicon targets of 16S rRNA genes that show high similarity across bacterial genera. The amplification of *Phaeobacter* or *Pseudophaeobacter* by the genus-specific qPCR primers is unsurprising considering the recent reclassification of the species in *Leisingera-Phaeobacter* to *Sedimentitalea* and *Pseudophaeobacter* (Breider et al., 2014). Therefore, it is proposed that the *Phaeobacter* genus-specific primers designed in our study be designated group-specific primers.

The primers for *Psychrobium* were highly specific and this genus was present in all analysed samples of eggs, larvae and environmental water and a positive correlation existed between the results of qPCR and 16S rRNA gene sequencing. The lowest number of species were detected for the *Psychrobium* genus compared to the other targeted genera and it was easier to design primers specific for this genus. The high specificity of the *Psychrobium* primers was assigned to the high variability of the *Psychrobium* 16S rRNA gene compared to other bacteria.

Six of the bacterial genera, *Massilia*, *Marinomonas*, *Polaribacter*, *Alkalimarinus*, *Enterovibrio*, and *Photobacterium*, for which genus-specific primers were designed, had low relative abundance in the 16S rRNA metagenome of eggs samples (Najafpour et al., 2021). Nonetheless, the identified genera were well represented in seabream and seabass larvae, rotifers, and tank water and therefore genus-specific primers were validated in qPCR. The designed genus-specific primers revealed a high relative abundance of the *Massilia* genus in seabream and seabass larvae, *Marinomonas* in rotifer, and *Alkalimarinus*, *Polaribacter*, Photobacterium, and *Enterovibrio* in larval seabream intestine. These genera have previously been reported in a wide diversity of samples and experiments. For example, *Massilia* spp. are widespread in soil (Zhang et al., 2006), drinking water (Gallego et al., 2006) and plants (Ofek et al., 2012) and was a dominant genus in the gastrointestinal microbiota of the herbivorous grass carp, *Ctenopharyngodon idellus* (Li et al., 2014). *Marinomonas* species are abundant in seawater (Ivanova et al., 2005; Yoon et al., 2005) and were highly abundant in the microbial community of rotifer cultures prepared as a feed for fish larvae (Rombaut et al., 2001). *Polaribacter* species have been isolated from Antarctic soil and in biofilms on stones from the North Sea (Choo et al., 2020; Kim et al., 2013) and the genus was detected in the intestine of marine organisms and algae (Hyun et al., 2014; Nedashkovskaya et al., 2013; Wei et al., 2018). *Alkalimarinus sediminis* was isolated from marine sediment in Shandong Province, China

(Zhao et al., 2015) and *Alkalimarinus* species were found in bone-eating worms (*Osedax mediterranea*) from the Mediterranean Sea (Hewitt et al., 2020). Healthy and diseased *Dentex dentex* and *Sparus aurata* (bony fishes) cultured in Spanish Mediterranean aquaculture contained *Enterovibrio coralii* strains and possibly *Enterovibrio* nigricans (Pascual et al., 2009). *Photobacterium* have been isolated from seawater, mussel, eggs of spiny lobster and fish intestine and the bioluminescence and pathogenicity of some species (e.g. *Photobacterium damselae*) has made studies of them a priority (Egerton et al., 2018; Labella et al., 2017; Osorio et al., 1999).

4.2.6. Conclusion

Genus-specific primers were developed for *Pseudomonas*, *Massilia*, *Psychrobium*, *Phaeobacter*, *Marinomonas*, *Polaribacter*, *Alkalimarinus*, *Enterovibrio*, *Photobacterium*, *Oleispira* and *Colwellia*. Most of the designed primers were highly efficient in qPCR and were genus-specific as shown by amplicon sequencing and phylogeny. Difficulty was encountered in the design of genus-specific primers for *Phaeobacter* and *Pseudophaeobacter* since the short amplicons of the 16S rRNA gene encompassed regions with 93-98% identity and so these primers were designated group-specific. The genus-specific primers designed in this study will be useful for rapid evaluation and quantification by qPCR of target bacterial genera in fish related samples and potentially other metagenomic samples since the genus targeted are ubiquitous in a diversity of environments. An interesting observation was that the 16S rRNA gene universal primers used to profile microbiomes do not cover all identified species of some bacterial genera such as Pseudomonas and this has the potential to create bias in metagenomics studies and reinforces the value of complimentary qPCR studies. Overall, the designed genus-specific primers provided a rapid and cost-effective evaluation of abundant bacterial genera in samples and can therefore contribute to understanding of the modulation of abundant microbiota in complex microbial communities with potentially high impact on host biology.

4.2.7. See ANNEX IV for Supplementary materials

Supplementary table 4.2.1. Evaluation of primer GC content, dimer and secondary structure formation, and length using the OligoAnalyzer™ Tool.

Supplementary table 4.2.2. Details of the genus-specific primers and their preparation for qPCR.

Supplementary table 4.2.3. List of samples used to correlate qPCR and 16s rRNA abundance estimates.

Supplementary table 4.2.4. The numbers of potentially amplified and unamplified bacterial species with each of the designed genus-specific primer set

Supplementary figure 4.2.1. **The alignment and identity of 16S rRNA genes and genus-specific primer sequences of different bacterial species of each target bacterial genus**. The forward (in green) and reverse (in pink) primers (5′ to 3′) highlighted at the top of the alignment are specific for each bacterial genus. The name of species that their 16S rRNA genes are not 100% identical to the primers and may not be amplified with the primers are colored with yellow. This is only available in on digital format in Annex IV because the figure is very extensive.

Supplementary figure 4.2.2. **Analysis of PCR products (5 µl) on 2% agarose gel electrophoresis after staining with GreenSafe**. The first lane (L) represents the DNA ladder and the size (base pair [bp]) of some markers is shown in red on the left-hand side. Each lane contains the amplicon generated using the genus-specific primer pairs (P) and sample type (S): P = *Massilia* (*Mass*), S = 49 days post hatch (dph) seabream larvae; P = *Phaeobacter* (*Phae*), S = 9 dph seabass larvae; P = *Pseudomonas* (*Pseu*), S = 46 dph seabass larvae; P = Psychrobium (Psyc), S = 39 dph seabream larvae; P = *Alkalimarinus* (*Alka*), S = intestine of seabream larvae, P = *Polaribacter* (*Pola*), S = intestine of seabream larvae; P = *Photobacterium* (*Phot*), S = intestine of seabream larvae; P = *Marinomonas* (*Mari*), S = rotifer; P = *Oleispira* (*Olei*), S = seabream eggs; P = *Enterovibrio* (*Entv*), S = intestine of seabream larvae; P = *Colwellia*

(*Colw*), S = seabream eggs. The size of the PCR products amplified with each primer pair is presented in red (e.g., Mass, product size 142 bp). PCR, poymerase chain reaction.

4.2.8. References

Baker, G.C., Smith, J.J., Cowan, D.A., 2003. Review and re-analysis of domain-specific 16S primers. J. Microbiol. Methods. https://doi.org/10.1016/j.mimet.2003.08.009

Bergmark, L., Poulsen, P.H.B., Al-Soud, W.A., Norman, A., Hansen, L.H., Sørensen, S.J., 2012. Assessment of the specificity of *Burkholderia* and *Pseudomonas* qPCR assays for detection of these genera in soil using 454 pyrosequencing. FEMS Microbiol. Lett. 333, 77–84. https://doi.org/10.1111/J.1574-6968.2012.02601.X

Breider, S., Scheuner, C., Schumann, P., Fiebig, A., Petersen, J., Pradella, S., Klenk, H.P., Brinkhoff, T., Göker, M., 2014. Genome-scale data suggest reclassifications in the *Leisingera-Phaeobacter* cluster including proposals for *Sedimentitalea* gen. nov. and *Pseudophaeobacter* gen. nov. Front. Microbiol. 5, 416. https://doi.org/10.3389/FMICB.2014.00416/ABSTRACT

Cao, Y., Fanning, S., Proos, S., Jordan, K., Srikumar, S., 2017. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. Front. Microbiol. 8. https://doi.org/10.3389/FMICB.2017.01829

Casa-Resino, I.D.L., Empl, M.T., Villa, S., Kolar, B., Fabrega, J., Lillicrap, A.D., Karamanlis, X.N., Carapeto-García, R., 2021. Environmental risk assessment of veterinary medicinal products intended for use in aquaculture in Europe: the need for developing a harmonised approach. Environ. Sci. Eur. 2021 331 33, 1–17. https://doi.org/10.1186/S12302-021-00509-8

Choo, S., Borchert, E., Wiese, J., Saha, M., Künzel, S., Weinberger, F., Hentschel, U., 2020. *Polaribacter septentrionalilitoris* sp. nov., isolated from the biofilm of a stone from the North Sea. Int. J. Syst. Evol. Microbiol. 70, 4305–4314. https://doi.org/10.1099/IJSEM.0.004290

D'Alvise, P.W., Lillebø, S., Wergeland, H.I., Gram, L., Bergh, Ø., 2013. Protection of cod larvae from vibriosis by *Phaeobacter* spp.: A comparison of strains and introduction times. Aquaculture 384–387, 82–86. https://doi.org/10.1016/J.AQUACULTURE.2012.12.013

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/NAR/GKH340

Egerton, S., Culloty, S., Whooley, J., Stanton, C., Ross, R.P., 2018. The gut microbiota of marine fish. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.00873

Gallego, V., Sánchez-Porro, C., García, M., Ventosa, A., 2006. *Massilia aurea* sp. nov., isolated from drinking water. Int. J. Syst. Evol. Microbiol. 56, 2449–2453. https://doi.org/10.1099/IJS.0.64389-0

Glöckner, F.O., 2019. The SILVA Database Project: An ELIXIR core data resource for high-quality ribosomal RNA sequences. undefined 3. https://doi.org/10.3897/BISS.3.36125

Golyshin, P.N., Ferrer, M., Chernikova, T.N., Golyshina, O. V., Yakimov, M.M., 2010. *Oleispira*. Handb. Hydrocarb. Lipid Microbiol. 1755–1763. https://doi.org/10.1007/978-3-540-77587-4_125

Grotkjær, T., Bentzon-Tilia, M., D'Alvise, P., Dourala, N., Nielsen, K.F., Gram, L., 2016. Isolation of TDA-producing *Phaeobacter* strains from sea bass larval rearing units and their probiotic effect against pathogenic *Vibrio* spp. in Artemia cultures. Syst. Appl.

Microbiol. 39, 180–188. https://doi.org/10.1016/J.SYAPM.2016.01.005

Hewitt, O.H., Díez-Vives, C., Taboada, S., 2020. Microbial insights from Antarctic and Mediterranean shallow-water bone-eating worms. Polar Biol. 2020 4310 43, 1605–1621. https://doi.org/10.1007/S00300-020-02731-1

Hyun, D.-W., Shin, N.-R., Kim, M.-S., Kim, P.S., Jung, M.-J., Kim, J.Y., Whon, T.W., Bae, J.-W., 2014. *Polaribacter atrinae* sp. nov., isolated from the intestine of a comb pen shell, *Atrina pectinata*. Int. J. Syst. Evol. Microbiol. 64, 1654–1661. https://doi.org/10.1099/IJS.0.060889-0

Ivanova, E.P., Onyshchenko, O.M., Christen, R., Lysenko, A.M., Zhukova, N. V., Shevchenko, L.S., Kiprianova, E.A., 2005. *Marinomonas pontica* sp. nov., isolated from the Black Sea. Int. J. Syst. Evol. Microbiol. 55, 275–279. https://doi.org/10.1099/IJS.0.63326-0

Janda, J.M., Abbott, S.L., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol. https://doi.org/10.1128/JCM.01228-07

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLoS One 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Kasai, Y., Kishira, H., Sasaki, T., Syutsubo, K., Watanabe, K., Harayama, S., 2002. Predominant growth of *Alcanivorax* strains in oil-contaminated and nutrient-supplemented sea water. Environ. Microbiol. 4, 141–147. https://doi.org/10.1046/J.1462-2920.2002.00275.X

Kim, B.C., Oh, H.W., Kim, H., Park, D.S., Hong, S.G., Lee, H.K., Bae, K.S., 2013. *Polaribacter sejongensis* sp. nov., isolated from Antarctic soil, and emended descriptions of the genus *Polaribacter*, *Polaribacter butkevichii* and *Polaribacter irgensii*. Int. J. Syst. Evol. Microbiol. 63, 4000–4005. https://doi.org/10.1099/IJS.0.047100-0

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1. https://doi.org/10.1093/nar/gks808

Labella, A., Arahal, D., Castro, D., Lemos, M., Borrego, J., 2017. Revisiting the genus *Photobacterium*: taxonomy, ecology and pathogenesis. Int. Microbiol. 20, 1–10. https://doi.org/10.2436/20.1501.01.280

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Li, J., Ni, J., Li, J., Wang, C., Li, X., Wu, S., Zhang, T., Yu, Y., Yan, Q., 2014. Comparative study on gastrointestinal microbiota of eight fish species with different feeding habits. J. Appl. Microbiol. 117, 1750–1760. https://doi.org/10.1111/jam.12663

Locatelli, L., Tarnawski, S., Hamelin, J., Rossi, P., Aragno, M., Fromin, N., 2002. Specific PCR amplification for the genus *Pseudomonas* targeting the 3' half of 16S rDNA and the whole 16S-23S rDNA spacer. Syst. Appl. Microbiol. 25, 220–227. https://doi.org/10.1078/0723-2020-00110

López, J.R., Lorenzo, L., Marcelino-Pozuelo, C., Marin-Arjona, M.C., Navas, J.I., 2016.

Pseudomonas baetica : pathogenicity for marine fish and development of protocols for rapid diagnosis. FEMS Microbiol. Lett. 364, fnw286. https://doi.org/10.1093/femsle/fnw286

Martens, T., Heidorn, R., Pukall, M., Simon, B., Tindall T., Brinkhoff, 2006. Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1999 as Marinovu. Int. J. Syst. Evol. Microbiol. 56, 1293–1304. https://doi.org/10.1099/IJS.0.63724-0

Martínez-Porchas, M., Vargas-Albores, F., 2017. Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. Rev. Aquac. 9, 42–56. https://doi.org/10.1111/RAQ.12102

Mason, O.U., Han, J., Woyke, T., Jansson, J.K., 2014. Single-cell genomics reveals features of a Colwellia species that was dominant during the Deepwater Horizon oil spill. Front. Microbiol. 5. https://doi.org/10.3389/FMICB.2014.00332

Matthijs, S., Coorevits, A., Gebrekidan, T.T., Tricot, C., Wauven, C. Vander, Pirnay, J.P., De Vos, P., Cornelis, P., 2013. Evaluation of oprI and oprL genes as molecular markers for the genus *Pseudomonas* and their use in studying the biodiversity of a small Belgian River. Res. Microbiol. 164, 254–261. https://doi.org/10.1016/J.RESMIC.2012.12.001

Najafpour, B., Pinto, P.I.S., Moutou, K.A., Canario, A.V.M., Power, D.M., 2021. Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream. Microorg. 9, 2275. https://doi.org/10.3390/MICROORGANISMS9112275

Nedashkovskaya, O.I., Kukhlevskiy, A.D., Zhukova, N. V., 2013. *Polaribacter reichenbachii* sp. nov.: A new marine bacterium associated with the green alga ulva fenestrata. Curr. Microbiol. 66, 16–21. https://doi.org/10.1007/S00284-012-0200-X

Ofek, M., Hadar, Y., Minz, D., 2012. Ecology of Root Colonizing *Massilia* (Oxalobacteraceae). PLoS One 7, e40117. https://doi.org/10.1371/JOURNAL.PONE.0040117

Osorio, C.R., Collins, M.D., Toranzo, A.E., Barja, J.L., Romalde, J.L., 1999. 16S rRNA gene sequence analysis of *Photobacterium damselae* and nested pcr method for rapid detection of the causative agent of fish pasteurellosis. Appl. Environ. Microbiol. 65, 2942.

Palleroni, N.J., 2015. *Pseudomonas*. Bergey's Man. Syst. Archaea Bact. 1–1. https://doi.org/10.1002/9781118960608.GBM01210

Parte, A.C., Carbasse, J.S., Meier-Kolthoff, J.P., Reimer, L.C., Göker, M., 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. Int. J. Syst. Evol. Microbiol. 70, 5607–5612. https://doi.org/10.1099/IJSEM.0.004332

Pascual, J., Macián, M.C., Arahal, D.R., Garay, E., Pujalte, M.J., 2009. Description of *Enterovibrio nigricans* sp. nov., reclassification of *Vibrio calviensis* as *Enterovibrio calviensis* comb. nov. and emended description of the genus *Enterovibrio* Thompson et al. 2002. Int. J. Syst. Evol. Microbiol. 59, 698–704. https://doi.org/10.1099/IJS.0.001990-0

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590.

https://doi.org/10.1093/NAR/GKS1219

Rombaut, G., Suantika, G., Boon, N., Maertens, S., Dhert, P., Top, E., Sorgeloos, P., Verstraete, W., 2001. Monitoring of the evolving diversity of the microbial community present in rotifer cultures. Aquaculture 198, 237–252. https://doi.org/10.1016/S0044-8486(01)00594-4

Sharpton, T.J., 2018. Role of the gut microbiome in vertebrate evolution. mSystems 3, e00174-17. https://doi.org/10.1128/msystems.00174-17

Simon, C., Daniel, R., 2011. Metagenomic analyses: Past and future trends. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.02345-10

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Wang, Y., Qian, P.Y., 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16s ribosomal DNA amplicons in metagenomic studies. PLoS One 4, e7401. https://doi.org/10.1371/journal.pone.0007401

Wei, N., Wang, C., Xiao, S., Huang, W., Lin, M., Yan, Q., Ma, Y., 2018. Intestinal microbiota in large yellow croaker, *Larimichthys crocea*, at different ages. J. World Aquac. Soc. 49, 256–267. https://doi.org/10.1111/JWAS.12463

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Won, B.W., 2009. Characterizing the role of *Phaeobacter* in the mortality of the squid, *Euprymna scolopes*. Honor. Sch. Theses.

Yoon, J.H., Kang, S.J., Lee, S.Y., Oh, T.K., 2007. *Phaeobacter daeponensis* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. Int. J. Syst. Evol. Microbiol. 57, 856–861. https://doi.org/10.1099/IJS.0.64779-0

Yoon, J.H., Kang, S.J., Oh, T.K., 2005. *Marinomonas dokdonensis* sp. nov., isolated from sea water. Int. J. Syst. Evol. Microbiol. 55, 2303–2307. https://doi.org/10.1099/IJS.0.63830-0

Zhang, D., Zhang, Q., Li, A., 2014. Development of a multiplex PCR assay for rapid and simultaneous detection of four genera of fish pathogenic bacteria. Lett. Appl. Microbiol. 59, 471–478. https://doi.org/10.1111/LAM.12303

Zhang, D.C., Li, H.R., Xin, Y.H., Liu, H.C., Chi, Z.M., Zhou, P.J., Yu, Y., 2008. *Phaeobacter arcticus* sp. nov., a psychrophilic bacterium isolated from the Arctic. Int. J. Syst. Evol. Microbiol. 58, 1384–1387. https://doi.org/10.1099/IJS.0.65708-0

Zhang, Y.Q., Li, W.J., Zhang, K.Y., Tian, X.P., Jiang, Y., Xu, L.H., Jiang, C.L., Lai, R., 2006. *Massilia dura* sp. nov., *Massilia albidiflava* sp. nov., Massilia plicata sp. nov. and *Massilia lutea* sp. nov., isolated from soils in China. Int. J. Syst. Evol. Microbiol. 56, 459–463. https://doi.org/10.1099/IJS.0.64083-0

Zhou, L., Li, H., Zhang, Y., Han, S., Xu, H., 2014. Development of genus-specific primers for better understanding the diversity and population structure of *Sphingomonas* in soils. J. Basic Microbiol. 54, 880–888. https://doi.org/10.1002/jobm.201200679

CHAPTER 4.3

Core microbiome profiles and their modification by environmental, biological, and rearing factors in aquaculture hatcheries

Core metagenomic profiles and their modification by environmental, biological, and rearing factors in aquaculture hatcheries

Manuscript published in Marine Pollution Bulletin, 193 (2023).

https://doi.org/10.1016/j.marpolbul.2023.115218.

Acknowledgments: The authors acknowledge all the aquaculture hatchery managers and technicians for supplying the samples and Soraia Santos for assisting in the laboratory work. This work and the PhD grant to BN were funded by the European Union through project H2020-SFS2016-2-727610 "PerformFISH: Consumer-driven production: Integrating Innovative Approaches for Competitive and Sustainable Performance across the Mediterranean Aquaculture Value Chain". This output reflects only the authors' views, and the European Union cannot be held responsible for any use that may be made of the information contained herein.

Core metagenomic profiles and their modification by environmental, biological, and rearing factors in aquaculture hatcheries

Babak Najafpour[1], Patricia I.S. Pinto[1], Eric Climent Sanz[2], Juan F. Martinez-Blanch[2], Adelino V. M. Canario[1,4], Katerina A. Moutou[3], and Deborah M. Power[1,4#]

[1]CCMAR - Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, Edifício 7, 8005-139 Faro, Portugal;

[2]ADM Biopolis, Parc Cientific Universidad De Valencia, Paterna, Spain.

[3]Department of Biochemistry & Biotechnology, University of Thessaly, Ploutonos 26, Larissa, Greece.

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

CReDit statement:

Babak Najafpour: Methodology - Lead, Formal analysis, Validation, Investigation - Lead, Bioinformatics – Lead, Visualization, Writing – Original Draft, Review & Editing. Patricia I.S. Pinto: Formal analysis, Validation, Investigation, Review & Editing. Eric Climent Sanz: Formal Analysis, Validation, Review & Editing. Juan F. Martinez-Blanch: Formal analysis, Validation, Review & Editing. Adelino V.M. Canario: Resources, Writing – Review & Editing, Supervision, Project Administration. Katerina A. Moutou: Investigation, Resources, Review & editing, Funding Acquisition. Deborah M. Power: Conceptualization, Methodology, Investigation, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition.

4.3.1. Abstract

Aquaculture farms are susceptible to pathogen outbreaks that lead to imbalanced microbiome profiles. Adverse effects on human health and changes to the natural environment's trophic status are possible consequences of the dominance of pathogens. In contrast, a diverse and healthy microbiota can reduce the risk of disease outbreaks. Here we used 16S rRNA gene sequencing to profile the microbial communities and their associated functions in water, live feed (microalgae, *Artemia*, and rotifer), and different stages of European sea bass and gilthead sea bream larvae from hatcheries in Greece and Italy. The microbiome of water, live feed, and fish larvae differed, but they shared the same core bacteria, which included *Vibrio* and *Pseudoalteromonas*. Rearing water and live feeds had a different contribute to the larval microbiota: *Vibrio* transfer was more likely with *Artemia* and rotifer than with algae and water. Season (January and May) had a minor effect on the microbial composition in hatcheries. Younger (up to 23 days post-hatching, dph) and older larvae (42-77 dph) had different microbial communities, and the bacterial load increased with age. Host selection of microbiota was evident at start-feeding compared to mid-metamorphosis, and early larvae were colonized by some bacteria e.g. *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, and *Blautia*, that may be beneficial for development. The functional analysis predicted significant interactions between the host and microbiota during larval development linked to the growth and resistance towards the host of bacteria or the response to changes in host physiology, and morphology at the mid-metamorphosis stages (older larvae). The presence of pathogenic bacteria in the core microbiota, the enrichment of bacterial resistance and biofilm formation, and the overall low beneficial bacteria load during larval ontogeny emphasizes the risks of disease outbreaks. A better understanding of the microbiota in early life stages can contribute new strategies for management, optimization of environmental conditions and boost fish health and hatchery sustainability.

Keywords: Common bacteria, hatchery, larvae, microbiota, ontogeny, season, *Vibrio*

4.3.2. Introduction

Aquaculture is an industry of increasing importance for food security and the blue economy across the world. The European sea bass (*Dicentrarchus labrax*) and the gilthead sea bream (*Sparus aurata*) are the main marine aquaculture fish species in the Mediterranean. Greece, Turkey, Spain, and Italy are major producers of these two species under intensive cage culture (FAO, 2018). Despite improvements in intensive culture conditions, the survival rate of sea bass and sea bream during production is still variable, and bacterial infections cause production losses ranging from 15 to 40 % (Lane et al., 2014). The genus *Vibrio* is at the top of the list of known pathogens in mariculture, and pathogenic *Vibrio* and *Photobacterium* species have been associated with mass mortalities of sea bream and sea bass (Abdel-Aziz et al., 2013; Kahla-Nakbi et al., 2006; Snoussi et al., 2008). High interactions between fish and microbiota are proposed since fish live immersed in water and are influenced by multiple environmental factors in aquaculture sites (Austin, 2006; Ringø et al., 2010). An example, is the effect of geographical location and species on the microbial community of European sea bass and gilthead sea bream eggs (Najafpour et al., 2021a). Larval production is a critical period and a bottleneck in fish hatcheries because of the high vulnerability of larvae to environmental stressors. However, there is a growing recognition of the importance of the fish microbiome, particularly in early developmental stages, because initial bacterial colonization in organs such as the gut can potentially affect microbiota establishment at later stages and have long-term effects on host development (Deng et al., 2022; Du et al., 2021).

The microorganisms that colonize animals and plants can have both positive and negative effects on their growth and health (Vandenkoornhuyse et al., 2015; Yukgehnaish et al., 2020). Fish like other organisms (e.g., human) host both pathogenic and non-pathogenic bacteria and an imbalance (dysbiosis) can shift the microbiota composition towards pathogenic species and facilitate disease outbreaks (Brugman et al., 2018; Wynne et al., 2020). The global characterization of fish microbiota can provide information about the relative abundance of pathogenic and non-pathogenic bacteria and their interaction with the host, contributing to the development of measures for the prevention and control of disease (Egerton et al., 2018). This explains the interest in probiotics that can modulate the microbiota and in this way positively influence the immune system, growth and digestion (Belkaid and Hand, 2014; Tanaka and Nakayama, 2017; Zorriehzahra et al., 2016). For example, in sea bass and sea bream administration of bacterial probiotics (e.g., *Lactobacillus*) via the diet had a positive effect on growth and the immune system (Abelli et al., 2009; Carnevali et al., 2006; Cordero et al., 2015; Suzer et al., 2008). Thus a particular focus has been placed on the gut microbiome due to its

importance in nutritional provisioning, metabolic homeostasis and immune defence (Egerton et al., 2018; Gómez and Balcázar, 2007; Sabree et al., 2009; Sullam et al., 2012).

In recent years, understanding of the composition, complexity and contribution of the microbiota to host biology (plants or animals) has been greatly facilitated by advances in next-generation sequencing (NGS) using common DNA markers like bacterial 16S ribosomal RNA (rRNA, Ghanbari et al., 2015). The advantages of 16S rRNA gene sequencing is that it is a culture-free method and gives a snapshot of microbial communities even in complex ecosystems (Petti et al., 2005; Simon and Daniel, 2011). Analysis of the core gut microbiota has been established under standard laboratory conditions for zebrafish (*Danio rerio*) and has provided insight into host-microbe interactions (Roeselers et al., 2011). The skin and gills are also identified niche of microbiota and a core component of the host mucosal barrier (Merrifield and Rodiles, 2015). During the development of fish, the water microbiota appears to influence the gut microbiota in early larval stages, and is later modified by the introduction of different food types (Dimitroglou et al., 2010; Ingerslev et al., 2014; Navarrete et al., 2013). Fish age is also proposed as a factor influencing the diversity of the gut microbiota in smolts and adult stages of farmed chinook salmon (Zhao et al., 2020). However, comprehensive studies of marine aquaculture species under routine production conditions are infrequent.

The overall aim of the present study was to understand the microbiome composition, and its establishment and development during early ontogeny and highlight mechanisms that influence bacterial load or enrich specific pathways in sea bass and sea bream under hatchery conditions. For this purpose, a comprehensive assessment of the microbiota of sea bass and sea bream larvae was carried out to identify: a) the core and diverse microbiota composition in larvae, food and water from different production sites; b) the effect of fish species, larval stage, and season on larval microbiota; c) the contribution of different foods (microalgae, rotifer and *Artemia*) and water sources (environmental water before and after UV treatment and rearing tank water) on the fish larvae microbiota in winter and spring; and d) the differentially abundant pathways with larval age (start feeding vs mid-metamorphosis), sample type (larvae vs food vs water), and food (microalgae, rotifer, *Artemia*) to identify the functional dynamics of the larval microbiota during early ontogeny and microbial-host interactions driven by food or water.

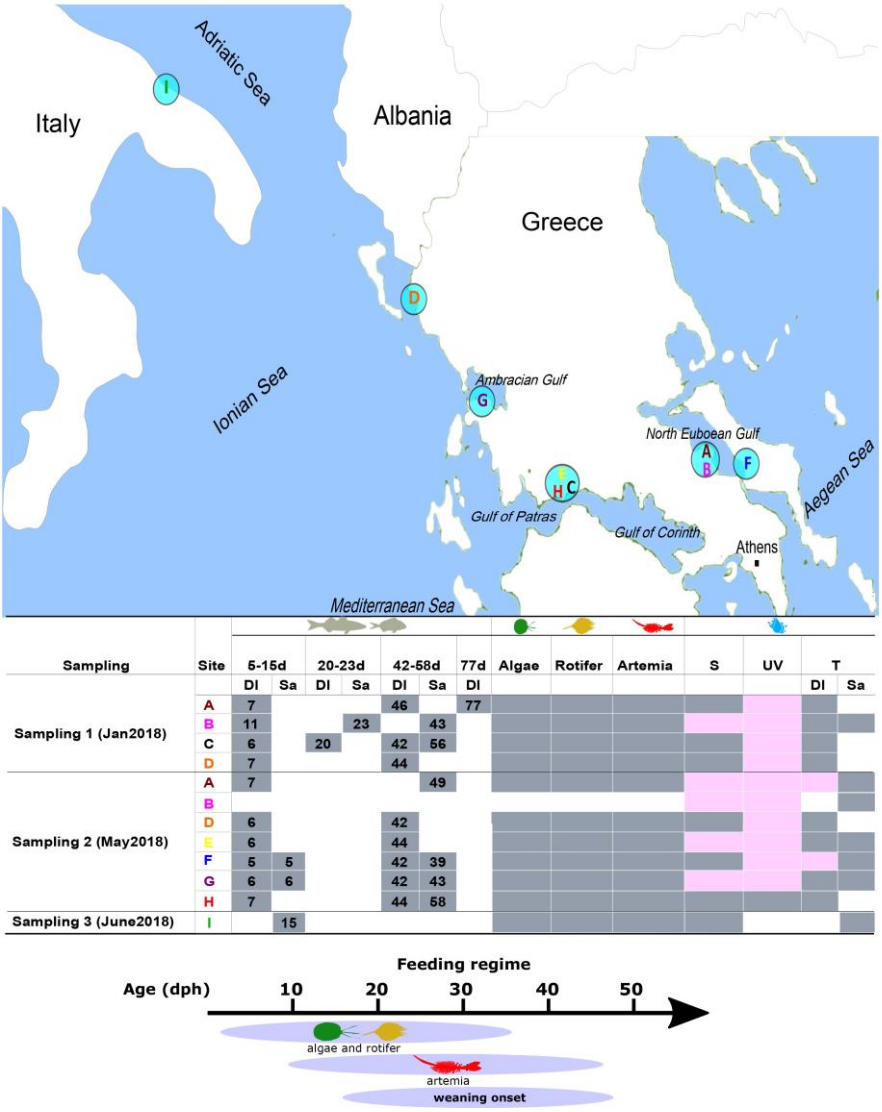
4.3.3. Material and methods

4.3.3.1. Sample collection

Samples were collected from aquaculture hatcheries in Greece and Italy, as part of their usual routine sampling procedures to verify production performance and animal welfare. The aquaculture hatcheries operated in compliance with the recommendations of the Directive 2009/58/EC of the European Parliament and of the Council of 13 July 2009 (protection of animals kept for farming). The companies held GLOBAL G.A.P. (Good Agricultural Practice) Certification, which included the demonstration of compliance with good animal welfare practices. Sampling was supervised by a qualified veterinary surgeon and was performed by hatchery technicians.

Fish larvae, food and water were sampled from eight marine hatcheries on the coast of Greece, with their approximate localization represented in **Figure 4.3.1** and detailed in **Supplementary table 4.3.1**. Samples from the Greek hatcheries were obtained in winter (January 2018) and late spring (May 2018) and designated Greece Sampling 1 (GrS1) and Greece Sampling 2 (GrS2), respectively. The first sampling exercise was carried out in four hatcheries in GrS1, designated A1 to D1, and the second sampling was performed from seven hatcheries in GrS2, named A2, B2, D2 to H2 (three hatcheries were common between GrS1 and GrS2 as indicated in **Figure 4.3.1**).

Samples were also obtained from one marine hatchery in Italy (Adriatic Sea) at the beginning of summer (July 2018) and were designated Italy Sampling (ItS). When possible, three types of water were collected from each sampling site: untreated water from the source (seawater or well) before ultra-violet (UV) disinfection; UV-disinfected water and water from the rearing tanks where the European sea bass or gilthead sea bream larval samples were collected (see **Figure 4.3.1** and **Supplementary table 4.3.1**). Approximately 400 ml of water was collected per sample into sterile 500-ml flasks containing 50 ml of DESS preserving solution (0.25M disodium EDTA pH 8.0, 20% dimethyl sulphoxide and NaCl to saturation). For the collection of the water microbiota, aliquots of 50 ml were sequentially centrifuged in the same falcon tube for 15 min at 15,500 x *g* at 4ºC, and the supernatant (water) was discarded after each centrifugation. The cumulative pellet for each sample was resuspended in 2 to 5 ml of RNA*later* (Sigma-Aldrich, Madrid, Spain) transferred to 2-ml microcentrifuge tubes and stored at -20ºC.



Sampling	Site	5-15d		20-23d		42-58d		77d	Algae	Rotifer	Artemia	S	UV	T	
		Dl	Sa	Dl	Sa	Dl	Sa	Dl						Dl	Sa
Sampling 1 (Jan2018)	A	7				46		77							
	B	11			23		43								
	C	6		20		42	56								
	D	7				44									
Sampling 2 (May2018)	A	7					49								
	B														
	D	6				42									
	E	6				44									
	F	5	5			42	39								
	G	6	6			42	43								
	H	7				44	58								
Sampling 3 (June2018)	I		15												

Figure 4.3.1. Sampling scheme presenting the samples collected from Greek hatcheries. The different hatchery sites that provided samples are presented on the map and the code allocated to each sampling site is indicated in the table using capital letters (A-I). Samples were collected at three different times of the year, January, May and June (2018) and the samples collected from the same hatcheries at the two time points are shown with the same letter and colour (A, B, D). The samples collected (larvae, food and water) at each sampling time (January and May) and at each hatchery are shaded in grey in the table; in the case of the larvae samples, the age (in days post-hatch) is specified in each cell, and the species is indicated above each column (Dl - European sea bass, *Dicentrarchus labrax* and Sa - gilthead sea bream, *Sparus aurata*). Analysed water samples from the source (S), after UV treatment (UV) and the hatchery tank water (T) for each species are presented in the table and some water samples that were extracted but did not yield enough DNA for sequencing are highlighted in light pink. The

feeding regime represents an approximate range of days post-hatch (blue ovals) that live feed fed to fish larvae across different hatcheries. See **Supplementary table 4.3.1** for detailed information on all collected samples.

Samples of microalgae (single or multiple species), rotifers and *Artemia* were collected from most hatcheries as they are the main feeds used in sea bass and sea bream hatcheries (**Figure 4.3.1** and **Supplementary table 4.3.1**). These samples were collected on the same day as the larvae from their suspension cultures into 15-ml Falcon tubes, gently centrifuged, resuspended in RNA*later* and stored at -20ºC. Sea bass and sea bream larvae, ranging in age from 5 to 77 dph (from first feeding to mid-metamorphosis) were collected directly from their tanks into 15-ml Falcon tubes containing 10 volumes of RNA*later* and stored at -20ºC (**Figure 4.3.1** and **Supplementary table 4.3.1**). All samples (water, food items and larvae) were shipped refrigerated to CCMAR (Faro, Portugal), where they were stored at -20ºC until DNA extraction.

4.3.3.2. DNA extraction

DNA was extracted from whole larvae, different food sources (microalgae, rotifer and *Artemia*) and water types (before and after UV treatment and fish tank water) using a DNeasy Blood & Tissue Kit (Qiagen, Germany) with some modifications.

For the water samples, the total volume received in RNA-later from the hatcheries (2-5 ml) was centrifuged for 10 min at 16,100 *x g* at room temperature in 2-ml sterile extraction tubes (Sarstedt, Germany). The water samples were sequentially added to the same tubes to concentrate collected material in a single tube. The collected pellet from the water was extracted in 400 µl of lysis mix (200 µl of lysis buffer 20 mM Tris-HCl, pH 8; 2 mM sodium EDTA; 1.2 % Triton X-100; 40 mg/ml lysozyme mixed with 200 µl of AL buffer from the Qiagen kit) and approx. 400 mg of 0.1 mm zirconia/silica beads (Biospec) were added. The tubes were maintained on ice until mechanical disruption, which was carried out at room temperature in a Tissue Lyser (Qiagen, Germany) for 3 cycles of 5 min at 25 Hz.

For food samples, 0.3 - 2 ml of microalgae, rotifer or artemia suspensions (20 - 300 mg wet weight) stored in RNA*later* were centrifuged for 10 min at 16,100 *x g* at room temperature. Lysis mix (400 µl) was added to the pelleted material and one iron bead (Qiagen stainless steel beads of 5 mm) per tube was used for the initial mechanical disruption carried out for 3 cycles of 30 sec at 30 Hz in a Tissue Lyser (Qiagen). After removal of the iron beads,

400 mg of 0.1 mm zirconia/silica beads were added and the disruption of the bacteria proceeded for 3 cycles of 5 min at 25 Hz.

For larvae, the amount of initial material and the disruption protocol was optimized according to the larval age. Pools of approx. 10 larvae from the first-feeding stages (6 -11 dph) weighing between 6 and 10 mg were disrupted in 400 µl of lysis mix with iron and glass beads as described above. Pools of 8 larvae were extracted for 20 - 23 dph larvae, and weighed between 10 and 14 mg. Single larvae were extracted for the metamorphosis stage (40 - 50 dph) and weighed between 9 and 35 mg. The bigger sea bass larvae of 77 dph, weighted approximately 79 mg and exceeded the proposed sample quantity of the extraction kit (< 30 mg) and so they were extracted in a double volume of lysis mix and only 1/3 of the disrupted material was used for genomic DNA extraction.

Following mechanical disruption of samples in the lysis mix they were incubated for 30 min at 37 ºC with the lysozyme (80 µl of 100 mg/ml lysozyme per sample, included in the lysis mix), followed by 30 min at 56 ºC with proteinase K (25 µl of 20 mg/ml). Tubes were centrifuged for 1 min at 4,300 *x g* to pellet the beads, and the lysate was collected into a clean microcentrifuge tube and treated with RNase (10 µl of 10 mg/ml) for 10 min at room temperature, and then 0.5 volumes of 100 % ethanol were added. The lysate mix was then purified using the column supplied in the DNeasy Blood & Tissue kit. For the food and larvae samples DNA was eluted into two tubes, the first fraction was eluted in 30 µl of Tris-HCl (10 mM, pH 8) and the second fraction was eluted in 70 µl Tris-HCl (10 mM, pH 8). In the case of the water samples the second eluate volume was 170 µl of Tris-HCl (10 mM, pH 8). DNA quality and integrity were analysed using a Nanodrop spectrophotometer and 1% agarose gel electrophoresis.

When the concentration of eluted DNA was low (< 5-10 ng/ µl), which was the case of the water samples and some of the microalgae extracts, the two fractions eluted from the column were combined and concentrated with a QIAamp DNA Micro kit (Qiagen) following the suppliers' protocol, and elution was carried out in 20 µl Tris-HCl (10 mM, pH 8).

4.3.3.3. 16S rRNA gene microbiome library preparation and sequencing

Each sequencing library was constructed using DNA from individual samples of algae, rotifers, *Artemia*, and water or triplicate samples corresponding to DNA pooled from 10 larvae at 1st feeding or from one older larva. Library preparation followed the 16S

Metagenomic Sequencing Library Preparation protocol for the Illumina MiSeq system, using optimized primers targeting the hypervariable V3 and V4 regions of the 16S rRNA gene (Klindworth et al., 2013). Libraries were paired-ended sequenced (300 bp x2) on an Illumina MiSeq platform at ADM Biopolis (previously Lifesequencing S.L.; Paterna, Valencia, Spain).

4.3.3.4. Sequence processing and bioinformatics

A pipeline developed by ADM Biopolis was used for initial processing of the sequenced reads. Through this pipeline, raw reads were cleaned and merged into paired-end reads, chimeras were filtered and eliminated, and sequences were used for classification of operational taxonomic units (OTU) as previously described (Codoñer et al., 2018), using BLASTn against the NCBI 16S rRNA database with a cut-off set at 97% identity. The software CD-HIT was used for hierarchy clustering (Fu et al., 2012).

The R Package pheatmap v. 1.0.12 was used to draw clustered heatmaps (Kolde, 2019). Nonmetric multidimensional scaling (NMDS) analysis was performed using the *metaMDS* function in the R package vegan v. 2.5-7 applying Bray–Curtis distance and data transformation (Oksanen et al., 2020). For visualizing intersecting (bacterial genera) size and the core microbiota in each sample (GrS1, GrS2 and ItS), the UpsetR package v. 1.4.0 was used (Conway et al., 2017).

To obtain standard OTU classifications for functional predictions, sequenced reads from GrS1 and GrS2 were also processed using the amplicon analysis pipeline of the SILVA project 'SILVAngs 1.4' (Quast et al., 2013). Reads shorter than 50 aligned nucleotides and reads with more than 2% ambiguities, or 2% homopolymers, respectively, were excluded from further analysis. After accomplishing initial quality control steps, the unique reads were clustered (OTUs) in each sample and the reference read of each OTU was classified using VSEARCH v. 2.14.2 (Rognes et al., 2016), applying identity criteria of 1.00 and 0.98, respectively. The classification was performed by BLASTn 2.2.30+ (Camacho et al., 2009) with standard settings using the non-redundant version of the SILVA SSU Ref dataset as the classification reference (release 138.1).

Functional annotation of the 16S rRNA metagenomics profiles were established using the web-based platform MicrobiomeAnalyst (Chong et al., 2020) and the Tax4Fun method (Aßhauer et al., 2015). To identify KEGG Orthologs (KO) with differential abundance, KOs with low counts and variance were filtered and normalized using the Cumulative Sum Scaling (CSS) method (Paulson et al., 2013). Two statistical methods were applied to perform

differential abundance analysis and minimise bias: the RNA-seq method using the edgR algorithm (Robinson et al., 2009) and the metagenomeSeq method using a 'zero-inflated Gaussian fit' model (Paulson et al., 2016). Significant KOs associated with each variable (age and sample type including larvae, water, *Artemia*, algae and rotifer) were selected for KEGG pathway analysis using the clusterProfiler R package v. 3.18.1 (Yu et al., 2012) when a confidence interval of 95% ($p < 0.05$) was confirmed by both methods.

4.3.3.5. Validation and quantitative measurement of total bacterial load and five abundant bacterial genera using qPCR analysis

Quantitative polymerase chain reaction (qPCR) was used to quantify total bacterial load and the changes in load of some abundant bacteria during larval ontogeny and to compare the results with the 16S metagenomic sequencing results. The genomic extracts of sea bass and sea bream larvae (from GrS1, GrS2 and ItS) of two age ranges (5-23 dph and 39-58 dph) were used in the qPCR reactions in triplicate. To enlarge the scope of the samples for the analysis of bacterial load and balance the number of samples of younger and older larvae, four additional samples (each corresponding to a pool of 10 larvae in triplicate) that were collected in February and March 2018 were also included in the qPCR analysis. The total bacterial load and five bacterial genera (*Vibrio*, *Massilia*, *Phaeobacter*, *Pseudomonas*, and *Psychrobium*) were quantified using genus specific 16S rRNA primers in a Bio-Rad CFX96 qPCR Instrument (Bio-Rad Laboratories, Hercules, CA, USA). Overall quantification of the 16S rRNA used the universal sense primer 16S-515fbY or 515F (Parada) 5′-GTGYCAGCMGCCGCGGTAA-3′ and antisense primer 16S-806rbN or 806R(Apprill) 5′-GGACTACNVGGGTWTCTAAT-3 (Apprill et al., 2015; Caporaso et al., 2012; Parada et al., 2016).

The qPCR amplification reactions for the 16S rRNA gene contained 30 ng of template DNA (in 2 µL), the universal primers and other reagents as previously described (Najafpour et al., 2022). *Vibrio* genus-specific primers targeting a 120-bp region of the 16S rRNA gene were used for qPCR quantification as previously described (Pinto et al., 2019; Tall et al., 2012). The genus-specific primers for *Massilia*, *Phaeobacter*, *Pseudomonas*, and *Psychrobium* have been recently developed and validated (Najafpour et al., 2022). The amount of template DNA used in qPCR was 80 ng (in 2 µL) and optimized reaction mixtures and conditions for each primer pair were used (Najafpour et al., 2022). Bacterial genera copy number was measured by creating genus-specific standard curves ranging from 10^2 to 10^7 copies of the target gene (in 10-fold increments). Bacterial copy number based on the 16S rRNA

gene was calculated as follows: copy number = $(X \times 6.022 \times 10^{23})/(Y \times 1 \times 10^{9} \times 650)$, where X is ng of the template gene, Y is the template length (bp), and 650 (Da) represented the average molecular weight of a base pair (Martyniuk et al., 2009).

4.3.3.6. Statistical analysis

The Shannon index and CHAO1 were used to estimate alpha-diversity and species richness in each sample. Data normality was checked using a Shapiro-Wilk normality test, and the data distribution for the Shannon index was normal across the samples (n = 82). To analyse CHAO1, outlier samples (n = 10, > 2 standard deviations) were eliminated to obtain a normal distribution for pairwise comparison analysis (n = 70). The analysis of variance (ANOVA) and Tukey's test for multiple pairwise comparisons were used to compare the statistical difference in bacterial richness and diversity between different samples (larvae, algae, rotifer and *Artemia*). To evaluate beta-diversity, data homogeneity was controlled using the *betadisper* function (evaluating beta-dispersion) and an ANOVA-like test applied, using the *permutest* function in R. Based on principal coordinate analysis (PcoA) of beta-diversity data, permutational analysis of variance (PERMANOVA) using the *adonis* function and Bray–Curtis distances were used to test whether the overall microbial community differed with each variable under analysis.

The R package metagenomeSeq v 1.32.0 was used to identify differentially abundant OTUs in the microbiome data, and different models based on different variables were created using the *fitZIG* function to determine differentially abundant OTUs with an adjusted p-value < 0.05 (Paulson et al., 2016). OTUs that were significantly changed in each comparison were visualized with heatmaps using the R package superheat v 1.0.0 (Barter and Yu, 2021).

4.3.4. Results

4.3.4.1. Sequencing statistics

Of the 112 samples extracted for microbiome analysis, 89 yielded adequate DNA quality for sequencing (**Figure 4.3.1**). A total of 7.5 million (M) raw read sequences were produced from the 89 libraries and generated 3.4 M paired-end reads. After quality control and trimming of low-quality sequences, 3.2 M filtered reads with an average length of 410-bp were obtained. The detailed sequencing statistics for each sample is provided in **Supplementary**

table 4.3.2.

The rarefaction curves indicated saturation or close to saturation coverage for diversity in each sample (**Supplementary figure 4.3.1-3**). Based on the number of detected OTUs, the rarefaction curves also revealed different bacteria richness (between different hatcheries, sample type or sampling time). In general, water samples had a higher richness (100-500 OTUs), with the greatest number of OTUs detected in the environmental source waters (SCJ, SDM, **Supplementary figure 4.3.1 - 4.3.3**).

4.3.4.2. Bacterial community composition

The size of the microbiome and common/core bacteria genera were analysed by counting the number of bacterial genera across all samples (**Supplementary figure 4.3.4**). In general, the samples of source water (SCJ and SDM) had the highest number of bacteria genera and *Artemia* and algae (AM, GM) had the lowest number (**Supplementary figure 4.3.4**). The distribution in all samples of the 61 bacterial genera identified in at least 55% of the samples is shown in **Supplementary figure 4.3.4**.

The most abundant phyla in the sequenced libraries were Proteobacteria (mean = 59.6%, range = 2.7 - 98.3%), followed by Bacteroidetes (mean = 12.2%, range = 0 - 82.6%) and to a lesser extent Firmicutes (mean = 3.4%, range = 0 - 44.1%) and Actinobacteria (mean = 0.8%, range = 0 - 10.4%, **Figure 4.3.2**, **Supplementary tables 4.3.3 – 4.3.5**). In general, a

higher percentage of Firmicutes was detected in 6 - 23 dph larvae (mean = 8.1%, range = 0.9 - 25.3%) and Proteobacteria (mean = 80.4%, range = 47.9 - 98.2%) in older larvae (**Supplementary tables 4.3.3 – 4.3.5**). The most abundant families were *Vibrionaceae* (mean = 21.4%, range = 0 - 96.9%), *Rhodobacteraceae* (mean = 8.7%, range = 0 - 45.7%), *Flavobacteriaceae* (8.1%, range = 0-54.8), *Pseudoalteromonadaceae* (mean = 5.9%, range = 0 - 45.8), *Oceanospirillaceae* (mean = 4.4%, range = 0 - 28.1%), *Alteromonadaceae* (mean = 4.3%, range = 0 - 39.8%), *Oxalobacteraceae* (mean = 3.4%, range = 0 - 89.6%), *Colwelliaceae* (mean = 2.7, range = 0 - 62.9%), *Shewanellaceae* (mean = 1.7%, range = 0 - 33.8%) and *Crocinitomicaceae* (mean = 1.7%, range = 0 - 80.4%, **Figure 4.3.2**, **Supplementary tables 4.3.3 – 4.3.5**).

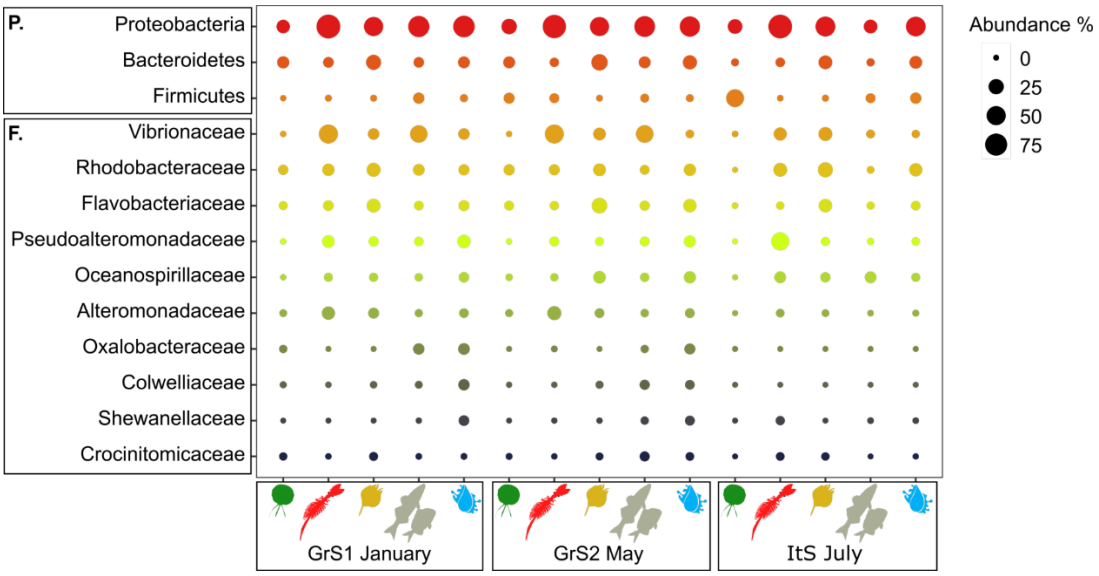


Figure 4.3.2. A bubble plot representing the most abundant phyla (P.) and families (F.) overall by sample type collected at the two sampling timepoints, January 2018 (GrS1) and May 2018 (GrS2) and June 2018 (ItS). Different sample types are represented with different symbols: European sea bass larvae (); gilthead sea bream larvae (); algae (); rotifer (); *Artemia* (); water ().

The microbiota abundance was determined up to the species level (**Supplementary tables 4.3.3 – 4.3.5**), but for higher confidence in OTU assignment, it is only discussed up to the genus level. The bacterial genera detected with an average abundance higher than 1% are listed in **Supplementary tables 4.3.3 – 4.3.5**. The clustered heatmaps of relative abundance only include the genera/OTUs with > 8% abundance in at least one of the 16S rRNA gene libraries (**Supplementary figure 4.3.5 - 4.3.7**). There was no strong clustering of samples unless the predominance of a bacterial genus led to cluster formation, e.g., *Vibrio* in fish larvae or in *Artemia* (mean = 21.5%, range = 0 - 96.8%, **Supplementary tables 4.3.3 – 4.3.5**, **Supplementary figure 4.3.5 - 4.3.7**).

The ten bacterial genera with the highest relative abundance in each of the samples are presented in **Supplementary table 4.3.6** and a graphical summary of the five most abundant genera per sample type is presented in **Figure 4.3.3**. *Vibrio* was the genera most abundant in larval samples, with only a few exceptions: *Epibacterium* was more abundant in sea bass 1BJ (29.4%), *Massilia* in sea bass 3AJ and sea bream 3CJ (36.1% and 35%, respectively) from GrS1, *Salinirepens* (80.2%) and *Thalassotalea* (58.9%) in sea bass 1FM and sea bream 1FM,

respectively, from GrS2 (**Figure 4.3.3**, **Supplementary table 4.3.6**). Some larval samples had a high percentage of "No hit" reads and retrieved no matches from the 16S rRNA gene database (e.g., sea bass 1DM, **Supplementary table 4.3.6**).

Vibrio (mean = 20.7%, range = 0.01 – 88.5%), *Alteromonas* (mean = 6.8%, range = 0.01 – 38.4%) and *Pseudoalteromonas* (mean = 3.5%, range = 0.004 – 27.8%) were the most abundant bacterial genera in the food samples from GrS1 and GrS2 (**Figure 4.3.3**, **Supplementary table 4.3.6**). In general, these genera were present at lower abundance in microalgae than in *Artemia* and rotifers (**Figure 4.3.3**, **Supplementary table 4.3.6**). High relative abundance of some genera was limited to a few food samples and food types, such as *Epibacterium* in *Artemia* AJ (35.4%), *Artemia* AM (30.6%) and *Artemia* IUJ (11.2%), *Donghicola* in rotifer AJ and AM (15%), *Kordia* in rotifer CJ (14.8%) and rotifer EM (14.6%) and *Aestuariibacter* in rotifer CJ (11.2%), *Yoonia* in rotifer BJ (21.8%) and rotifer IJU (23.3%), *Algoriphagus* algae DJ and DM (10.7% and 7.7%, **Supplementary table 4.3.6**). *Mesoflavibacter* was one of the top 10 most abundant genera in some rotifer samples and showed high abundance in the samples AJ, FM, AM, and HM (Figure 3, Suppl. Table 6). *Marinomonas* was common in food samples (mean = 3.2%, range = 0 – 24.3%) and was found in most *Artemia* and rotifer samples from GrS1, GrS2 and ItS (**Figure 4.3.3**, **Supplementary table 4.3.6**). Some clusters of unknown identity had high abundance in algae samples, and BLAST searches indicated they resulted from non-specific PCR amplification of microalgae chloroplast sequences.

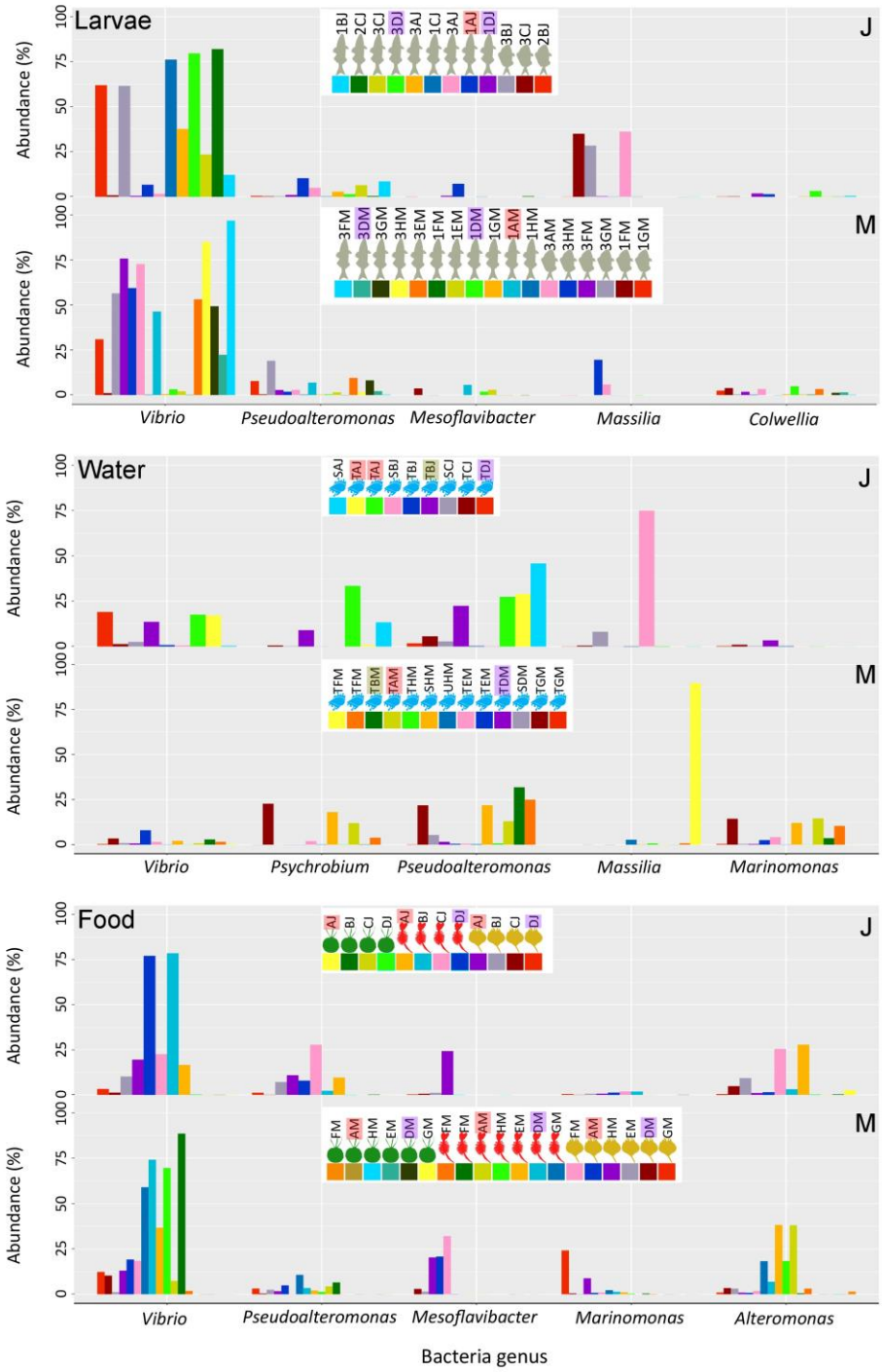


Figure 4.3.3. Selection of the five most abundant bacterial genera in each type of sample, obtained from the detailed analysis of the ten genera represented in the highest proportion in all analysed microbiomes. The percentage of each identified genus relative to the total reads per library is presented for each sample, organized by sample type (larvae, water or food) for Greece Sampling 1 in January (panel J in each sample type), and Greece Sampling 2 in May (panel M). Different sample types are represented with different symbols: European sea bass

larvae (); gilthead sea bream larvae (); algae (); rotifer (); *Artemia* (); water (). In the larvae panel, the different age ranges are represented with a code: 1 (5-11 dph), 2 (15-23 dph) and 3 (42-77 dph). Different letters (A, B, C, D, E, F, G, H, I) identify the different hatcheries and different times of sampling are shown by 'J' (January), and 'M' (May) after each site. The type of water is represented by “S” for source water “U” for water after UV treatment and “T” for water from hatchery tanks (see **Supplementary table 4.3.1** for more details). The codes of the collected samples in January and their equivalents from the same site in May are highlighted using common colours.

Vibrio was also among the most abundant bacteria in water samples, particularly in tank water (TAJ, TBJ, TDJ, TEM) - Figure 3 and Suppl. Table 6. *Pseudoalteromonas* (mean = 11.6%, range = 0.01-45.8%) was among the top 10 genera of most water samples from GrS1 and GrS2. This genus represented 22 - 45% abundance in water samples SAJ, TAJ, TBJ, TFM, and TBM. *Massilia, Fucophilus, Thalassotalea, Psychrobium, Marinomonas, Amphritea, Colwellia and Polaribacter* were also among the most abundant bacterial genera of water samples (**Figure 4.3.3**).

Comparing the relative abundance of each of the main five genera from 16S rRNA gene sequencing and their 16S rRNA gene copy number estimated by quantitative PCR in larvae samples using genus-specific primers, they were positively correlated (**Supplementary figure 4.3.8**). The total 16S rRNA gene copy numbers (bacterial loads quantified by qPCR with 16s rRNA universal primers, **Figure 4.3.4**) had a normal distribution across the larval samples (p-value = 2.33 e^{-05}) and was not significantly different between sea bass and sea bream samples ($p > 0.05$). A significant increase in 16S rRNA gene copy number with larvae age reflected the increase in the total bacterial load irrespective of species ($p = 0.009$, **Figure 4.3.4**). When genera were considered separately, species and age did not significantly affect the *Vibrio* population ($p > 0.05$) (Figure 4); *Massilia* was significantly more abundant in sea bream (age range = 39 - 58 dph) compared to sea bass (age range = 42 - 46 dph). *Phaeobacter, Pseudomonas* and *Psychrobium* abundance was not dependent on species or age (**Figure 4.3.4**).

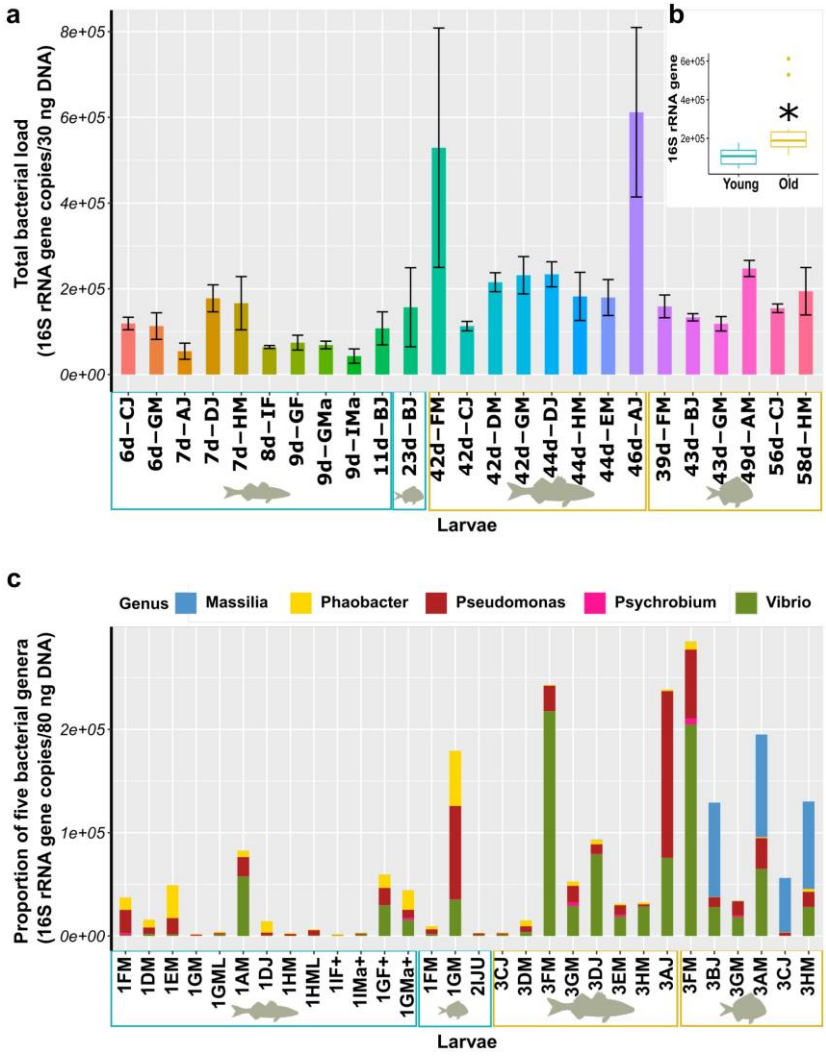


Figure 4.3.4. Results from quantitative PCR (qPCR). The graphs show the quantitative measurements of the total bacterial load and five bacterial genera during gilthead sea bream (🐟) and European sea bass (🐟) larval ontogeny using specific primers. a) The total bacterial load was measured by quantifying the 16S rRNA gene in larvae of two age ranges: younger larvae (6-23 dph) and older larvae (42-58 dph); b) The significant change identified between the total bacterial load of younger and older larvae (younger larvae were mainly related to European sea bass species); c) The relative proportion of the five bacterial genera quantified using genus-specific primers amplifying bacterial 16s rRNA genes in the different larvae samples. The age range of the European sea bass and gilthead sea bream larvae are denoted with a code: 1 (5-11 dph), 2 (15-23 dph) and 3 (39-77 dph). The letters A, B, C, D, E, F, G, H, and I identify the different hatchery sites (see figure 4.3.1 for more details). The main sampling times of the material used for 16S rRNA gene metabarcoding analysis and qPCR are shown with J (January), and M (May). A larvae sample obtained at the beginning of Summer

(June 2018) from one aquaculture site in Italy (Adriatic Sea) is shown by JU. Additional samples (+) from other sampling time points that were analysed only with the genus specific primers and 16S rRNA primers are indicated by F (related to February 2018) and Ma (related to March 2018); for more details and sample codes see **Supplementary table 4.3.1**.

4.3.4.3. Alpha diversity

There was a high variation in CHAO1 richness and Shannon diversity for equivalent sample types from GrS1 and GrS2 (**Supplementary figure 4.3.9**, **Supplementary table 4.3.7**). Multiple pairwise comparisons showed divergent bacterial richness between rotifer - algae (difference = 194.5, $p < 0.05$), water - algae (difference = 190.1, $p < 0.05$) and larvae -algae (difference = 169.7, $p < 0.05$). The CHAO1 richness in *Artemia* samples was not significantly different ($p > 0.05$) from water (difference = 152.9), larvae (difference = 132.4) or rotifer (difference = 157.3).

There was a Shannon diversity trend from low to high values in food, water and larvae in both GrS1 and GrS2 (**Supplementary figure 4.3.9**, **Supplementary table 4.3.7**). The average Shannon diversity index was around 2.85 and multiple pairwise comparisons revealed that it was only significantly different between larvae and *Artemia* (difference = 0.93, $p < 0.05$). Bacterial richness and diversity did not differ significantly between samples collected in January (GrS1) and May (GrS2, **Supplementary figure 4.3.9**, **Supplementary table 4.3.7**). In some samples with low Shannon diversity, there was a high CHAO1 richness index (e.g., sea bream 3BJ), suggesting that high bacterial richness does not necessarily lead to a balanced bacterial community (**Supplementary figure 4.3.9**, **Supplementary table 4.3.7**).

4.3.4.4. Beta diversity

Overall, a comparison of the beta diversity composition of the microbiome between samples using non-metric multidimensional scaling (NMDS) showed sample type clustering in GrS1 and GrS2 (e.g., microalgae, *Artemia*, rotifer and larvae, **Figure 4.3.5**). The effect of species/rearing practice, geographical location, and type of water sample (source water, rearing tank water and UV treatment water) caused divergence in the microbiome. Comparing the larvae samples within GrS1 (**Figure 4.3.5a**) or GrS2 (**Figure 4.3.5b**) showed that the younger larvae (ranging between 5 - 7 dph, code 1) were generally differentiated from older larvae (ranging between 39 - 77 dph, code 3). Larvae of 15 - 23 dph (code 2) tended to cluster with the younger larvae (code 1). In GrS1, sea bream larvae of 56 dph (3CJ) and sea

bass larvae of 77 dph (3AJ) were discriminated from the younger larvae (codes 1, 2). An obvious separation of the sea bream and sea bass larvae beta diversity was not observed. PERMANOVA analysis (Bray–Curtis distance) supported the outcome of the PcoA that indicated different microbial communities in different sample types (larvae, water, algae, *Artemia*, and rotifer) and larval ages, irrespective of the sampling time (GrS1 and GrS2, **Tables 4.3.1 and 4.3.2**).

Table 4.3.1. **PERMANOVA analysis (permutations = 1000) between different types of samples collected in January 2018.**

	Df	Sum Sq	Mean Sq	F. Model	R2	Pr (>F)
Algae vs *Artemia*	1	0.8713	0.87135	2.5639	0.2993	0.0309 *
Algae vs rotifer	1	0.7674	0.7674	2.1285	0.2618	0.0219 *
Algae vs larvae	1	1.0503	1.0503	3.0402	0.1784	0.0019 **
Algae vs water	1	0.9153	0.9152	2.5281	0.1868	0.0019 **
Artemia vs rotifer	1	0.5188	0.5188	1.4358	0.1930	0.0269 *
Artemia vs larvae	1	0.8214	0.8214	2.3753	0.1450	0.0009 ***
Artemia vs water	1	0.8231	0.8231	2.2709	0.1711	0.0009 ***
Rotifer vs larvae	1	0.7019	0.7019	1.9790	0.1238	0.0089 **
Rotifer vs water	1	0.7078	0.7077	1.8937	0.1468	0.0019 **
Larvae vs water	1	0.6611	0.6611	1.8515	0.0887	0.0079 **
Larvae1 vs larvae2	1	0.4533	0.4532	1.4116	0.2608	0.1333
Larvae2 vs larvae3	1	0.3920	0.3920	1.1594	0.1619	0.3037
Larvae1 vs larvae3	1	0.7800	0.7800	2.4916	0.2374	0.0069 **

Df: degrees of freedom; Sq: square; Significance code: ***p < 0.001, **p < 0.01, *p < 0.05

Table 4.3.2. **PERMANOVA analysis (permutations = 1000) between different types of samples collected in May 2018.**

	Df	Sum Sq	Mean Sq	F. Model	R2	Pr (>F)
Algae vs *Artemia*	1	0.9745	0.9745	2.5766	0.1897	0.0029 **
Algae vs rotifer	1	0.9195	0.9195	2.4826	0.1988	0.0019 **
Algae vs larvae	1	1.0488	1.0487	2.8707	0.1255	0.0009 ***
Algae vs water	1	0.9183	0.9182	2.4226	0.1247	0.0009 ***
Artemia vs rotifer	1	1.0970	1.0969	3.5733	0.2452	0.0019 **
Artemia vs larvae	1	1.0083	1.0083	2.9955	0.1198	0.0019 **
Artemia vs water	1	1.3523	1.3522	3.9794	0.1810	0.0009 ***
Rotifer vs larvae	1	1.1441	1.1441	3.4575	0.1413	0.0009 ***
Rotifer vs water	1	1.2674	1.2673	3.8064	0.1829	0.0009 ***
Larvae vs water	1	1.0928	1.0927	3.1580	0.1013	0.0009 ***
Larvae1 vs larvae3	1	1.1686	1.1685	3.7561	0.2002	0.0009 ***

Df: degrees of freedom; Sq: square; Significant code: ***p < 0.001, **p < 0.01

4.3.4.5. Differential abundance according to sample type

The beta diversity analysis revealed distinct microbial communities between sample types (larvae, water, rotifer, and algae), time of year (May and January), hatchery, larval species, and age, prompting the development of models of bacterial abundance by sample type.

Microalgae and rotifer bacterial populations were more similar between themselves compared to *Artemia*, water or larvae samples (evidenced in the heatmaps and differential abundance analyses in **Figure 4.3.6**) and this was confirmed by the clear separation between algae and rotifer microbial communities in the PcoA analyses. For example, the *Planctomycetaceae* and *Phyllobacteriaceae* families were more abundant in algae and rotifer samples than in *Artemia*, larvae and water samples ($p < 0.05$, **Figure 4.3.6a**). Lower abundance of *Vibrionaceae* and *Pseudoalteromonadaceae* and higher abundance of *Rhodospirillaceae* were also observed in algae compared to the other sample types rotifer, *Artemia*, water and larvae ($p < 0.05$, **Figure 4.3.6a**). *Erysipelotrichaceae*, *Coriobacteriaceae*, *Bifidobacteriaceae* were of significantly higher abundance ($p < 0.05$) in younger larvae (age range = 5 - 23 dph) compared to older larvae (age range = 39 - 77 dph;

Figure 4.3.6b).

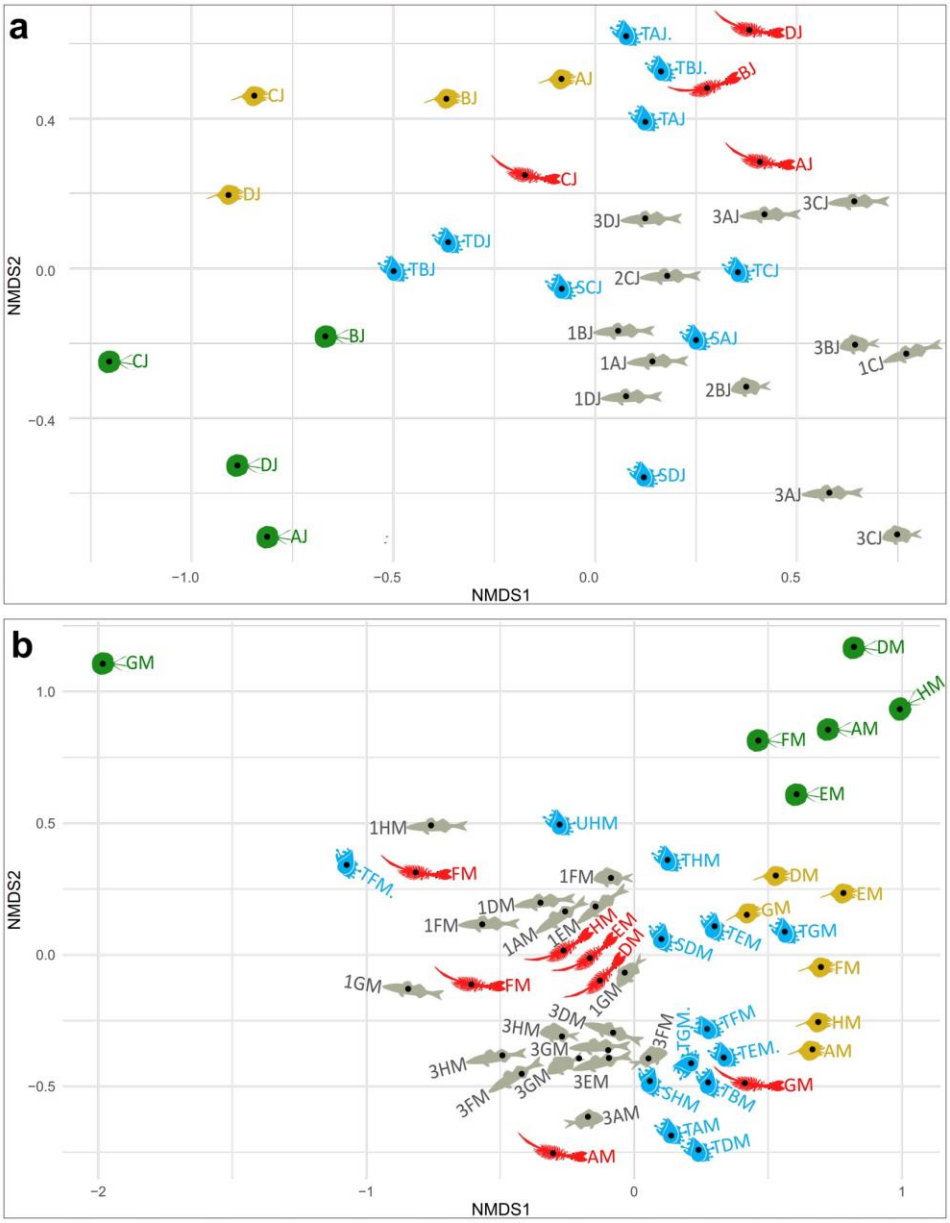


Figure 4.3.5. Nonmetric multidimensional scaling (NMDS) analysis of the samples collected in January (panel a, Greece Sampling 1) and May (b, Greece Sampling 2). Bray–Curtis distance and data transformation were applied for the analyses. Different types of samples are shown with different symbols: European sea bass larvae ( ); gilthead sea bream larvae ( ); algae ( ); rotifer ( ); *Artemia* ( ); water ( ). Samples are coded so that age, site, and time of sampling are perceptible. Larvae of different ages are indicated as follows: 1 (5-11 dph), 2 (20-23 dph) and 3 (39-77 dph). The letters A, B, C, D, E, F, G, H, and

I identify the different hatchery sites (see **Figure 4.3.1** for more details). The letters J and M represent the different sampling times, January, and May, respectively. Water from source (S), after UV treatment (U) and in hatchery tanks (T) are also specified (See also **Supplementary table 4.3.1**).

Some bacterial genera were more abundant (e.g., *Streptococcus*, *Bifidobacterium*, *Lactobacillus*, *Bacteroides* and *Blautia*) and others less abundant (e.g., *Alkalimarinus*, *Agarivorans*, *Leucothrix* and *Pseudomonas*) in younger larvae (5 - 23 dph) compared to older larvae (39 - 77 dph, **Figure 4.3.7a**). Most of the bacterial genera that were differentially distributed between larvae and water samples had a higher abundance in water (e.g., *Psychrobium*, *Marinomonas*, *Amphritea*, *Pseudofulvibacter* (**Figure 4.3.7b**). Relatively few genera had lower abundance in water (e.g., *Blautia*, *Bacteroides*, *Bifidobacterium*) than in larvae (Figure 7b). Generally, in larvae there was higher abundance of *Polaribacter*, *Colwellia*, *Blautia*, *Bacteroides*, *Alkalimarinus*, *Leucothrix* and *Psychrobium* and less abundance of *Alteromonas*, *Roseovarius*, *Winogradskyella*, *Aestuariibacter* compared to food irrespective of type (**Figure 4.3.7c**; **Supplementary figure 4.3.10**). Marker bacterial genera that separated the microbial community of different food types were identified in differential abundance analysis. There was a differential distribution of bacterial genera between the food type algae, *Artemia*, and rotifer ($p < 0.05$, **Figure 4.3.8**), and some more abundant genera in each food type were also more likely transferred to larvae (**Figures 4.3.7** and **4.3.8**, **Supplementary tables 4.3.3 – 4.3.5**).

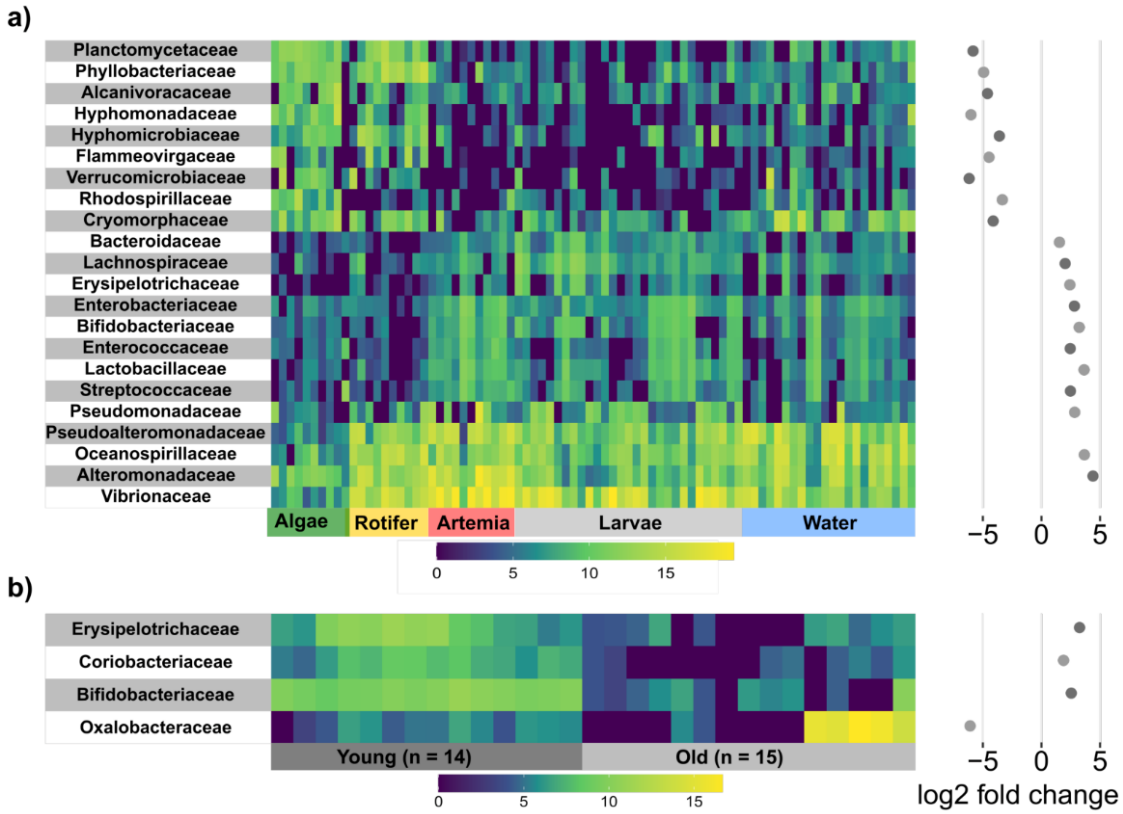


Figure 4.3.6. Differentially abundant bacterial families across different types of samples and larvae; a) the model factors included site, time of sampling (January and May) and sample type (algae, rotifer, *Artemia*, larvae); b) the factors included in the model included species (European sea bass and gilthead sea bream), site, time of sampling (January and May) and larval age range (younger = 5-23 dph and older = 39-77 dph). The R package metagenomeSeq v 1.32.0 was used to identify differentially abundant OTUs with an adjusted p-value < 0.05 (Paulson et al., 2016). Bacterial families with significant changes in each comparison are visualized with heatmaps using the R package superheat v 1.0.0 (Barter and Yu, 2021).

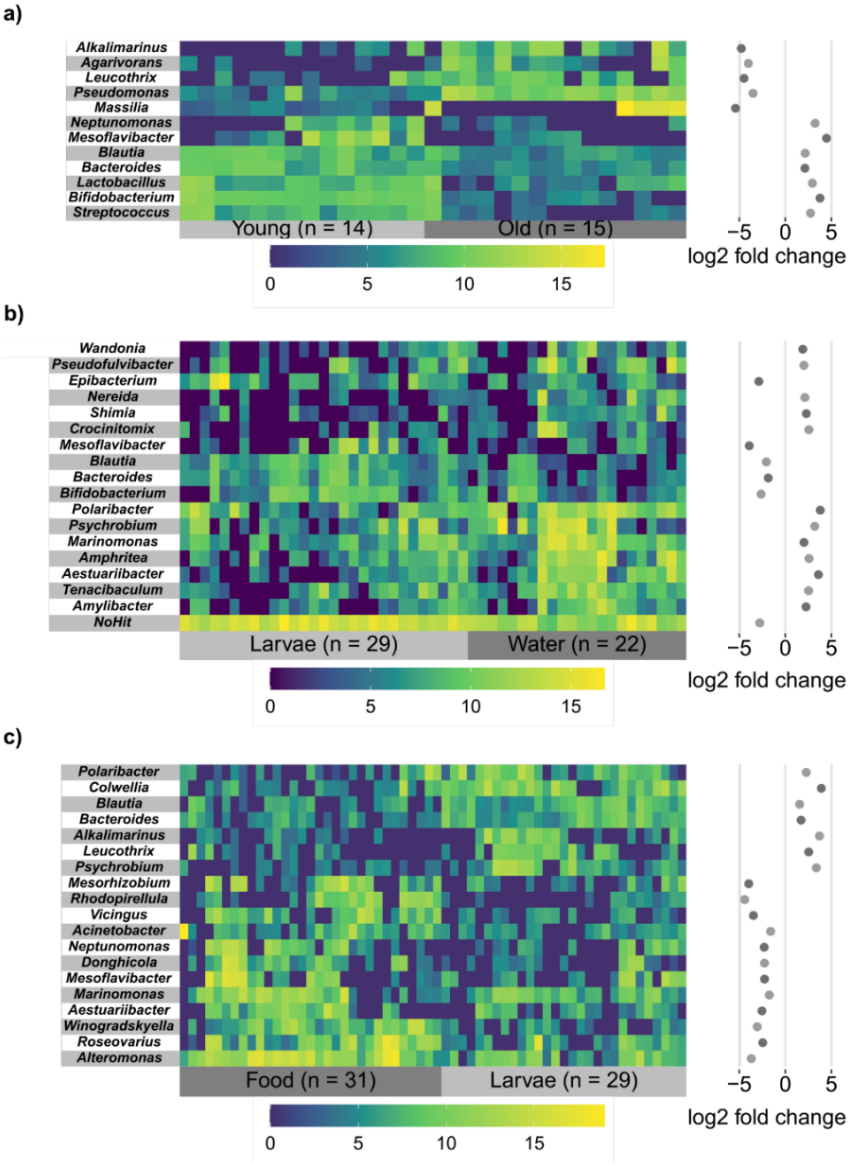


Figure 4.3.7. Differentially abundant bacterial genera comparing younger and older larvae, larvae and water, and larvae and food; a) the model factors included species (European sea bass and gilthead sea bream), site, time of sampling (January and May), and larvae age range (younger = 5-23 dph and older = 39-77 dph); b) the model factors included site, time of sampling (January and May) and sample type (larvae and water); c) the model factors included site, time of sampling (January and May) and sample type (larvae and food), the separate analysis of differentially abundant genera between larvae and each food type (algae, rotifer, and *Artemia*) is presented in suppl. Fig. 10. The R package metagenomeSeq v 1.32.0 was used to identify differentially abundant OTUs with an adjusted p-value < 0.05 (Paulson et al., 2016). Bacterial families with significant changes in each comparison are visualized in heatmaps generated with the R package superheat v 1.0.0 (Barter and Yu, 2021).

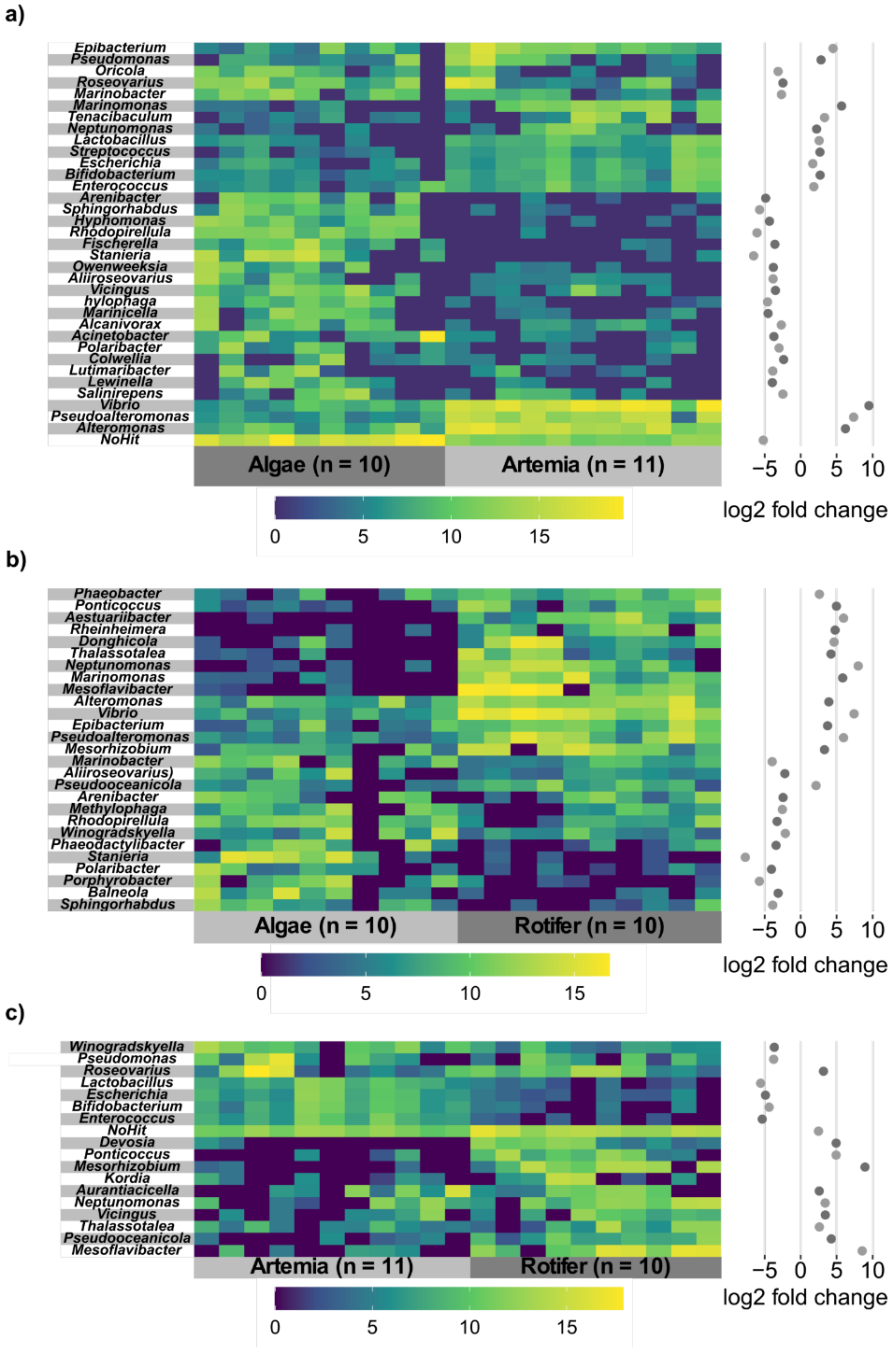


Figure 4.3.8. Differentially abundant bacterial genera comparing different types of foods; a) the factors in the model included site, time of sampling (January and May), and food type (algae and *Artemia*); b) the model factors included site, time of sampling (January and May), and food type (algae and rotifer); c) the model factors included site, time of sampling (January and May), and food type (*Artemia* and rotifer). The R package metagenomeSeq v 1.32.0 was used to identify differentially abundant OTUs with an adjusted p-value < 0.05 (Paulson et al., 2016). Bacterial families with significant changes in each comparison are visualized with heatmaps using the R package superheat v 1.0.0 (Barter and Yu, 2021).

There were no significant differences in the microbial abundance of the rearing water of sea bream and sea bass. Differential abundance was present, however, between source water and tank rearing water: *Maliponia, Ichthyenterobacterium, Oleispira, Marivita,* and *Phaeobacter* were more abundant in larval rearing water and a higher abundance of *Psychrobium, Bifidobacterium,* and *Escherichia* was found in source water of some hatcheries.

4.3.4.6. Functional prediction

Common KEGG orthologs (KOs) identified by both RNA-seq method-edgR algorithm and metagenomeSeq were used in KEGG enrichment analysis (**Supplementary table 4.3.8**). Enriched KEGG pathways were identified by comparison of larvae at different ages (**Figure 4.3.9a**), different foods (**Figure 4.3.9b**) and larvae - food comparisons (**Supplementary figure 4.3.11**). "Quorum sensing", "peptidoglycan biosynthesis", "starch and sucrose metabolism", "oxidative phosphorylation", and "terpenoid backbone biosynthesis" were enriched in younger larvae, while "biofilm formation", "bacterial secretion system", "lipopolysaccharide biosynthesis", "cationic antimicrobial peptide (CAMP) resistance", "biosynthesis of amino acids", "homologous recombination and bacterial invasion of epithelial cells" were enriched in older larvae (**Figure 4.3.9a**). Generally, a greater number of enriched pathways were found in *Artemia* compared to rotifer and algae. Enriched pathways in *Artemia* included "two−component system", "*Staphylococcus aureus* infection", "peptidoglycan biosynthesis", "quorum sensing", "fructose and mannose metabolism", "starch and sucrose metabolism", and "galactose metabolism". "Bacterial secretion system" was enriched in rotifers and microalgae compared to *Artemia* (**Figure 4.3.9b**). "Two-component system", "phosphotransferase system (PTS)", "peptidoglycan biosynthesis", "biosynthesis of amino acids", "starch and sucrose metabolism", "fructose and mannose metabolism", "beta-lactam resistance", "ubiquinone and other terpenoid-quinone biosynthesis", "terpenoid backbone biosynthesis" were enriched in larvae compared to rotifer and algae (**Supplementary figure 4.3.11**). "Quorum sensing" was enriched in larvae compared to rotifer and "oxidative phosphorylation" was enriched in larvae compared to microalgae (**Supplementary figure 4.3.11**). "Legionellosis", "cell cycle – *Caulobacter"*, and "arginine and proline metabolism" were enriched in algae compared to larvae (**Supplementary figure 4.3.11**). There was no significant pathway enrichment when larvae were compared to *Artemia*.

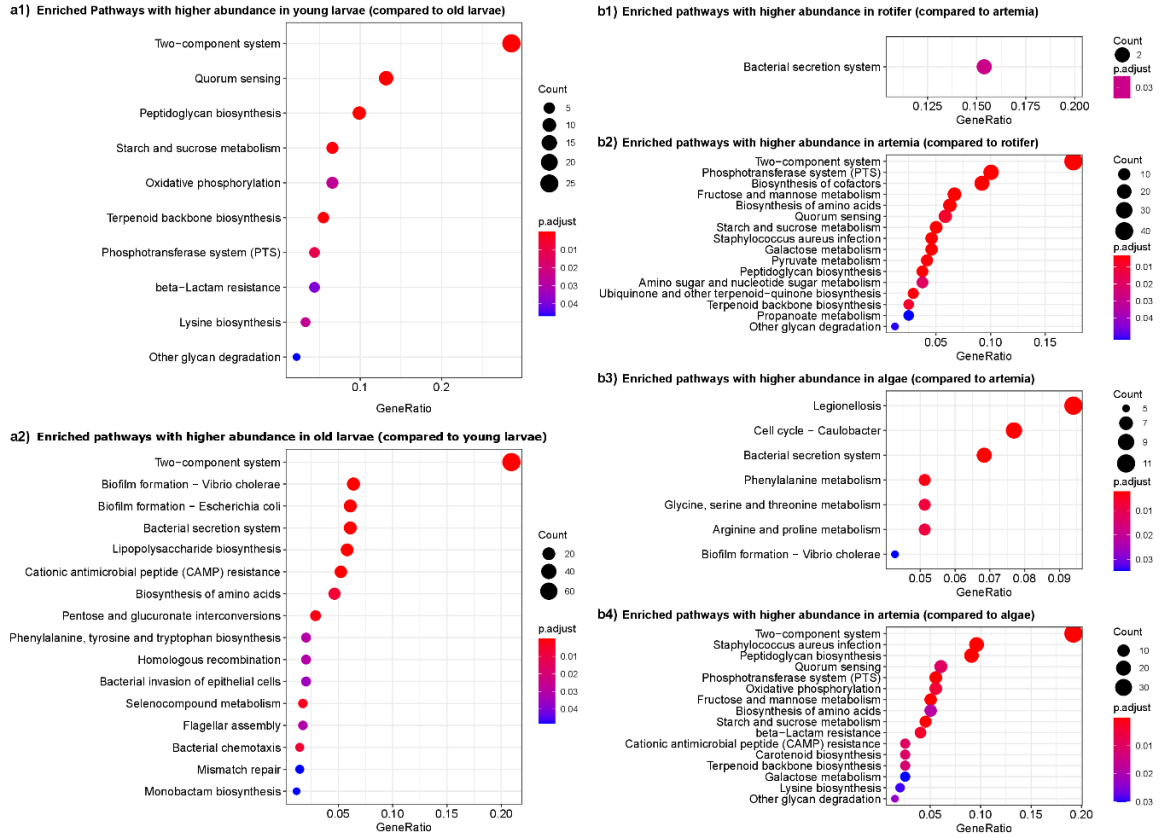


Figure 4.3.9. The enriched KEGG pathway analyse. a1) The pathways enriched in younger larvae (age range = 5-23 dph), a2) the pathways enriched in older larvae (age range = 39-77 dph). The larvae samples included both European sea bass and gilthead sea bream species. b1-b4) The enriched KEGG in each food sample using pairwise comparisons of different types of food (algae, rotifer, *Artemia*). Differentially abundant KOs were found using two methods, the RNA-seq method-edgR algorithm and metagenomeSeq. Based on age and food type variables, the common KOs obtained by the two methods and that showed higher abundance in each age group (a = younger, b = older) or food (algae, rotifer, *Artemia*) were used in KEGG enrichment analysis.

4.3.5. Discussion

The present metagenomic study shows that there are significant differences in microbial communities between larvae development stages, food items and source of water, and that some communities are abundant irrespective of sample type, geographical location or time of year. The reported large-scale approach has the advantage of identifying a common core microbiota independent of rearing conditions in each hatchery. The most abundant or core

bacteria identified in the current study will be a valuable guide for understanding microbiome development during sea bream and sea bass larval ontogeny and the modulation of the microbiota by the fish or external factors. Moreover, the diverse microbiota abundance and function across different sample types provides the basis for a robust tool for microbiome management in aquaculture.

4.3.5.1. Core and most abundant bacterial communities

The most abundant and common phyla Proteobacteria, Bacteroidetes, Firmicutes, and Actinobacteria, detected across all samples, including larval species, food and water in the present study, may be considered core or common bacterial phyla in aquaculture (seawater or freshwater), since they have also been identified among the topmost bacterial phyla in other studies of fish microbiomes. Previous studies that pointed to all or some of these phyla included the microbiomes of sea bream and sea bass egg and water (Najafpour et al., 2021a), adult skin and gill (Pimentel et al., 2017; Rosado et al., 2019, 2021), sea bream larvae and live feeds (Califano et al., 2017), cod larvae (*Gadus morhua*) and their live feeds (Bakke et al., 2015) and adult rainbow trout (*Oncorhynchus mykiss*) gut (Ingerslev et al., 2014). The presence in almost all samples of the present study (larvae, rotifer, *Artemia*, microalgae, water) of *Vibrio*, *Pseudoalteromonas, Alteromonas*, *Escherichia*, *Marinomonas*, *Phaeobacter*, *Bifidobacterium*, *Pseudomonas, Lactobacillus* and *Bacteroides* is most likely explained by their high adaptability to marine environments. This may stem from the fact they are symbiotic and contribute to nutrient exchange between the host and microbiota, e.g. *Bifidobacterium* (Turroni et al., 2008), *Lactobacillus* (Walter et al., 2011), *Bacteroides* (Comstock, 2009) and *Phaeobacter* (Fuentes et al., 2016).

The high abundance of *Vibrio* and *Pseudoalteromonas*, is in line with other studies and indicates that these two genera are among the most successful bacterial genera in marine environments (Baker-Austin et al., 2018; Parrilli et al., 2021). The identification of *Vibrio* and *Pseudoalteromonas* pathogens in hatcheries may be problematic if dysbiosis occurs, as it may lead to disease outbreaks. Together with commonalities, we also found discrepancies with previous studies of fish microbiota. For example, a study of sea bream food microbiomes identified *Loktanella* and *Paracoccus* as dominant in rotifer and *Artemia*, respectively (Califano et al., 2017). In contrast, in our data *Loktanella* had a maximum abundance of 1 % in one *Artemia* sample from ItS (*Artemia*-IUJ) and *Paracoccus* was among the top bacteria in only two *Artemia* samples (*Artemia*-HM and GM, GrS2). These differences emphasize the need

when studying fish microbiota from an aquaculture perspective to have samples with sufficient breadth and scope to overcome bias driven by specific and restricted environmental conditions, particularly if the aim is to develop comprehensive disease management tools. For example, the use of the most efficient probiotics in diverse hatchery conditions, the identification of the core microbiota, or potential beneficial interactions between bacteria and the host in different environments should be specified by comprehensive studies that consider multiple variables across multiple sites.

4.3.5.2. Diverse bacteria across larvae, food and water

Although a common core microbiota existed across different samples there was a divergence in the global microbiome of the sea bass and sea bream larvae that suggested that sample type (microalgae, *Artemia*, rotifer, and water), larval age, and geographical location influenced the composition of the microbiota, as previously observed (Bakke et al., 2015; Califano et al., 2017). A higher richness and lower Shannon diversity was reported for rearing water compared to sea bream (Califano et al., 2017) although in our study the relationship did not achieve statistical significance. The difference in microbial abundance and composition of different types of food is suggested as a potential modulator of the larval microbiome and highlights one of the challenges of feeding larva with live food in hatcheries. For example, the high bacterial richness or diversity of live feeds (marine invertebrates) means that in addition to providing essential nutrients for larvae they may also have positive benefits for the establishment of the larval microbiota (Vázquez-Salgado et al., 2020). The lower bacterial richness of algae compared to other live foods and their capacity to disrupt bacterial communication or quorum sensing (Natrah et al., 2011) may be advantageous in relation to pathogens. The lower abundance of *Vibrionaceae* (*Vibrio* spp.) and *Pseudoalteromonadaceae* (*Pseudoalteromonas* spp.) in algae compared to *Artemia* and rotifers may indicate that algae have a minor contribution to the establishment of these two bacterial genera in the larval microbiome. The lifecycle and culture condition of rotifers, such as very high organic loads due to rotifer feed, dead rotifers, and faecal matter (Dhont et al., 2013), may explain the higher bacterial richness of rotifers compared to algae and *Artemia*. The contribution of *Artemia* for the transfer of beneficial gut microbiota in fish larvae such as *Bifidobacterium* (Turroni et al., 2008), *Lactobacillus* (Walter et al., 2011) and *Enterococcus* (Ringø et al., 2010) is probably high since these bacteria are more abundant in *Artemia* compared to rotifer and algae.

4.3.5.3. Seasonal effects on the microbial community

The influence of season and temperature change on the gut microbiota has been reported in several studies (Bereded et al., 2021; Dulski et al., 2020; Hagi et al., 2004). Overall, in the present study there were small changes in the microbial community between January and May, possibly because of the consistent and controlled environmental parameters (e.g. temperature, oxygen) in closed circuit / semi-closed circuit of the aquaculture units of the sea bream and sea bass hatcheries. Nonethless, a significant change in abundance of some bacterial genera was observed such as the increase in May compared to January of the relative abundance of *Aestuariibacter* (mainly in water, larvae, and *Artemia*) and *Amphritea* (mainly in water and larvae) or the decrease of *Blautia* (mainly in water, larvae, microalgae, and rotifers).

4.3.5.4. Effect of geographical location on the microbial community

The divergence in bacterial communities between sites was less pronounced, possibly because of the limited number of replicated samples from each site or the higher differences in the microbiol community of different sample types that are included in a single PcoA. However, the differences in the relative abundance of bacteria in samples from different sites and the separation between specimens that belonged to the same group (e.g., the separation between algae samples in January) suggest an effect of the geographical location on the abundance of microbial communities. An effect of geographical location on the microbial community of water and eggs of sea bass and sea bream has previousy been reported (Najafpour et al., 2021a). The presence of a different abundance of bacterial genera in the rearing tank water and source water in some sites is most likely a consequence of the water treatment regime, management practices and the introduction of feed and even larvae.

4.3.5.5. Age related microbial community

Developmental stage, diet, and rearing water are three factors previously reported to modulate the gut microbiome with age (Bakke et al., 2015; Llewellyn et al., 2014; Wong and Rawls, 2012) and the difference in the microbial composition of sea bass and sea bream larvae at different developmental stages in this study corroborated previous observations in zebrafish adults (Yan et al., 2012), cod larvae (Bakke et al., 2015), and sea bream and sea bass adults (Rosado et al., 2021). In the present study some bacterial genera typical of the adult gut, such as *Lactobacillus*, *Streptococcus* and *Blautia* (Firmicutes), and *Bifidobacterium* and *Bacteroides* were at higher abundance in early larval stages, and we speculate this may indicate a significant

change in the gut microbiota occurs with age.

Host selection of beneficial bacteria to establish a mutualistic relationship may explain the higher abundance of beneficial bacteria (e.g., *Bifidobacterim* and *lactobacillus*) during early development (Bakke et al., 2015; Sullam et al., 2012), and provide the selective pressure needed to boost the host immune system (Swain et al., 2006). Live foods with a high abundance of beneficial bacteria may play an important role in modulating the larval microbiome composition. An example of this is *Mesoflavibacter*, which was abundant in larvae at start of feeding compared to mid-metamorphosis and presumably originated from the rotifers where they are also abundant and, which are fed to early larvae. Most of the abundant bacteria in early larval stages are characteristic of the gut microbiota and are anaerobic, while the shift in dominance to other genera (e.g., *Alkalimarinus, Leucothrix, Pseudomonas, and Agarivorans*) in older larvae (e.g., metamorphosis) may reflect modulation of the microbiota by species-specific morpho-physiochemical changes (e.g., increasing intestinal folds, Najafpour et al., 2021b) and the change in feed regime that shifts from algae to *Artemia*. The abundance of bacteria genera in larvae, e.g., *Bifidobacterium, Lactobacillus, Bacteroides,* and *Blautia* suggest they may colonize larvae at early stages and prevent host intestinal colonization by other microorganisms such as the anaerobic bacteria, *Blautia producta* and *Clostridium bolteae,* and in this way prevent infection by vancomycin-resistant enterococci (Barbeiro et al., 2020; Caballero et al., 2017). The consequence of the high abundance of *Streptococcus* when larvae start to feed needs to be further studied since this genus includes species that may be pathogens or probiotics.

4.3.5.6. Host - bacteria interactions during larval development

Bacteria use quorum sensing to regulate a diversity of physiological activities, e.g., symbiosis, virulence, competence, conjugation, antibiotic production, motility, sporulation, and biofilm formation (Miller and Bassler, 2001). Since the most abundant bacteria at start feeding (compared to mid-metamorphosis) were associated with the gut, the "quorum sensing" enrichment in younger larvae (5-23 dph) may indicate symbiosis and the colonization of the gut environment by bacteria (Jimenez and Sperandio, 2019). Enrichment in "biofilm formation" pathways in mid-metamorphosis (39-77 dph) larvae may be related to the increased surface area and villosity of the gut as it matures. Although this process is expected to be beneficial it is also associated with infections and pathogenesis in humans (Bjarnsholt, 2013; Hall-Stoodley and Stoodley, 2009). Considering the abundance of *Vibrio* and that some are

pathogenic biofilm forming species (Tan et al., 2015) they could pose a disease risk in dysbiosis. Enrichment of the "bacterial invasion of epithelial cells" pathway further emphasizes the potential for pathogenic bacteria colonization in early larval stages. It should be noted, however, that bacterial metabolic pathways are mainly named based on human nomenclature. For instance, biofilm formation during sea bream and sea bass larval development in this study is likely to be by other bacteria species rather than *V. cholerae* and *E. coli*, which are associated with the biofilm pathway in human.

Enrichment in fish larvae microbiota of "peptidoglycan biosynthesis" and "beta−Lactam resistance" pathways in younger larvae and "lipopolysaccharide biosynthesis" and "cationic antimicrobial peptide (CAMP) resistance" pathways at mid-metamorphosis is suggestive of bacterial resistance caused by major modifications in the microbial cell wall and defense against potential host antimicrobial peptides or antibiotic usage in aquaculture (Anaya-López et al., 2013; Chen et al., 2018; Nikolaidis et al., 2014; Rebl and Goldammer, 2018; Ryu et al., 2012). Although, a further interesting possibility linked to the enrichment terms that requires further investigation is the developmental status of the larval immune system. Different resistance mechanisms of bacteria across larval stages may be a consequence of changes in bacterial composition, e.g., a change in gram-positive and gram-negative bacterial communities. Exposure to the host oxidative stress defenses can lead to bacterial DNA damage followed by homologous recombination DNA repair (Michod et al., 2008), as suggested by enrichment of the "homologous recombination" pathway at the mid-metamorphosis stage.

The enrichment of the "starch and sucrose metabolism" pathway in younger larval stages (5-23 dph) and "biosynthesis of amino acids" and "pentose and glucuronate interconversions" pathways at the mid-metamorphosis stage suggests a potential trade-off in the metabolism of the host and microbiota during larval development. "Enrichment of phenylalanine, tyrosine, and tryptophan biosynthesis" pathways at mid-metamorphosis suggests that an increase in biosynthesis of specific amino acids could be a key factor in increasing bacterial load/abundance and colonization of the host. Biosynthesis of several amino acids was previously reported to be enriched in older age groups of sea bass and sea bream, when comparing the gill and skin microbiota of mature and juvenile fish (Rosado et al., 2021) although this is not directly comparable with larvae microbiota it may suggest an increasing trend in amino acids biosynthesis with age. Enhanced biosynthesis and metabolism pathways for carbohydrates, amino acids, and lipids in ryegrass-fed grass carp was proposed as a response to a low-protein diet (Ni et al., 2014). The pairwise comparison in the present study of the

microbiota functional prediction between algae, *Artemia* and rotifer suggests diverse impacts of different live foods on the fish microbiota. The *Artemia* microbiota was enriched in the "biosynthesis of amino acids" pathway, which is essential for bacterial growth and survival (Amorim Franco and Blanchard, 2017) and allows pathogen growth in amino acid-deficient environments (Li et al., 2019; Zhang and Rubin, 2013). Potential deficiency in amino acids may result from a host cell response to deplete intracellular amino acid (e.g., tryptophan) during pathogen invasion (Silva et al., 2002) or a result of the lack of essential amino acids (e.g., phenylalanine and tryptophan) in fish food.

Overall, our data shows that microbial abundance in larvae and *Artemia* are more similar and that more pathways are enriched in *Artemia* than the other live feeds. Both beneficial and harmful effects on larval microbiota are predicted from the enriched pathways in *Artemia* (compared to other live feeds). Beneficial would be supporting the host with carbohydrate metabolism (e.g., the enrichment of "fructose and mannose metabolism", "starch and sucrose metabolism", and "galactose metabolism" pathways). Harmful would be potential vectors to increase bacterial resistance or transfer of pathogens (e.g., the enrichment in "peptidoglycan biosynthesis" or "beta-lactum resistance" pathways).

4.3.6. Conclusions

A large-scale field study of aquaculture hatcheries geographically separated in the Mediterranean identified the core (e.g., *Vibrio* and *Pseudoalteromonas*) and most abundant bacteria in different types of samples (larvae, algae, rotifer, *Artemia*, water). Seasonality (January vs May) had a minor effect on the microbial community possibly because physico-chemical conditions are kept constant in the closed recirculating systems of hatcheries. Geographical location influenced the abundance of some bacteria. The microbiota of Euroepan sea bass and gilthead sea bream larvae was similar with some exceptions while the microbial community of their tank water did not differ. Both live foods and rearing water contributed to the larval microbiota during larval ontogeny, but with differing importance. There was a higher possibility of *Vibrio* transfer to larvae with *Artemia* and rotifer than with algae and water. This study suggests strong host selection of beneficial bacteria to establish a mutualistic relationship in very early larvae (at the start of feeding) compared to mid-metamorphosis larvae and suggested *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, and *Blautia* as part of the beneficial communities which colonize larvae at early stages. Functional analyses of significantly higher abundance microbiota predicted a potential interaction between the host and microbiota during

larval development that may be linked to requirements of bacteria for growth and resistance and to changes in host demand, physiology, immune system and morphology. In respect to the effects of feeds, *Artemia* microbiota seemed to have a higher contribution to biosynthesis of amino acids and carbohydrate metabolism. The low relative abundance of the beneficial bacteria in live feeds and tank water indicates manipulation of bacteria in aquaculture facilities using food or water enrichment techniques (e.g., probiotics) may be an effective way to counter negative effects of abundant of potentially pathogenic *Vibrio* and *Pseudoalteromonas* outbreaks.

4.3.7. See ANNEX IV for Supplementary materials

Supplementary table 4.3.1. The list and codes of samples used for DNA extraction and microbiome analysis from different locations and sampling times: Greece sampling 1 (GrS1, January 2018), Greece Sampling 2 (GrS2, May 2018) and Italy Sampling (ItS, June 2018). This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.2. Sequencing statistics of the 16S rRNA gene libraries, compiled for different locations and sampling dates. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.3. The relative proportions of the detected bacteria (present at a minimum of 1% in the sum of all libraries) at the level of the phylum, family, genus or species in the microbiome libraries obtained from Greece Sampling 1. Sample codes are descried in **Supplementary table 4.3.1**. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.4. The relative proportions of the detected bacteria (present at a minimum of 1% in the sum of all libraries) at the level of the phylum, family, genus or species in the microbiome libraries obtained from Greece Sampling 2. Sample codes are descried in **Supplementary table 4.3.1**. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.5. The relative proportions of the detected bacteria (present at a minimum of 1% in the sum of all libraries) at the level of the phylum, family, genus or species in the microbiome libraries obtained from Italy sampling. Sample codes are descried in **Supplementary table 4.3.1** This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.6. The top ten bacterial genera detected by 16s rRNA gene sequencing in each sample, separated by sample type and samplings. Sample codes are descried in **Supplementary table 4.3.1.** This is only available in digital format in Annex IV because the

table is very extensive.

Supplementary table 4.3.7. Chao1 Richness Estimation and Shannon value from different sampling time in Greece (Jan 2018 and May 2018) and in Italy (June) as an out group. Sample codes are descried in **Supplementary table 4.3.1**. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary table 4.3.8. Differential functional analysis of metagenomeSeq method using KEGG orthologous predicted by Tax4Fun in larvae samples (January and May) based on different factors. This is only available in digital format in Annex IV because the table is very extensive.

Supplementary figure 4.3.1. Rarefaction plots of sequencing data from the 16S rRNA microbiome libraries prepared with the Greek samples (GrS1) collected in January 2018. The plot gives an indication of the bacterial diversity within the samples, determined using all operational taxonomic units (OTUs) found at the genus level, including annotated and non-annotated sequences (no hits). The vertical-coloured lines represent the range of the rarefaction curves for each sample type: larvae - green; water - pink; and food (algae - red, artemia - yellow and rotifer – blue). This is only available in digital format in Annex IV.

Supplementary figure 4.3.2. Rarefaction plots of sequencing data from the 16S rRNA microbiome libraries prepared with the Greek samples (GrS2) collected in May 2018. The plot gives an indication of the bacterial diversity within the samples, determined using all operational taxonomic units (OTUs) found at the genus level, including annotated and non-annotated sequences (no hits). The vertical-coloured lines represent the range of rarefaction curves for each sample type: larvae - green; water - pink; and food (algae - yellow, artemia - red and rotifer – blue). This is only available in digital format in Annex IV because the table is very extensive.

Supplementary figure 4.3.3. Rarefaction plots of sequencing data from the 16S rRNA microbiome libraries prepared from the Italian samples (ItS) collected in June 2018. The plot gives an indication of the bacterial diversity within the samples, determined using all

operational taxonomic units (OTUs) found at the genus level, including annotated and non-annotated sequences (no hits). This is only available in digital format in Annex IV.

Supplementary figure 4.3.4. UpSetR plot of the core bacteria genera in different types of the combined samples from three sampling times (Greece Sampling 1, Greece Sampling 2, Italy Sampling). The bars at the left represent the total number of the detected bacteria genera in each sample (set size). The upper panel bars represent the number of bacteria genera that are common (connected dots) among most of the samples. The intersects are ordered by degree and only 60 intersects are shown. The identity of the bacteria genera for the 60 presented intersects are listed at the legend at the right side of the figure and the related intersects numbering are shown on the plot (using the same number of genera) in blue at the bottom of the figure. Finally, all samples also contained No Hit reads (grouped in intersect number 1), for which no related hit was retrieved from the 16S rRNA gene database, although in general these represented a low proportion of the described microbiomes (**Supplementary table 4.3.2**; average of No hits 3.3%, maximum 22%). Samples codes description is provided in **Supplementary table 4.3.1**. This is only available in digital format in Annex IV.

Supplementary figure 4.3.5. Heatmap representing the relative abundance of the bacteria genera identified in the microbiome libraries from Greece Sampling 1 (January), with a hierarchical clustering tree grouping them according to microbiome similarities. Each horizontal line represents one genus identified through 16S rRNA gene sequencing, with labels on the right panel for the genera identified at more than 8%, and the shading of each cell represents its percentage relative to the total microbiome in each library (according to the coloured scale). “No hits” corresponds to the relatively low proportion of sequences that were not annotated by the NCBI 16S rRNA gene database. Different types of samples are shown with different symbols: European sea bass larvae (); gilthead sea bream larvae (); algae (); rotifer (); *Artemia* (); water (). For larvae, different age ranges are presented with code 1 (6-11 dph), 2 (20-23 dph) and 3 (42-77 dph). Different letter (A, B, C, D) denote different hatcheries and the times of sampling (January) is indicated with 'J' after each hatchery. Water from source (S) and fish tanks (T) are specified with related letter (See also **Supplementary table 4.3.1**). This is only available in digital format in Annex IV.

Supplementary figure 4.3.6. Heatmap representing the relative abundance of the bacteria

genera identified in the microbiome libraries from Greece Sampling 2 (May), with a hierarchical clustering tree grouping them according to microbiome similarities. Each horizontal line represents one genus identified through 16S rRNA gene sequencing, with labels on the right panel for the genera identified at more than 8 %, and the shading of each cell represents its percentage relative to the total microbiome in each library (according to the coloured scale). “No hits” corresponds to the relatively low proportion of sequences that were not annotated by the NCBI 16S rRNA gene database. Different types of samples are shown with different symbols: European sea bass larvae (); gilthead sea bream larvae (); algae (); rotifer (); *Artemia* (); water (). For larvae, different age ranges are presented with code 1 (5-7 dph) and 3 (39-58 dph). Different letter (A, B, D, E, F, G, H) show different hatcheries and the times of sampling (May) shown with 'M' after each hatchery. Water from source (S), UV treatment (U) and fish tank (T) are specified with related letter (See also **Supplementary table 4.3.1**). This is only available in digital format in Annex IV.

Supplementary figure 4.3.7. Heatmap representing the relative abundance of the bacteria genera identified in the microbiome libraries from the Italy sampling (June), with a hierarchical clustering tree grouping them according to microbiome similarities. Each horizontal line represents one genus identified through 16S rRNA gene sequencing, with labels on the right panel for the genera identified at more than 8 %, and the shading of each cell represents its percentage relative to the total microbiome in each library (according to the coloured scale). “No hits” correspond to the relatively low proportion of sequences that were not annotated by the NCBI 16S rRNA gene database. Different types of samples are shown with different symbols: gilthead sea bream larvae (); algae (); rotifer (); *Artemia* (); water ( ). Based on age classification in this study, 15 dph larvae is shown with code 2. Sampling hatchery (I) and time (JU) are specified for each sample. Water from source (S) and fish tank (T) are presented with related letter (See also **Supplementary table 4.3.1**). This is only available in digital format in Annex IV.

Supplementary figure 4.3.8. The correlation between genus-specific copy number established by qPCR quantification and the relative abundance of five bacterial genera determined by 16S rRNA gene sequencing of gilthead sea bream and European sea bass larvae samples. The scatter plots were generated using the “Spearman” method in an R environment. This is only available

in digital format in Annex IV.

Supplementary figure 4.3.9. Alpha diversity measured by the Chao 1 Richness Estimator and Shannon indexes for each sample. Different types of samples are shown in different colours: *Artemia* in red; Algae in green; Rotifer in gold; Water in blue; Larvae (L.) in grey. Two boxes for each type of sample represents the different sampling times: Greece Sampling 1 in January (without background) and Greece Sampling 2 in May (with dots background). This is only available in digital format in Annex IV.

Supplementary figure 4.3.10. Differentially abundant bacterial genera comparing larvae and live foods (algae, rotifer and *Artemia*). a) the model factors included site, time of sampling (January and May), and sample type (algae, larvae); b) the model factors included site, time of sampling (January and May) and sample type (larvae and rotifer); c) the model factors included site, time of sampling (January and May) and sample type (larvae and *Artemia*). The R package metagenomeSeq v 1.32.0 was used to identify differentially abundant OTUs with an adjusted p-value < 0.05 (Paulson et al., 2016). Bacterial genera with significant changes in each comparison are visualized with heatmaps using the R package superheat v 1.0.0 (Barter and Yu, 2021). This is only available in digital format in Annex IV because the table is very extensive.

Supplementary figure 4.3.11. The enriched KEGG pathways comparing larvae and food type (rotifer and algae). The larvae samples included both European sea bass and gilthead sea bream species. The differentially abundant KOs were found using two methods (RNA-seq method-edgR algorithm and metagenomeSeq) based on the food type variable and common KOs obtained from the two methods that showed higher abundance in larvae compared to rotifer (a) and algae (b) or showed higher abundance in algae compared to larvae (c) were used in KEGG enrichment analysis. This is only available in digital format in Annex IV.

4.3.8. References

Abdel-Aziz, M., Eissa, A.E., Hanna, M., Okada, M.A., 2013. Identifying some pathogenic *Vibrio*/*Photobacterium* species during mass mortalities of cultured Gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) from some Egyptian coastal provinces. Int. J. Vet. Sci. Med. 1, 87–95. https://doi.org/10.1016/J.IJVSM.2013.10.004

Abelli, L., Randelli, E., Carnevali, O., Picchietti, S., 2009. Stimulation of gut immune system by early administration of probiotic strains in *Dicentrarchus labrax* and *Sparus aurata*. Ann. N. Y. Acad. Sci. 1163, 340–342. https://doi.org/10.1111/j.1749-6632.2008.03670.x

Amorim Franco, T.M., Blanchard, J.S., 2017. Bacterial branched-chain amino acid biosynthesis: structures, mechanisms, and drugability. Biochemistry 56, 5849. https://doi.org/10.1021/ACS.BIOCHEM.7B00849

Anaya-López, J.L., López-Meza, J.E., Ochoa-Zarzosa, A., 2013. Bacterial resistance to cationic antimicrobial peptides. Crit. Rev. Microbiol. 39, 180–195. https://doi.org/10.3109/1040841X.2012.699025

Apprill, A., McNally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75, 129–137. https://doi.org/10.3354/ame01753

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Austin, B., 2006. The bacterial microflora of fish, revised. Sci. World J. 6, 931–945. https://doi.org/10.1100/tsw.2006.181

Baker-Austin, C., Oliver, J.D., Alam, M., Ali, A., Waldor, M.K., Qadri, F., Martinez-Urtaza, J., 2018. *Vibrio* spp. infections. Nat. Rev. Dis. Prim. 2018 41 4, 1–19. https://doi.org/10.1038/s41572-018-0005-8

Bakke, I., Coward, E., Andersen, T., Vadstein, O., 2015. Selection in the host structures the microbiota associated with developing cod larvae (*Gadus morhua*). Environ. Microbiol. 17, 3914–3924. https://doi.org/10.1111/1462-2920.12888

Barbeiro, H.V., Barbeiro, D.F., Soriano, F.G., 2020. Sepsis: future role of omics in diagnosis and therapy. Precis. Med. Investig. Pract. Provid. 281–289. https://doi.org/10.1016/B978-0-12-819178-1.00027-7

Barter, R., Yu, B., 2021. superheat: agraphical tool for exploring complex datasets using heatmaps. R package version 1.0.0.

Belkaid, Y., Hand, T., 2014. Role of the microbiota in immunity and inflammation. Cell 157, 121. https://doi.org/10.1016/J.CELL.2014.03.011

Bereded, N.K., Abebe, G.B., Fanta, S.W., Curto, M., Waidbacher, H., Meimberg, H., Domig, K.J., 2021. The impact of sampling season and catching site (wild and aquaculture) on gut microbiota composition and diversity of nile tilapia (*Oreochromis niloticus*). Biol. 10, 180. https://doi.org/10.3390/BIOLOGY10030180

Bjarnsholt, T., 2013. The role of bacterial biofilms in chronic infections. APMIS. Suppl. 1–51. https://doi.org/10.1111/APM.12099

Brugman, S., Ikeda-Ohtsubo, W., Braber, S., Folkerts, G., Pieterse, C.M.J., Bakker, P.A.H.M., 2018. A comparative review on microbiota manipulation: lessons from fish, plants, livestock, and human research. Front. Nutr. 5, 80. https://doi.org/10.3389/FNUT.2018.00080/BIBTEX

Caballero, S., Kim, S., Carter, R.A., Leiner, I.M., Sušac, B., Miller, L., Kim, G.J., Ling, L., Pamer, E.G., 2017. Cooperating commensals restore colonization resistance to vancomycin- resistant *Enterococcus faecium*. Cell Host Microbe 21, 592. https://doi.org/10.1016/J.CHOM.2017.04.002

Califano, G., Castanho, S., Soares, F., Ribeiro, L., Cox, C.J., Mata, L., Costa, R., 2017. Molecular taxonomic profiling of bacterial communities in a gilthead seabream (*Sparus aurata*) hatchery. Front. Microbiol. 8, 204. https://doi.org/10.3389/FMICB.2017.00204/BIBTEX

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10, 1–9. https://doi.org/10.1186/1471-2105-10-421/FIGURES/4

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Carnevali, O., de Vivo, L., Sulpizio, R., Gioacchini, G., Olivotto, I., Silvi, S., Cresci, A., 2006. Growth improvement by probiotic in European sea bass juveniles (*Dicentrarchus labrax*, L.), with particular attention to IGF-1, myostatin and cortisol gene expression. Aquaculture 258, 430–438. https://doi.org/10.1016/J.AQUACULTURE.2006.04.025

Chen, B., Lin, L., Fang, L., Yang, Y., Chen, E., Yuan, K., Zou, S., Wang, X., Luan, T., 2018. Complex pollution of antibiotic resistance genes due to beta-lactam and aminoglycoside use in aquaculture farming. Water Res. 134, 200–208. https://doi.org/10.1016/J.WATRES.2018.02.003

Chong, J., Liu, P., Zhou, G., Xia, J., 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat. Protoc. 15, 799–821. https://doi.org/10.1038/s41596-019-0264-1

Codoñer, F.M., Ramírez-Bosca, A., Climent, E., Carrión-Gutierrez, M., Guerrero, M., Pérez-Orquín, J.M., Horga De La Parte, J., Genovés, S., Ramón, D., Navarro-López, V., Chenoll, E., 2018. Gut microbial composition in patients with psoriasis. Sci. Reports 2018 81 8, 1–7. https://doi.org/10.1038/s41598-018-22125-y

Comstock, L.E., 2009. Importance of Glycans to the Host-Bacteroides Mutualism in the Mammalian Intestine. Cell Host Microbe 5, 522–526. https://doi.org/10.1016/J.CHOM.2009.05.010

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cordero, H., Guardiola, F.A., Tapia-Paniagua, S.T., Cuesta, A., Meseguer, J., Balebona, M.C., Moriñigo, M.Á., Esteban, M.Á., 2015. Modulation of immunity and gut microbiota after

dietary administration of alginate encapsulated *Shewanella* putrefaciens Pdp11 to gilthead seabream (*Sparus aurata* L.). Fish Shellfish Immunol. 45, 608–618. https://doi.org/10.1016/J.FSI.2015.05.010

Deng, Y., Verdegem, M.C.J., Eding, E., Kokou, F., 2022. Effect of rearing systems and dietary probiotic supplementation on the growth and gut microbiota of Nile tilapia (*Oreochromis niloticus*) larvae. Aquaculture 546, 737297. https://doi.org/10.1016/J.AQUACULTURE.2021.737297

Dhont, J., Dierckens, K., Støttrup, J., Van Stappen, G., Wille, M., Sorgeloos, P., 2013. Rotifers, Artemia and copepods as live feeds for fish larvae in aquaculture. Adv. Aquac. Hatch. Technol. 157–202. https://doi.org/10.1533/9780857097460.1.157

Dimitroglou, A., Merrifield, D.L., Spring, P., Sweetman, J., Moate, R., Davies, S.J., 2010. Effects of mannan oligosaccharide (MOS) supplementation on growth performance, feed utilisation, intestinal histology and gut microbiota of gilthead sea bream (*Sparus aurata*). Aquaculture 300, 182–188. https://doi.org/10.1016/J.AQUACULTURE.2010.01.015

Du, S., Chen, W., Yao, Z., Huang, X., Chen, C., Guo, H., Zhang, D., 2021. *Enterococcus faecium* are associated with the modification of gut microbiota and shrimp post-larvae survival. Anim. Microbiome 2021 31 3, 1–15. https://doi.org/10.1186/S42523-021-00152-X

Dulski, T., Kozłowski, K., Ciesielski, S., 2020. Habitat and seasonality shape the structure of tench (*Tinca tinca* L.) gut microbiome. Sci. Reports 2020 101 10, 1–11. https://doi.org/10.1038/s41598-020-61351-1

Egerton, S., Culloty, S., Whooley, J., Stanton, C., Ross, R.P., 2018. The gut microbiota of marine fish. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.00873

FAO, 2018. State of fisheries and aquaculture in the world 2018. Meeting the sustainable development goals. Rome.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. https://doi.org/10.1093/BIOINFORMATICS/BTS565

Fuentes, J.L., Garbayo, I., Cuaresma, M., Montero, Z., González-Del-Valle, M., Vílchez, C., 2016. Impact of microalgae-bacteria interactions on the production of algal biomass and associated compounds. Mar. Drugs 2016, 14, 100. https://doi.org/10.3390/MD14050100

Ghanbari, M., Kneifel, W., Domig, K.J., 2015. A new view of the fish gut microbiome: advances from next-generation sequencing. Aquaculture. https://doi.org/10.1016/j.aquaculture.2015.06.033

Gómez, G.D., Balcázar, J.L., 2007. A review on the interactions between gut microbiota and innate immunity of fish. FEMS Immunol. Med. Microbiol. 52, 145–154. https://doi.org/10.1111/j.1574-695X.2007.00343.x

Hagi, T., Tanaka, D., Iwamura, Y., Hoshino, T., 2004. Diversity and seasonal changes in lactic acid bacteria in the intestinal tract of cultured freshwater fish. Aquaculture 234, 335–346. https://doi.org/10.1016/J.AQUACULTURE.2004.01.018

Hall-Stoodley, L., Stoodley, P., 2009. Evolving concepts in biofilm infections. Cell. Microbiol. 11, 1034–1043. https://doi.org/10.1111/J.1462-5822.2009.01323.X

Ingerslev, H.C., von Gersdorff Jørgensen, L., Lenz Strube, M., Larsen, N., Dalsgaard, I., Boye, M., Madsen, L., 2014. The development of the gut microbiota in rainbow trout (*Oncorhynchus mykiss*) is affected by first feeding and diet type. Aquaculture 424–425, 24–34. https://doi.org/10.1016/J.AQUACULTURE.2013.12.032

Jimenez, A.G., Sperandio, V., 2019. Quorum sensing and the gut microbiome. Quor. Sens. Mol. Mech. Biotechnol. Appl. 151–169. https://doi.org/10.1016/B978-0-12-814905-8.00006-X

Kahla-Nakbi, A. Ben, Chaieb, K., Besbes, A., Zmantar, T., Bakhrouf, A., 2006. Virulence and enterobacterial repetitive intergenic consensus PCR of *Vibrio alginolyticus* strains isolated from Tunisian cultured gilthead sea bream and sea bass outbreaks. Vet. Microbiol. 117, 321–327. https://doi.org/10.1016/j.vetmic.2006.06.012

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1. https://doi.org/10.1093/nar/gks808

Kolde, R., 2019. pheatmap: Pretty Heatmaps. R package version 1.0.12.

Lane, A., Hough, C., Bostock, J., 2014. The long-term economic and ecologic impact of larger sustainable aquaculture.

Li, T., Zhan, Z., Lin, Y., Lin, M., Xie, Q., Chen, Y., He, C., Tao, J., Li, C., 2019. Biosynthesis of amino acids in xanthomonas oryzae pv. oryzae is essential to its pathogenicity. Microorg. 7, 693. https://doi.org/10.3390/MICROORGANISMS7120693

Llewellyn, M.S., Boutin, S., Hoseinifar, S.H., Derome, N., 2014. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front. Microbiol. 5, 1–1. https://doi.org/10.3389/FMICB.2014.00207

Martyniuk, C.J., Kroll, K.J., Porak, W.F., Steward, C., Grier, H.J., Denslow, N.D., 2009. Seasonal relationship between gonadotropin, growth hormone, and estrogen receptor mRNA expression in the pituitary gland of largemouth bass. Gen. Comp. Endocrinol. 163, 306–317. https://doi.org/10.1016/j.ygcen.2009.04.028

Merrifield, D.L., Rodiles, A., 2015. The fish microbiome and its interactions with mucosal tissues, in: Mucosal Health in Aquaculture. Elsevier Inc., pp. 273–295. https://doi.org/10.1016/B978-0-12-417186-2.00010-8

Michod, R.E., Bernstein, H., Nedelcu, A.M., 2008. Adaptive value of sex in microbial pathogens. Infect. Genet. Evol. 8, 267–285. https://doi.org/10.1016/J.MEEGID.2008.01.002

Miller, M.B., Bassler, B.L., 2001. Quorum sensing in bacteria. Annu. Rev. Microbiol. 55, 165–199. https://doi.org/10.1146/ANNUREV.MICRO.55.1.165

Najafpour, B., Dorafshan, S., Heyrati, F.P., Canario, A.V.M., Power, D.M., 2021a. Comparative ontogeny of the digestive tract of *Oncorhynchus mykiss* ♀ x *Salmo trutta caspius* ♂ triploid hybrids to their parental species. Aquac. Nutr. 27, 427–438. https://doi.org/10.1111/ANU.13196

Najafpour, B., Pinto, P., Canario, A., Power, D., 2022. Quantifying dominant bacterial genera detected in metagenomic data from fish eggs and larvae using genus-specific primers. Authorea Prepr. https://doi.org/10.22541/AU.164652122.26782894/V1

Najafpour, B., Pinto, P.I.S., Moutou, K.A., Canario, A.V.M., Power, D.M., 2021b. Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream. Microorg. 9, 2275. https://doi.org/10.3390/MICROORGANISMS9112275

Natrah, F.M.I., Defoirdt, T., Sorgeloos, P., Bossier, P., 2011. Disruption of bacterial cell-to-cell communication by marine organisms and its relevance to aquaculture. Mar. Biotechnol. (NY). 13, 109–126. https://doi.org/10.1007/S10126-010-9346-3

Navarrete, P., Fuentes, P., De la Fuente, L., Barros, L., Magne, F., Opazo, R., Ibacache, C., Espejo, R., Romero, J., 2013. Short-term effects of dietary soybean meal and lactic acid bacteria on the intestinal morphology and microbiota of Atlantic salmon (*Salmo salar*). Aquac. Nutr. 19, 827–836. https://doi.org/10.1111/anu.12047

Ni, J., Yan, Q., Yu, Y., Zhang, T., 2014. Factors influencing the grass carp gut microbiome and its effect on metabolism. FEMS Microbiol. Ecol. 87, 704–714. https://doi.org/10.1111/1574-6941.12256

Nikolaidis, I., Favini-Stabile, S., Dessen, A., 2014. Resistance to antibiotics targeted to the bacterial cell wall. Protein Sci. 23, 243. https://doi.org/10.1002/PRO.2414

Oksanen, J., F. Guillaume Blanchet, M.F., Kindt, R., Pierre Legendre, McGlinn, D., Minchin, P.R., O'Hara, R.B., L.Simpson, G., Solymos, P., Stevens, M.H.H., And, E.S., Wagner, H., 2020. vegan: community ecology package. R package version 2.5-7.

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parrilli, E., Tedesco, P., Fondi, M., Tutino, M.L., Lo Giudice, A., de Pascale, D., Fani, R., 2021. The art of adapting to extreme environments: The model system *Pseudoalteromonas*. Phys. Life Rev. 36, 137–161. https://doi.org/10.1016/J.PLREV.2019.04.003

Paulson, J.N., Colin Stine, O., Bravo, H.C., Pop, M., 2013. Robust methods for differential abundance analysis in marker gene surveys. Nat. Methods 10, 1200. https://doi.org/10.1038/NMETH.2658

Paulson, Joseph Nathaniel, Talukder, H., Pop, M., Corrada, H., Maintainer, B., Paulson, Joseph N, 2016. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. bioconductor.statistik.tu-dortmund.de.

Petti, C.A., Polage, C.R., Schreckenberger, P., 2005. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. J. Clin. Microbiol. 43, 6123–6125. https://doi.org/10.1128/JCM.43.12.6123-6125.2005

Pimentel, T., Marcelino, J., Ricardo, F., Soares, A.M.V.M., Calado, R., 2017. Bacterial communities 16S rDNA fingerprinting as a potential tracing tool for cultured seabass *Dicentrarchus labrax*. Sci. Reports 2017 71 7, 1–10. https://doi.org/10.1038/s41598-017-11552-y

Pinto, P.I.S., Guerreiro, C.C., Costa, R.A., Martinez-Blanch, J.F., Carballo, C., Codoñer, F.M., Manchado, M., Power, D.M., 2019. Understanding pseudo-albinism in sole (*Solea senegalensis*): a transcriptomics and metagenomics approach. Sci. Rep. 9, 1–16.

https://doi.org/10.1038/s41598-019-49501-6

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590. https://doi.org/10.1093/NAR/GKS1219

Rebl, A., Goldammer, T., 2018. Under control: The innate immunity of fish from the inhibitors' perspective. Fish Shellfish Immunol. https://doi.org/10.1016/j.fsi.2018.04.016

Ringø, E., Løvmo, L., Kristiansen, M., Bakken, Y., Salinas, I., Myklebust, R., Olsen, R.E., Mayhew, T.M., 2010. Lactic acid bacteria vs. pathogens in the gastrointestinal tract of fish: a review. Aquac. Res. 41, 451–467. https://doi.org/10.1111/j.1365-2109.2009.02339.x

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K., Rawls, J.F., 2011. Evidence for a core gut microbiota in the zebrafish. ISME J. 5, 1595–1608. https://doi.org/10.1038/ismej.2011.38

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4. https://doi.org/10.7717/PEERJ.2584

Rosado, D., Pérez-Losada, M., Pereira, A., Severino, R., Xavier, R., 2021. Effects of aging on the skin and gill microbiota of farmed seabass and seabream. Anim. Microbiome 2021 31 3, 1–14. https://doi.org/10.1186/S42523-020-00072-2

Rosado, D., Pérez-Losada, M., Severino, R., Cable, J., Xavier, R., 2019. Characterization of the skin and gill microbiomes of the farmed seabass (*Dicentrarchus labrax*) and seabream (*Sparus aurata*). Aquaculture 500, 57–64. https://doi.org/10.1016/J.AQUACULTURE.2018.09.063

Ryu, S.H., Park, S.G., Choi, S.M., Hwang, Y.O., Ham, H.J., Kim, S.U., Lee, Y.K., Kim, M.S., Park, G.Y., Kim, K.S., Chae, Y.Z., 2012. Antimicrobial resistance and resistance genes in *Escherichia coli* strains isolated from commercial fish and seafood. Int. J. Food Microbiol. 152, 14–18. https://doi.org/10.1016/J.IJFOODMICRO.2011.10.003

Sabree, Z.L., Kambhampati, S., Moran, N.A., 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc. Natl. Acad. Sci. 106, 19521–19526. https://doi.org/10.1073/pnas.0907504106

Silva, N.M., Rodrigues, C. V., Santoro, M.M., Reis, L.F.L., Alvarez-Leite, J.I., Gazzinelli, R.T., 2002. Expression of indoleamine 2,3-dioxygenase, tryptophan degradation, and kynurenine formation during in vivo infection with Toxoplasma gondii: induction by endogenous gamma interferon and requirement of interferon regulatory factor 1. Infect. Immun. 70, 859–868. https://doi.org/10.1128/IAI.70.2.859-868.2002

Simon, C., Daniel, R., 2011. Metagenomic analyses: past and future trends. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.02345-10

Snoussi, M., Hajlaoui, H., Noumi, E., Zanetti, S., Bakhrouf, A., 2008. Phenotypic and genetic diversity of *Vibrio alginolyticus* strains recovered from juveniles and older *Sparus aurata*

reared in a Tunisian marine farm. Ann. Microbiol. 58, 141–146. https://doi.org/10.1007/BF03179458

Sullam, K.E., Essinger, S.D., Lozupone, C.A., O'Connor, M.P., Rosen, G.L., Knight, R., Kilham, S.S., Russell, J.A., 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: A meta-analysis. Mol. Ecol. 21, 3363–3378. https://doi.org/10.1111/j.1365-294X.2012.05552.x

Suzer, C., Çoban, D., Kamaci, H.O., Saka, Ş., Firat, K., Otgucuoğlu, Ö., Küçüksari, H., 2008. *Lactobacillus* spp. bacteria as probiotics in gilthead sea bream (*Sparus aurata*, L.) larvae: Effects on growth performance and digestive enzyme activities. Aquaculture 280, 140–145. https://doi.org/10.1016/J.AQUACULTURE.2008.04.020

Swain, P., Dash, S., Bal, J., Routray, P., Sahoo, P.K., Sahoo, S.K., Saurabh, S., Gupta, S.D., Meher, P.K., 2006. Passive transfer of maternal antibodies and their existence in eggs, larvae and fry of Indian major carp, *Labeo rohita* (Ham.). Fish Shellfish Immunol. 20, 519–527. https://doi.org/10.1016/J.FSI.2005.06.011

Tall, A., Teillon, A., Boisset, C., Delesmont, R., Touron-Bodilis, A., Hervio-Heath, D., 2012. Real-time PCR optimization to identify environmental *Vibrio* spp. strains. J. Appl. Microbiol. 113, 361–372. https://doi.org/10.1111/j.1365-2672.2012.05350.x

Tan, D., Dahl, A., Middelboe, M., 2015. Vibriophages Differentially influence biofilm formation by *Vibrio anguillarum* strains. Appl. Environ. Microbiol. 81, 4489–4497. https://doi.org/10.1128/AEM.00518-15

Tanaka, M., Nakayama, J., 2017. Development of the gut microbiota in infancy and its impact on health in later life. Allergol. Int. 66, 515–522. https://doi.org/10.1016/J.ALIT.2017.07.010

Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D., Ventura, M., 2008. Human gut microbiota and bifidobacteria: From composition to functionality. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 94, 35–50. https://doi.org/10.1007/S10482-008-9232-4/FIGURES/2

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., Dufresne, A., 2015. The importance of the microbiome of the plant holobiont. New Phytol. 206, 1196–1206. https://doi.org/10.1111/nph.13312

Vázquez-Salgado, L., Olveira, J.G., Dopazo, C.P., Bandín, I., 2020. Role of rotifer (*Brachionus plicatilis*) and Artemia (*Artemia salina*) nauplii in the horizontal transmission of a natural nervous necrosis virus (NNV) reassortant strain to Senegalese sole (*Solea senegalensis*) larvae. https://doi.org/10.1080/01652176.2020.1810357 40, 205–214. https://doi.org/10.1080/01652176.2020.1810357

Walter, J., Britton, R.A., Roos, S., 2011. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proc. Natl. Acad. Sci. U. S. A. 108, 4645–4652. https://doi.org/10.1073/PNAS.1000099107/-/DCSUPPLEMENTAL

Wong, S., Rawls, J.F., 2012. Intestinal microbiota composition in fishes is influenced by host ecology and environment. Mol. Ecol. 21, 3100–3102. https://doi.org/10.1111/J.1365-294X.2012.05646.X

Wynne, J.W., Thakur, K.K., Slinger, J., Samsing, F., Milligan, B., Powell, J.F.F., McKinnon, A., Nekouei, O., New, D., Richmond, Z., Gardner, I., Siah, A., 2020. Microbiome profiling

reveals a microbial dysbiosis during a natural outbreak of tenacibaculosis (yellow mouth) in atlantic salmon. Front. Microbiol. 11, 2582. https://doi.org/10.3389/FMICB.2020.586387/BIBTEX

Yan, Q., van der Gast, C.J., Yu, Y., 2012. Bacterial community assembly and turnover within the intestines of developing zebrafish. PLoS One 7, e30603. https://doi.org/10.1371/JOURNAL.PONE.0030603

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. Omi. A J. Integr. Biol. 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Yukgehnaish, K., Kumar, P., Sivachandran, P., Marimuthu, K., Arshad, A., Paray, B.A., Arockiaraj, J., 2020. Gut microbiota metagenomics in aquaculture: factors influencing gut microbiome and its physiological role in fish. Rev. Aquac. https://doi.org/10.1111/raq.12416

Zhang, Y.J., Rubin, E.J., 2013. Feast or famine: the host–pathogen battle over amino acids. Cell. Microbiol. 15, 1079–1087. https://doi.org/10.1111/CMI.12140

Zhao, R., Symonds, J.E., Walker, S.P., Steiner, K., Carter, C.G., Bowman, J.P., Nowak, B.F., 2020. Salinity and fish age affect the gut microbiota of farmed Chinook salmon (*Oncorhynchus tshawytscha*). Aquaculture 528, 735539. https://doi.org/10.1016/J.AQUACULTURE.2020.735539

Zorriehzahra, M.J., Delshad, S.T., Adel, M., Tiwari, R., Karthik, K., Dhama, K., Lazado, C.C., 2016. Probiotics as beneficial microbes in aquaculture: an update on their multiple modes of action: a review. Vet. Q. 36, 228–241. https://doi.org/10.1080/01652176.2016.1172132

General discussion

Overall, the microbiome and transcriptome profiles of fish larvae and their association with larval development under hatchery condition is not well studied. The microbiome and trasncriptome profiles of fish larvae under hatchery conditions that were developed in the present thesis have revealed valuable insights into their association with core modifications occurring during early development. In aquaculture farms, where disease outbreaks and suboptimal larval quality are major concerns, the interaction between the host, bacteria, and the environment plays a pivotal role in determining the outcome of production. This study addresses all three factors, providing a foundation for improved disease management strategies, guiding future research directions and providing knowledge of direct relevance for production. The vast amounts of data generated through next-generation sequencing techniques serves as a crucial resource for identifying effective aquaculture practices and disease prevention efforts.

A comprehensive overview of the state of the art in relation to the microbiome and transcriptome (**Chapter 1**) during larval ontogeny was the basis for our large-scale transcriptome (**Chapter 2**) and microbiome studies (**Chapter 4**). The outcomes and applications of the knowledge generated with these high throughput approaches underscore the significance of omics tools in aquaculture. It also highlights the substantial modifications in microbiome and transcriptome profiles in a range of species and under different situations and offers insight into potential host-bacteria interactions.

Chapter 2 focuses on utilizing RNA-seq profiles to elucidate the transcriptional responses of larval development under hatchery conditions. This research identifies significant changes associated with age and weight and compares different species, namely gilthead sea bream and Senegalese sole. This broader approach provides insight into the core molecular processes involved in metamorphosis. In-depth analysis of the expression profiles of endocrine genes and differentially expressed immune genes provides valuable information. The study highlights the critical role of the endocrine system and specific endocrine factors in modulating the larval-juvenile transition. Additionally, the expression profiles of the endocrine genes and differentially expressed immune genes were analyzed in depth. The global importance of the endocrine system at different stages of development and endocrine factors that play the most significant role in the modulation of larval-juvenile transition were identified. The low expression of immune-related gene transcripts in the early stages of gilthead seabream and Senegalese sole suggests they had a poorly established immune response in very early larval stages. Indicating that vaccination or attempts to stimulate the immune system in early larval

stages will be counterproductive. Furthermore, the results obtained suggest significant activation of the complement system in older larvae, although if the system is functional and activated by immunostimulants in early larval stages needs to be established. Similarly, pattern recognition receptors (PRR) were abundant at mid-metamorphosis but had a low abundance in early developmental stages. The suppression of immune-associated GOs and KEGG pathways, including antigen processing and presentation and lysosome, further underscore that the gilthead seabream (21 and 46 dph) and Senegalese sole (9 dph) larvae are not fully immunocompetent at very early stages. Nonetheless, since functional challenge studies were not carried out caution is needed since the status of immune responsiveness was based on the results of transcriptome studies and not animal trials.

The next step after the transcriptome and functional profile analysis involved targeting genes related to processes or pathways significantly modified, which be candidate biomarkers of immune status. Given the pivotal role of complement proteins in innate immunity and the significant modifications in the abundance of gene transcripts for this system during ontogeny, the complement system was further characterized. A comprehensive analysis and comparative evolutionary assessment of C3/C4/C5 and the regulatory factor cfh in deuterostomes with specific attention to fish was performed (**Chapter 3**). This analysis revealed different isoforms of complement proteins and their potential functional diversity. Specific primers for each isoform were designed for gilthead sea bream and revealed the tissue specific expression pattern of complement 3 gene expression. This knowledge will be crucial for assessing the response of this system in future studies in which larvae or fish are challenged with pathogens as it may be a useful biomarker in skin, as suggested in Chapter 3, that gives insight into immune status and teleost immune physiology.

Notably, this study represents a pioneering effort to examine the microbiome profile of fish larvae under hatchery conditions on a large scale. It comprehensively analyzes the microbiome at three critical stages (egg, first-feeding, mid-metamorphosis), providing insight into the biological, ecological, and rearing factors the influence microbiome dynamics during early development (**Chapter 4**). Profiling the microbiome identifies potential risks in marine hatcheries, such as *Vibrio*, and underscores the substantial influence of factors like hatchery location, species, diet, and age on the microbiota. These outputs suggest a robust and comprehensive disease management strategy for hatcheries needs to take into consideration all the factors that determine bacterial communities. Egg disinfection and UV treatment of water are essential practices for preventing disease outbreaks during the hatchery phase. However,

the study suggests a necessity to optimize disinfection usage and UV treatments (**Chapter 4.1, Chapter 4.3**) or to adopt complementary approaches to modulate the hatchery holobiont. It emphasizes that achieving a completely sterile environment is impractical, emphasizing the importance of controlling bacteria through methods that promote beneficial bacteria with antagonistic effects on pathogens. Additionally, like the transcriptome analysis, profiling bacteria serves as an initial step in exploiting NGS techniques, but it should be followed up with approaches for further tracking potential pathogens or beneficial bacteria or to associate the enrichment of specific microbiota pathways with larval performance. Cost effective and rapid exploitation of the microbiome data involved the design of genus-specific primers for the development of rapid and cost-effective monitoring of bacterial genera in hatcheries (**Chapter 4.2**).

Concluding remarks and recommendations for executive and research action

1- Microbiome and host transcriptomic responses during early fish development: the investigation of microbiome and host transcriptomic responses during the early development of fish revealed dynamic profiles in both microbiota and transcriptomes. This overall output suggests there is potential for the modulation of microbiome profiles and transcriptome responses, by exploiting probiotics to enhance larval performance and increase their capacity in later stages.

2-Transcriptome analysis for larval performance enhancement: transcriptome analysis of fish larvae reared in different production sites identified core gene sets that undergo modifications during larval development. These gene sets were associated with the weight and age of the larvae and can serve as applicable markers future in trials to assess and enhance larval development and performance. The approach use based on multiple larvae from several different hatcheries in different geographic location specified core transcriptional responses irrespective of the management techniques, food, broodstock etc. and minimized “noise” that results from insufficient samples collected from an insufficient number of experimental groups.

3- Immune responses in early development: analysis of the transcriptome revealed low levels of immune-related gene transcripts in very early development stages, including immune genes associated with the complement system and pathogen recognition receptors. This funding, coupled with the observation that most immune genes exhibit expression in both early and late developmental stages, but with differing abundance suggests that early stages of fish larvae are not sufficiently immunocompetent to counter high pathogen loads. This underscores the need

for alternative approaches to immunomodulators during early development unless they serve to upregulate the immune response and competence.

4- Diversity in complement system: an in-depth analysis of the complement system in fish has identified different isoforms of complement genes, indicating structural diversity exists. These differences in sequence and protein domains of deduced C3 proteins suggest functional diversity associated with structure and may be linked to the high diversity of the microbiota in which fish are in contact. The diverse C3 and Cfh in teleosts underscores an enhanced capacity to activate complement through direct interaction of C3 isoforms with pathogenic agents.

5- Microbiota in fish eggs: profiling the microbiota of fish eggs and evaluating different variables provided a baseline to increase egg quality and the potential of egg colonizing bacteria so they persist into later stages. Commonly used iodine-based disinfectants as they are currently used have a minimal impact on the bacterial load and composition of the eggs. Given these findings and the influence of site and species on egg bacterial communities, site-specific disease management and the need for species- and site-specific optimization of disinfection protocols are recommended.

6- Factors influencing microbiota establishment and microbiome modulation: several factors influence microbiota establishment during larval development, including a significant impact of the geographical location of the fish hatchery, water microbiota, and feeding regime. Hence, it is imperative to take these factors into account when developing protocols for the modulation and management of larval microbial communities in future studies. Additionally, this study profiles the microbiome of live foods, such as microalgae and *Artemia*, which could serve as a vehicle of microbiota modulation. Notably, the higher abundance of *Vibrio* in *Artemia* and rotifer than algae and water suggest a higher likelihood of the transfer to larvae of potentially pathogenic *Vibrio* species by the formers live feeds, irrespective of geographic location. The output of this investigation suggests a strong host selection of beneficial bacteria to create a mutually beneficial relationships in very early larvae, particularly at the onset of feeding. Accordingly, *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, and *Blautia* were identified as potential future probiotics to enhance larval performance.

7- Pathogens in hatcheries: potential pathogens such as *Vibrio* are abundant and a part of the core microbiome profile across hatcheries. The enrichment of biofilm formation of pathogens with age and various antibiotic resistance pathways are risks for the increase of disease outbreaks and decreased larval quality during larval development. These outputs of the thesis work highlight that current rearing procedures may need further optimization, including re-

evaluation of feeding regimes and disease management practices in European sea bass and gilthead sea bream hatcheries in the Mediterranean.

8- Interactions between microbiota and host biological pathways: transcriptome and microbiome profiles were significantly modified in parallel during larval development, suggesting potential interactions between bacteria function and host biological responses. The association between host and bacteria included biological pathways involved in muscle development, the thyroid and GH-IGF1 axes, energy homeostasis, gastrointestinal maturation, immune response, and tissue homeostasis. Research is needed that will target microbiome and meta transcriptomes to better relate and integrate the outcome of microbiome and transcriptome studies. Nonetheless, the pathways identified are suggested as targets to decipher and further study the impact of microbiota on host development in future studies.

In summary, the data indicate that while there are core and common modifications in both transcriptome and microbiome profiles across different hatcheries, variations in larvae performance are driven by rearing practices in hatcheries, including feeding regimes and water management, and the geographical site. The latter are factors that vary from hatchery to hatchery and may explain differences in quality. The output of this thesis highlights the importance of not only adopting optimal rearing protocols but also considering factors such as the biological capacity of larvae and the risks associated with the microbiota and antibiotic resistance, which can lead to disease outbreaks and persistence in some hatcheries, ultimately impacting their sustainability and production costs.