MELINA COELHO DA SILVA

EVOLUTION OF DESMOSOMAL GENES IN TELEOST FISH AND EXPRESSION ANALYSIS OF DESMOSOMAL CADHERINS IN THE GILTHEAD SEA BREAM (*Sparus aurata*)

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologias

2016

MELINA COELHO DA SILVA

EVOLUTION OF DESMOSOMAL GENES IN TELEOST FISH AND EXPRESSION ANALYSIS OF DESMOSOMAL CADHERINS IN THE GILTHEAD SEA BREAM (*Sparus aurata*)

Mestrado em Biotecnologia

Trabalho efetuado sob orientação de:

Dr.ª Deborah M Power

Dr.ª Katerina A Moutou

Co-orientador Dr. João CR Cardoso

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologias

2016

EVOLUTION OF DESMOSOMAL GENES IN TELEOST FISH AND EXPRESSION ANALYSIS OF DESMOSOMAL CADHERINS IN THE GILTHEAD SEA BREAM (*Sparus aurata*)

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.

Assinatura:_____

Acknowledgments/Agradecimentos

First of all, I would like to thank Professor Deborah Power for the opportunity to embark on this project and for her guidance.

Secondly, to Professor Katerina Moutou for the hospitality and constant support on this project. Also a big ευχαριστώ to everybody from the Laboratory of Genetics, Comparative and Evolutionary Biology from the University of Thessaly.

Às minhas coleguinhas, Ana L. e Catarina, tudo de bom para vocês, está?

Um obrigada à minha família e amigos, com destaque especial à minha mãe, à minha prima Élodie e às minhas amigas Andreia, Diana e Margarida.

Quero agradecer também à boa companhia e ajuda do pessoal do laboratório de Endocrinologia Comparativa e Biologia Integrativa do CCMAR. Obrigada à Rita Costa pela ajuda com a estatística e por fornecer algumas amostras.

Por fim, mas definitivamente não menos importante, quero agradecer ao meu co-orientador Professor João Cardoso, pela paciência, dedicação, correções preciosas e conhecimentos partilhados.

The work received funds from FCT – Foundation for Science and Technology under Pluriannual UID/Multi/04326/2013.

"We ourselves feel that what we are doing is just a drop in the ocean. But the ocean would be less because of that missing drop." Mother Teresa

Abstract

Fish aquaculture is a viable and profitable industry worldwide and monitoring fish larvae health and growth is a key step. The gilthead sea bream (*Sparus aurata*) is an economically important species reared mainly under mesocosm and intensive aquaculture systems. However, heterogeneous larvae growth and anatomic deformations are observed using these systems and few molecular markers to monitor fish growth are currently in use.

Desmosomes are a type of cell-cell junction formed by a protein complex that promotes strong connections between cells, playing a crucial role in the maintenance of tissue integrity. Desmosomal genes are well studied in human but poorly in fish, thus their potential as an additional molecular marker to monitor fish growth and health in aquaculture remains to be explored.

The aim of this project was to study the desmosomal genes evolution in fish and characterize their potential involvement in fish development and physiology in aquaculture. The results showed that desmosomal genes largely expanded in fish when compared to other vertebrates and teleosts possess a single desmocollin (*dsc*); three desmogleins (*dsg a*,*b* and *c*); two plakoglobins (*pg a* and *b*); two desmoplakins (*dp a* and *b*) and seven plakophilins (*pkp 1a*, *1b*, *2*, *3a*, *3b*, *4a* and *4b*). *Dsc* and *dsg b* were the candidates and their expression in *Sparus aurata* was analyzed in various experiments. In general their expression in larvae was not affected by the aquaculture systems. However, these genes seem to be associated with growth, in larvae exhibiting heterogeneous growth, the smaller individuals have the *dsg b* expression significantly up-regulated ($p < 0.05$) compared to the *dsc*. Thus these genes have potential to be used as molecular markers of growth. Different *dsc* profile expressions were found among the experiments, revealing that genetic background may also influence the expression of this gene.

Keywords: desmosomes, evolution, teleosts, expression, gilthead sea bream, growth.

Resumo

A aquacultura tem crescido vastamente ao longo das últimas décadas e emergido como uma alternativa às práticas pesqueiras. Os peixes são o principal organismo produzido e atualmente a sua produção mundial é aproximadamente 160 milhões de toneladas por ano. Consoante a densidade larvar a aquacultura pode ser classificada em diferentes categorias: extensiva, mesocosmo e intensiva. A dourada (*Sparus aurata*) é uma espécie bastante importante a nível económico no Mediterrâneo cuja produção em aquacultura tem registado um crescimento exponencial desde o início dos anos 90 até à atualidade. A sua produção é essencialmente sob regime intensivo e mesocosmo. Porém, estudos revelaram que esta espécie produzida em sistemas intensivos exibe deformações morfológicas e taxas de sobrevivência larvares menores quando comparadas com o sistema mesocosmo. Assim a seleção do sistema de aquacultura mais apropriado representa um desafio devido ao impacto que estes sistemas têm na saúde e crescimento dos indivíduos. Outro dos maiores desafios associado aquacultura é o crescimento não homogéneo das larvas de peixe. A origem deste problema permanece desconhecido e provoca perdas económicas elevadas devido às agressões e comportamentos canibalescos dos indivíduos maiores sobre os menores.

As formas clássicas de avaliar o crescimento larvar baseiam-se na caracterização do músculo, executando-se colorações por forma a contar as fibras e as dimensões das mesmas. Novas formas de avaliação têm vindo a emergir, recentemente o gene *mlc2* foi descrito como um marcador molecular de performance de crescimento de larvas de dourada em aquacultura.

Os desmossomas são um tipo de junção celular constituídos por um vasto complexo de proteínas que promove ligações fortes entre as células. Em humanos estão envolvidos em vários processos, tais como: proliferação, diferenciação e morfogénese. A disrupção destas estruturas tem efeitos drásticos na integridade celular, desta forma o seu mau emparelhamento tem vindo a ser associado a uma serie de doenças. Estruturalmente estas junções celulares são constituídas por proteínas de três famílias: caderinas (desmocolinas e desmogleínas), armadilho (plocoglobinas e placofilinas), e plaquinas (desmoplaquinas). A arquitetura dos desmossomas baseia-se na ligação heterofílica das proteínas transmembranares (desmocolinas e desmogleínas) na zona extracelular das células, as suas

caudas intracelulares por sua vez ligam-se às proteínas armadilo que interagem com as desmoplaquinas que finalizam a conexão da estrutura com a rede de filamentos intermédios das células. O papel destas proteínas é vastamente estudado em mamíferos, porém em peixes um único estudo foi realizado. Os resultados do silenciamento das caderinas desmossomais provocou fenótipos severos no peixe zebra e redução da integridade celular, revelando assim também um papel de extrema importância nos peixes.

Este projeto focou-se na biotecnologia azul e em termos gerais pretendeu-se enriquecer o conhecimento sobre este grupo de genes, desvendar como evoluíram nos peixes teleósteos e correlacionar com o seu desenvolvimento e crescimento, utilizando como modelo a dourada devido ao seu elevado interesse comercial. A nível de metodologia o projeto foi dividido em duas partes:

a) Análises bioinformáticas – identificação *in silico* de genes desmossomais em peixes e outros vertebrados para execução de análises filogenéticas;
b) Técnicas de biologia molecular – avaliação da expressão de caderinas desmossomais durante o desenvolvimento larvar, em larvas cultivadas em diferentes sistemas de aquacultura (intensivo e mesocosmo) e em larvas que exibissem crescimentos heterogéneos, por forma a correlacionar estes genes com os principais problemas da aquacultura acima referidos, utilizando q-RT-PCR.

Foram encontrados homólogos dos genes desmossomais humanos em várias espécies. No geral foram identificados em teleósteos: uma desmocolina (*dsc*); três desmogleínas (*dsg a*,*b* e *c*); duas placoglobinas (*pg a* e *b*); duas desmoplaquinas (*dp a* e *b*); sete placofilinas (*pkp 1a*, *1b*, *2*, *3a*, *3b*, *4a* e *4b*). Os genes desmossomais apresentam trajetórias evolutivas distintas de espécie para espécie, tendo sido identificadas duplicações específicas de espécies que sugerem ter um papel funcional especifico em cada espécie, potencialmente associado à sua adaptação ao meio. Devido ao seu papel crucial na formação da estrutura desmossomal e da sua evolução complexa, as caderinas desmossomais foram selecionadas pra se efetuar os estudos de expressão. A *dsc, dsg b* e *dsg c* foram isoladas na dourada, permanecendo o duplicado *a* por ser isolado e sendo o *c* uma nova descoberta alcançada neste projeto. A expressão tecidular em conjunto com uma pesquisa em bases de dados de EST revelou que a *dsc* é amplamente expressa enquanto as *dsg* apresentam uma distribuição mais restrita. Em humano, diferentes combinações destas proteínas foram descritas consoante o tipo de célula e camada celular, o mesmo parece se

verificar nos teleósteos. A *dsc* e *dsg b* foram identificadas em larvas enquanto a *dsg c* parece estar ausente nesta fase, o que indica também uma combinação das caderinas desmossomais associada à fase desenvolvimento dos peixes. Análises de expressão de *dsc* e *dsg b* em larvas que apresentava crescimento heterogéneo, revelou que os indivíduos menores expressam níveis maiores de *dsg b* e esta é significativamente sobre expressa em comparação à *dsc* ($p < 0.05$). Correlações entre estes genes e genes associados ao desenvolvimento do músculo foram observadas (*myog, igf-2* e *fst*). Após eclodirem, as larvas apresentam um período intenso de hiperplasia (recrutamento de fibras de músculo) que ocorre entre os 15-25 dias após eclosão e posteriormente um período de hipertrofia (aumento do tamanho das fibras existentes). Curiosamente, no geral foi no período de hiperplasia que níveis mais elevados de *dsc* e *dsg b* foram observados. Isto indica que a expressão destes genes está possivelmente mais associado ao recrutamento de fibras musculares do que ao aumento das mesmas, justificando assim o facto dos indivíduos menores apresentarem uma no geral níveis maiores de *dsg b* e uma sobre expressão desta em relação à *dsc*. Estes resultados sugerem que estes genes poderão servir como marcadores moleculares de crescimento em larvas de peixe em aquacultura. No geral os sistemas de aquacultura não parecem influenciar a diferença de expressão de *dsc* e *dsg b,* o que sugere que o tipo de sistema de aquacultura não tem impacto na integridade celular neste período (5-60 dias após ecolosão). Os resultados mais surpreendentes foram as diferenças de expressão de *dsc* obtidas nas diferentes experiencias, sugerindo que fatores genéticos podem influenciar fortemente a expressão destes genes visto que que as amostras eram provenientes de *stocks* diferentes.

Palavras-chave: desmossomas, evolução, teleósteos, expressão, dourada, crescimento.

Abbreviations

1R: first whole genome duplication

2R: second whole genome duplication

3R: third whole genome duplication

aa: amino acid

ANOVA: analysis of variance

aqp4: aquaporin 4

ATP: adenosine triphosphate

BLAST: Basic Local Alignment Search Tool

BME: β-mercaptoethanol

bp: base pair

BrEt: ethidium bromide

b4galt6: β-1,4-galactosyltransferas

Ca^{2+}: calcium ion

CAR: cell adhesion recognition site

cDNA: complementary deoxyribonucleic acid

cdh2: cadherin 2

cm: centimeters

CM: cytoplasmic membrane

col1a1: collagen type I alpha 1

Ct: cycle threshold

CTNNB1: human β-catenin 1 protein

CTNND1: human δ-catenin 1 protein

DNA: deoxyribonucleic acid

DP: desmoplakin

dph: days post hatch

Dsc: desmocollin

Dsg: desmoglein

DTD: desmoglein terminal domain

EA: extracellular anchor

EC: extracellular cadherin repeat

E. coli: Escherichia coli

ef1a: elongation factor 1-alpha

EST: expressed sequence tag

FAO: Food and Agriculture Organization

fst: follistatin

Fw: forward

g: gram

GSR: glycine–serine–arginine

HCMR: Hellenic Center for Marine Research

HKG: housekeeping gene

IA: intracellular anchor

ICS: cadherin-like sequence

IDP: inner dense plaque

IF: intermediate filament

Ig: immunoglobulin

Igf-2: insulin-like growth factor 2

Ind/L: individual per liter

IPL: proline rich linker region

IPTG: isopropyl β-D-1-thiogalactopyranoside

kb: kilobase

kctd1: potassium channel tetramerization domain containing 1

kg/m^3: kilogram per cubic meter

kg/ha/year: kilogram per hectare per year

L^{-1}: per liter

LB: Luria-Bertani

M: molar

m: meter

m^3: cubic meter

MCS: multiple cloning site

mg/ml: milligram per milliliter

Mg^{2+}: magnesium ion

min: minute

ml: milliliter

mlc2: myosin light chain 2

mm: millimeters

mrf4: myogenic regulatory factor 4

mst: myostatin

mt-DNA: mitocondrial deoxyribonucleic acid

MYA: million years ago

myog: myogenin

NCBI: National Center for Biotechnology Information

NF: normalization factor

Ng: nanogram

n.i. : not identified

nm: nanometers

nmol/L: nanomol per liter

NSRF: National Strategic Reference Framework

ODP: outer dense plaque

P: p value

PCR: polymerase chain reaction

PG: plakoglobin

pH: hydrogen potential

PKP: plakophilin

PLC: hepatocellular carcinoma-derived cell line

PLEC: human plectin protein

PM: plasmatic membrane

psma8: proteasome subunit alpha 8

PRD: plakin repeat domain

q-RT-PCR: quantitative real time polymerase chain reaction

R_0: starting fluorescence

RAPD: random amplified polymorphic DNA

RNA: ribonucleic acid

RPL13A: ribosomal protein L13a

rpm: rotations per minute

rps18: gene of ribosomal protein S18

RUD: repeat unit domain

Rv: reverse

SDS: sodium lauryl sulfate

SEACASE: sustainable extensive and semi-intensive coastal aquaculture in Southern Europe

sec: second

S.E.M: standard error of the mean

SNP: single nucleotide polymorphism

ss18: nBAF chromatin remodeling complex subunit

TAE: Tris-Acetate-EDTA

taf4b: gene of TATA-box binding protein associated factor 4b

Tm°: melting temperature

TSWGD: teleost-specific whole genome duplication

trappc8: gene of trafficking protein particle complex 8

ttr: transthyretin

tubb2: tubulin β-2A class IIa

U: unit

UV: ultraviolet

U/µl: unity per microliter

V: volt

WGD: whole genome duplication

X-gal: 5-bromo-4-chloro-3-indolyl-ß-D-galactoside

znf521: gene of zinc finger protein 521

°C- Celsius degrees

λ: wave length

µg: microgram

µl: microliter

µg/mL: microgram per milliliter

µg/µL: microgram per microliter

µm: micrometers

µM: micromolar

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Aquaculture an expanding industry worldwide

Aquaculture is defined as the large scale husbandry or rearing of aquatic organisms for commercial purposes and is a viable and profitable industry worldwide [1]. The origins of aquaculture as a form of farming trace back to more than 2000 years when the Roman and Chinese Empires were the pioneers [2]. In the last few decades, aquaculture has become a global practice and is one of the most rapidly evolving and technically innovative sectors of food production, with a significant investment, scientific input, technical development and production [3]. Aquaculture has been emerging as an alternative to fisheries that imposes a significant pressure on fish stocks reducing average age and body size, which in return reduces the percentage of the fish to spawn and maintain the natural stock density [4].

According to the Food and Agriculture Organization (FAO) it is estimated that more than 600 aquatic organisms are cultured and this includes several fish species, crustaceans, molluscs, amphibians, reptiles, aquatic invertebrates and also plants and algae [5]. Most aquaculture products are for human consumption, although they can have other applications, such as extraction of bioactive compounds from algae for the cosmetic and pharmaceutical industries. A small proportion of the aquaculture products are used to produce fishmeal and to extract fish oils for use in the manufacturing of fish diets with high protein content for aquaculture [5].

Fish are one of the main aquaculture products and they are the most traded food commodities in the world. Over the last sixty years, fish production in aquaculture has grown vastly (Figure 1.1), increasing at an average annual growth rate of 6.1 percent. In 1950 approximately 20 million tonnes of fish were produced in aquaculture, but nowadays the production has increased eight times and has almost reached 160 million tonnes per year [6]. In 2012, approximately 200 countries exported aquaculture products and these exports are especially important for developing nations as they represent more than half of the total value of traded commodities in the majority of the cases [5].

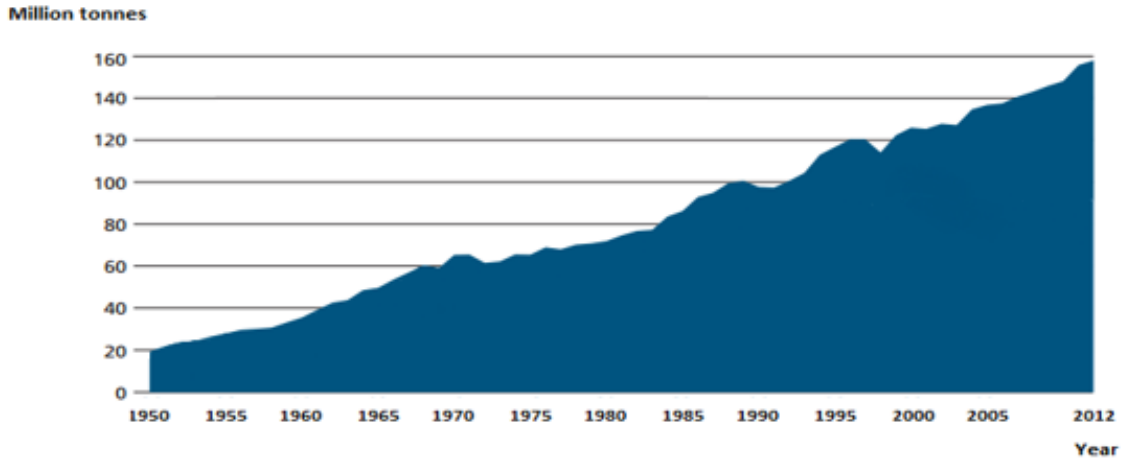


Figure 1.1: Global aquaculture fish production increment (in million tonnes) from 1950 until 2012. Data from FAO [6].

World fish consumption per capita increased from an average of 9.9 kg in 1960 to 19.2 kg in 2012. This increment was triggered by population growth, urbanization, expansion of fish production and the existence of efficient distribution channels [6]. Asia dominates the market and China is by far the largest global aquaculture producer and exporter [6,7]. In 2012, aquaculture provided almost half of the fish for human consumption and China, India and Vietnam were the top leading producers, with 41.1, 4.2 and 3.1 million tonnes of fish produced, respectively [5,8].

Fish is a rich source of protein and a 150 g portion provides about 50–60% of an adult human daily protein requirements. In densely populated countries where total protein intake levels are low, fish proteins represent a crucial nutritional component [6]. In 2050 it is expected that the global human population will reach 9.6 billion inhabitants with more than 800 million suffering from chronic malnutrition. This fact will challenge the capacity of the planet to feed the human population while safeguarding its natural resources for future generations. In this context, aquaculture plays a crucial role in eliminating hunger, promoting health and developing the economy in order to support the global development [3,5].

1.2 Fish farming systems

Aquaculture production of fish starts in hatcheries with the production of fry from broodstock. Fish hatchery systems may follow different approaches, but a decisive parameter is the fish larvae stocking density. Accordingly, hatchery systems are defined as: extensive, mesocosms, semi-intensive, intensive and hyper intensive (Figure 1.2) [9]. Despite this classification, commonly three main categories are defined: extensive, mesocosms and intensive systems that are described in more detail below [9].

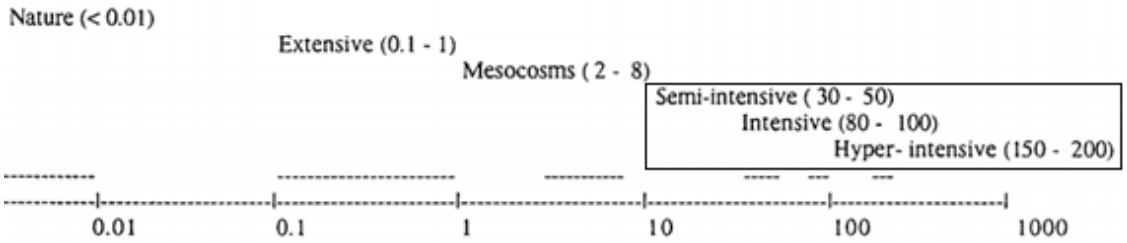


Figure 1.2: Classification of the aquaculture systems according to larval density. Values expressed in number of larvae L^{-1}. Larvae stocking density is indicated in brackets [9]. Boxed are the three types that constitute the intensive system.

Other parameters that can be considered are the prey source and availability (wild or cultured; phytoplankton or zooplankton; live or inert) and quality of water and hydroid system (clean water, green water, pseudo-green water, open or closed water circuit) [9]. Table 1.1 summarizes the main differences between the aquaculture systems.

Table 1.1: Main differences between extensive, mesocosm and intensive aquaculture techniques [9].

Parameters	Techniques		
	Extensive	**Mesocosm**	**Intensive***
Rearing enclosures	Ponds or bags	Tanks or bags	Tanks
Localisation	Outdoor	Indoor **	Indoor
Rearing volume (m^3)	>100	30-100	<20
Rearing density (ind/l)	0.1-1	2-8	30-200
Food chain	Endogenous	Mixed	Exogenous
Infrastructures	Light	Medium	Sophisticated
Environment	Natural	Mixed	Controlled
Autonomy and autarky	High	Medium	Low to nil
Dependence on man and technique	Light	Medium	High to very high
Need for specific biological knowledge	Light	Medium	High to very high
Validity for new species	Very high	High	Medium to low

*Includes semi-intensive, intensive and hyper-intensive techniques
**Sometimes outdoor (with bags) or semi-outdoor

1.2.1 Extensive system

The extensive system is based on the natural migration of euryhaline fish (organisms that are able to adapt to a wide range of salinities) and typical fishing traps are used to capture juveniles of 2-3 g. The juveniles captured are seeded into lagoons and costal ponds where they complete growth and reproduction [10]. Extensive systems also operate for fish eggs and larvae, in this case, feeding is based on phytoplankton and zooplankton that promote a natural environment, creating a food chain that provide the necessary nutrients for fish development [9]. Larviculture under this system is performed at low densities (0.1-1 larvae L^{-1}) and usually is carried out in small pouds or even bags [9].

In this type of farming the fish density generally does not exceed 0.0025 kg/m^3 and the output is less than 1000kg/ha/year [10]. This system is highly dependent on the environment conditions and this together with the lower outputs are the main disadvantages. An advantage of extensive system is that is a highly environmental-friendly practice [9].

1.2.2 Mesocosm system

The mesocosm system sits between extensive and semi-intensive systems (Figure 1.2), thus it is also denominated as semi-extensive [9]. The goal of the mesocosm system is to mimic the natural environment using both extensive and intensive techniques without their disadvantages [11]. This may include enrichment of the farming area with oxygen supplementation and exogenous feed, consequently requiring extra costs in more advanced technologies. In mesocosm system, the individuals obtain most of their food from the natural environment, yet they can also receive supplementary feed, which allow fish to grow faster than in the extensive system [10]. This system is based on a natural phytoplankton or zooplankton bloom that is stimulated prior to the yolk-sac larvae (larvae already hatched from the egg but not feeding yet and still absorb the yolk that contain nutritive compounds) and also water fertilization to stimulate algae growth [11].

This system can be subdivided according to the source and quality of the food in mesocosms with extensive and intensive philosophies. In the extensive philosophy, fish use natural source of food (phytoplankton or zooplankton) and ocasionally receive an exogenous input to complement the diet. In the intensive philosophy, the source of food is the natural food chain derived from the phytoplankton and zooplankton and supplementation with exogenous inputs always occur [9].

Larviculture in these systems is performed at relatively low densities (2-8 larvae L^{-1}) in relatively large (30-100 m^3) and deep (1.5-2.5 m) tanks, usually with circular shape [9]. Mesocosm systems are used throughout the world, yet there is no standard protocol that can be followed, although high-value marine food can be produced in such systems [11]. However it's the rearing methodology that exhibits improved results in relation to the production of high quality juveniles with lower ecological footprint compared with the intensive system. The final production varies a lot, according to the size of the juveniles stocked and the amount of feed provided. Normally the density in mesocosm systems does not exceed 1 kg/m^3 and the output is around 10000kg/ha/year [12].

The white sea bream, sea bass and the gilthead sea bream are the three main species that perform the best in this aquaculture system when compared with intensive or extensive techniques. These species when in mesocosm systems exhibit

homogenous behavior, normal wild coloration and low levels of deformities (1-5% in sea bream and white sea bream and 2-3% in sea bream) [9].

Sea bream larvae in mesocosm system exhibit higher growth and also higher rates of larval survival than in intensive systems. Anatomic deformities in the vertebral column, anal and caudal fins of some species in intensive systems were also observed [13]. Other advantage of this system is the environmental stability that prevents drastic changes in the water quality due to the larger volumes used and its lower economic impact when compared with intensive system. The main disadvantage is the inexistence of a protocol that can be applied to different species in different regions, since nutritional and environmental requirements varies from species to species and phytoplankton and zooplankton organisms vary widely from place to place [11].

1.2.3 Intensive systems

Intensive systems include the semi-intensive, intensive and hyper-intensive systems (Figure 1.2). These aquaculture systems are the most sophisticated and are highly dependent on skilled personnel and technology [9]. They are characterized by high larval densities in small tanks under strict specific hydraulic, thermal, light and feed conditions, which accelerate growth and fish biomass. Under this systems high rates of deformities and other abnormalities, such as alterations in the coloration, cannibalism and abnormal sex ratio, are observed [9].

These systems were implemented during the 1980s and consist of four phases: reproduction, larval rearing, fattening and grow-out. In intensive farming there is a brood stock that initiates the reproduction and then larval rearing phase occurs, followed by the fattening phase and terminating with the growing-out. There is no wild capture and the reproductive phase assures the existence of a new generation, thus gonad development, spawning and egg quality are significantly influenced by brood stock robustness [14]. The larval rearing phase is performed under highly controlled and optimized conditions in order to guarantee a good development of the fish. Fattening is very important and takes place in land-based installations. There is a pre-fattening phase applied to fry and an intensive fattening phase applied to juveniles. The growing phase may occur in land-based installations and also in sea cages, either

in sheltered or semi-exposed sites (floating cages) or totally exposed sites (semi-submersible or submersible cages) [14].

1.3 The gilthead sea bream and its farming

The gilthead sea bream (*Sparus aurata*) is an economically important species in European aquaculture and an intensive increment in its production is evident over the recent years (Figure 1.3). Since the early 90's a continuous exponential growth in sea bream production occurred, starting from less than 100 tonnes in 1986 to more than 150,000 tonnes in 2013 worldwide (Figure 1.3) [14].

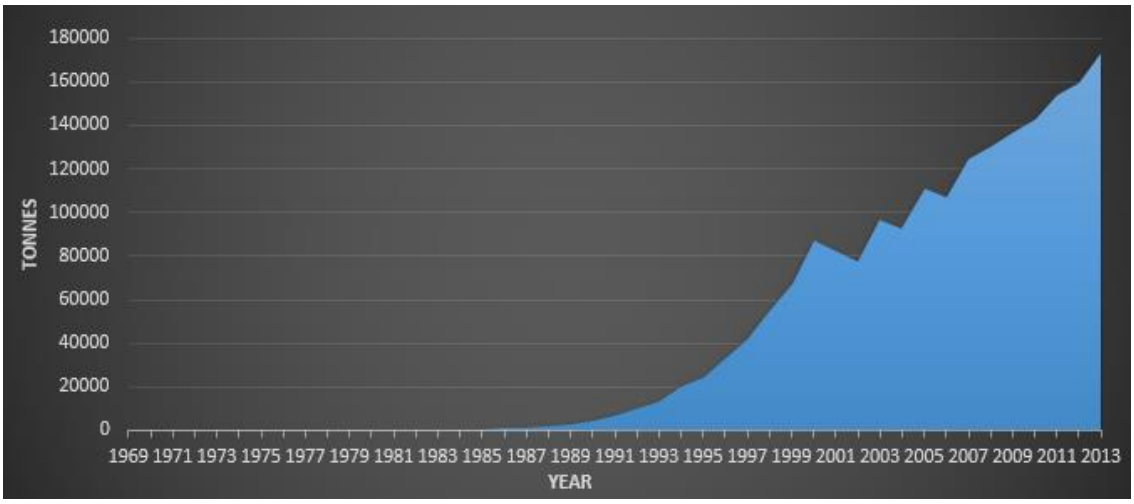


Figure 1.3**:** Increase growth in global aquaculture production of gilthead sea bream from 1969 to 2013. Adapted from FAO - Fisheries and Aquaculture Statistic 2013 [14].

Gilthead sea bream can be farmed in coastal ponds and lagoons with extensive and mesocosm systems or in land-based installations and in sea cages with intensive farming systems. In extensive systems gilthead sea bream reach the first commercial size (350-400 g) in approximately 20 months while in the intensive system it occurs in only 15 months [10].

Gilthead sea bream is a subtropical carnivorous teleost fish of the *Sparidae* family and is very common in the Mediterranean Sea, its presence extends from the Eastern Atlantic coasts of the Great Britain to Senegal in Northern Africa (Figure 1.4) [10].

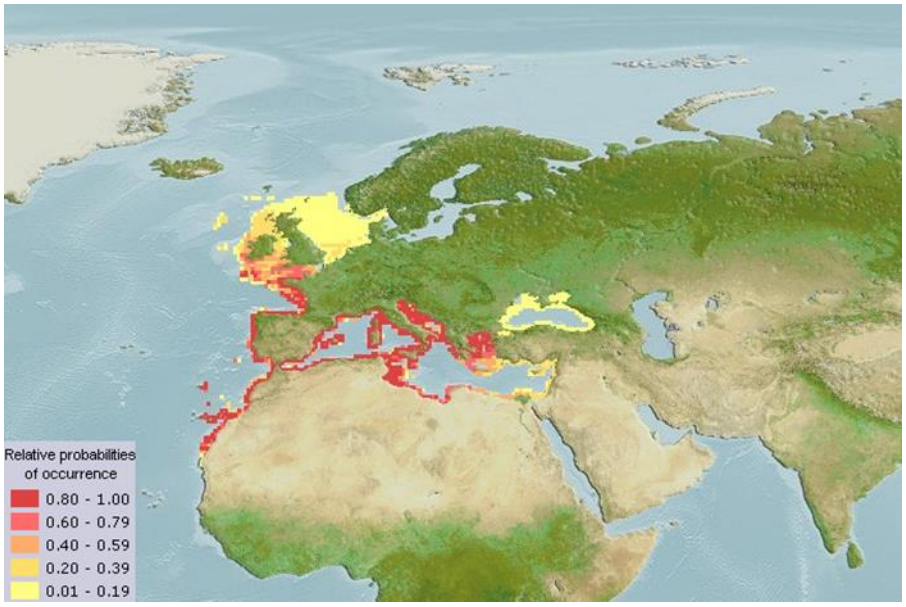


Figure 1.4: World distribution of the gilthead sea bream. Colours indicate the relative probability of gilthead sea bream occurrence in the area [15].

The average length of an adult sea bream is 35 cm with a maximum length of 70 cm. It is regularly curved with small eyes and the mouth has thick lips. The body shape is oval and its colour is silvery grey with a large black blotch at the origin of the lateral line extending from the upper margin of the opercula. On the body sides dark longitudinal lines are often present and the dorsal fin has a dark band while the tips of caudal fin are edged with black (Figure 1.5) [17,10].

Figure 1.5: Photograph of an adult gilthead sea bream [10].

This species is a protandrous hermaphrodite (organisms that develop into males first, then possibly to females), in the wild sexual maturity in males is achieved at 2 years of age (20-30 cm) and in females at 2-3 years (33-40 cm). In captivity sex reversal can be conditioned by social factors and hormones [10]. The breeding season

under natural environmental conditions occurs in late Autumn, eggs size are 0.9-1.1 mm and their incubation lasts about 2 days at 16-17°C [16]. Larval length at hatching is 2.5-3.0 mm and metamorphosis occurs around 50 days post hatch (dph) at 17.5°C or about 43 days post hatch at 20°C [16]. Moretti et al characterized the ontogeny of larval development at 17-18 °C in the intensive system and the principal characteristics during this stage are described in Table 1.2 [17].

Table 1.2: Developmental stages of gilthead sea bream larvae at 17-18 °C [17].

Day	Size (mm)	Characteristics
1	3.0	Hatching
2	3.5	Pectoral fins appear
3	3.8	Exotrophy starts
4	3.9	Eyes pigmented 60% of yolk sac absorbed, 40% of oil drop absorbed
5	4.0	Primary swim bladder inflation 100% of the yolk sac absorbed, 70% of oil drop absorbed
15	5.0	End of the primary swim bladder inflation 100% of the oil drop reabsorb, caudal fin
17	7.0	Anal fin
20	7.5	Stomach starts developing
45	11.0	Second dorsal fin
50	15.0	First dorsal and ventral fin
60-70	20.0	Scales
90	30.0	Definite morphology

Since this species is very sensitive to low temperatures, 4°C is their lethal temperature limit and juveniles typically migrate in early Spring to areas where they can find milder temperatures [10]. Due to its euryhaline (capacity to tolerate a wide range of salt water concentrations) and eurythermal habits, this species is found in both marine and brackish water environments [10].

1.4 Bottlenecks of fish farming and molecular markers

Selection of the aquaculture system may impact on fish health and may affect fish growth performance [19,20]. Different growth and larval survival rates are observed according to the type of aquaculture system selected, however the reasons behind this remain unclear and constitute a major problem in fish aquaculture [13]. The non-homogeneous larvae growth and consequently the rearing of juveniles of uneven sizes, increases aggressive behavior within a population provoking losses in the culture and this has a negative impact in the aquaculture production. In the wild, the

size-related dominancy promotes faster growth of the larger dominants since they obtain preferential access to food [20]. The cause of non-homogeneous growth in larvae remains to be explored and it may be a genetic consequence or a social event effect. Several exogenous factors, such as temperature, salinity, light intensity, food availability and quality also have an impact on growth performance [21]. To avoid non-homogenous growth in aquaculture a size grading occurs at regular intervals by separating larvae according to the size and each group of larvae is cultured in different tanks [15,23]. This prevents cannibalistic behavior and consequently improves growth, survival rates and biomass gain of the cohort reducing economic losses [23,20]. However, size sorting increases stress and is an extra cost for the producer [18].

Available methodologies to assess larvae growth and quality are based on the characterization of the muscle structure using histology. Staining of muscle sections (with haematoxylin and eosin) and counting of muscle fibers and their dimensions have been used extensively to determine the growth dynamics of fast muscle fibers in several aquaculture species [23]. However, the development of technology and the availability of sequenced genomes and transcriptomes for many fish species will permit the identification of candidate regulatory genes involved to monitor larval growth. This would prevent in the future the costs associated with the production of individual with low growth potential [19].

Recently, molecular markers based on the fish genetic traits have been developed and applied to aquaculture for genetic identification and discrimination of aquaculture stocks (mt-DNA or nuclear DNA such as microsatellites, SNP or RAPD) and to compare hatcheries and wild fish stocks [24]. However, few markers linked with aquaculture fish performance and production have been identified. Recently a marker for fish muscle growth, myosin light chain 2 protein (mlc2), a structural protein of the muscle, has been described and is a good candidate gene for assessing gilthead sea bream growth performance [19]. The development and growth of muscle comprises two fundamental processes: hyperplasia (the recruitment of new fibers) and hypertrophy (increase in size of existing fibers) and in gilthead sea bream *mlc2a* is highly expressed during hyperplasia while *mlc2b* is up-regulated during hypertrophy [25].

1.5 The desmosomes

Multicellular organisms have four types of cell-cell junctions: 1) gap junctions, intercellular channels that permit the free passage between the cells of ions and small molecules; 2) tight junctions, that regulate the passage of molecules and ions through the space between cells; 3) adherens junctions, that provide strong mechanical attachments between adjacent cells; and 4) desmosomes, localized patches that hold two cells tightly together [26]. Desmosomes (derived from the greek "desmos" meaning bond and "soma" meaning body [27]) are multiprotein complexes assembled in the plasma membrane and provide a connection between intermediate filaments of the cell cytoskeletons of adjacent cells, giving strength to tissues [28]. They are important in the maintenance of tissue architecture and form a transcellular web that allows cells to resist to mechanical stress [29]. In humans they are involved in fundamental processes such as cell proliferation, differentiation and morphogenesis [30].

Disruption of desmosomal structure has devastating consequences for tissue integrity and can lead to death [27] and there is a correlation between desmosomal diseases and aberrant cell signalling [31]. Several human diseases have been associated with functional impairment of desmosomes. As expected, skin and its appendages, heart and mucous membranes are the most affected, since desmosomal defects affect tissues and organs subjected do mechanical stress [29].

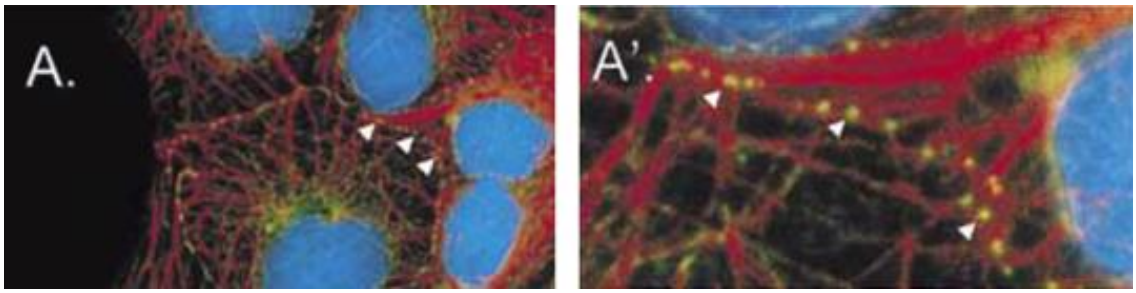


Figure 1.6: Immunofluorescence microscopy photograph of an epithelial cell line (PLC, hepatocellular carcinoma-derived cell line) stained with antibodies against desmoplakin (yellow) and keratin intermediate filaments (red). The cell nuclei are shown in blue. Arrowheads point to individual desmosomes that are lined up along the plasma membranes of adjacent cells. A': Higher magnification of cell-cell contacts shown in A [29].

In other vertebrates, the gene repertoire and importance of the desmosomes in animal physiology remain largely unknown and in teleosts, that represent the largest and successful group of vertebrates, there is a sole study in the zebrafish. The desmosomal proteins were knocked down leading to tissue integrity reduction and the animal developed severe phenotypes such as shortened body axis, severely reduced or absent head and or tail, absence of clearly defined somites and sometimes blebbing of the epidermis [18]. This suggests that as in humans, these protein complexes also play an important role in tissue development and animal growth, however they remain to be explored and their importance in the maintenance of tissue integrity and potential usefulness as an additional molecular marker to monitor fish growth in aquaculture remains to be explored.

1.5.1 Desmosomes structure

Desmosomes are classically “spot welds” and their disc-shaped-like structures are highly organized and resistant to dissolution, pH extremes and most detergents [26]. They are dynamic cell structures of size between 0.1 to 0.5 μm and can be assembled and disassembled in response to signals from the micro-environment that are essential to allow morphogenic process (Figure 1.6) [28]. They are abundant in cells derived from the ectodermal lineages [31], and present in all epithelial cell and a few non-epithelial cell types, including myocardial and Purkinje fiber cells, meningeal cells and follicular dendritic cells of lymph nodes [29]. This type of cell junction is composed by three major protein families: cadherins, armadillo and plakins [32] (Figure 1.7).

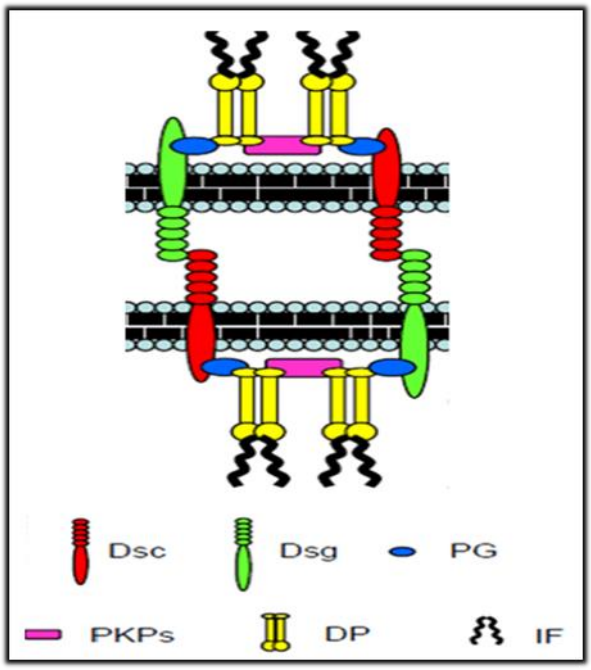


Figure 1.7: Structure of a desmosome. The three major protein families that are responsible for maintaining the desmosome structure are represented: the cadherin family members: Dsc-Desmocollin (red) and Dsg-Desmoglein (green); the armadillo family: PG-Plakoglobin (blue) and PKPs-Plakophilins (pink); plakin family: DP-Desmoplakin (yellow) that links to intermediate filaments (IF). Adapted from [28].

This structure may vary according to the cell type and layer and also during embryonic and post-embryonic development. For example during epidermal differentiation, smaller and less-organized desmosomes in the basal layer are replaced by larger and more electrondense desmosomes [29,32]. Structurally, desmosomes can be divided into three parts: desmoglea or extracellular core, inner dense plaque (IDP) and outer dense plaque (ODP) [33]. To form this structure, each cell provides “half” of the adhesion complex [34]. The N-terminal domains of Dsc and Dsg are located in the extracellular core, the apparent intercellular space between the two “half-desmosomes” of neighbouring cells, of approximately 34nm of thickness (Figure 1.8) [29,33,35]. The C-terminal of Dsc and Dsg are located in the ODP, which is 15-20 nm thick. ODP consists of the Dsc and Dsg tails interacting with plaque proteins, PKP, DP and PG. DP binds to IF within the IDP, serving to tether the IF to the plasma membrane (Figure 1.8) [35].

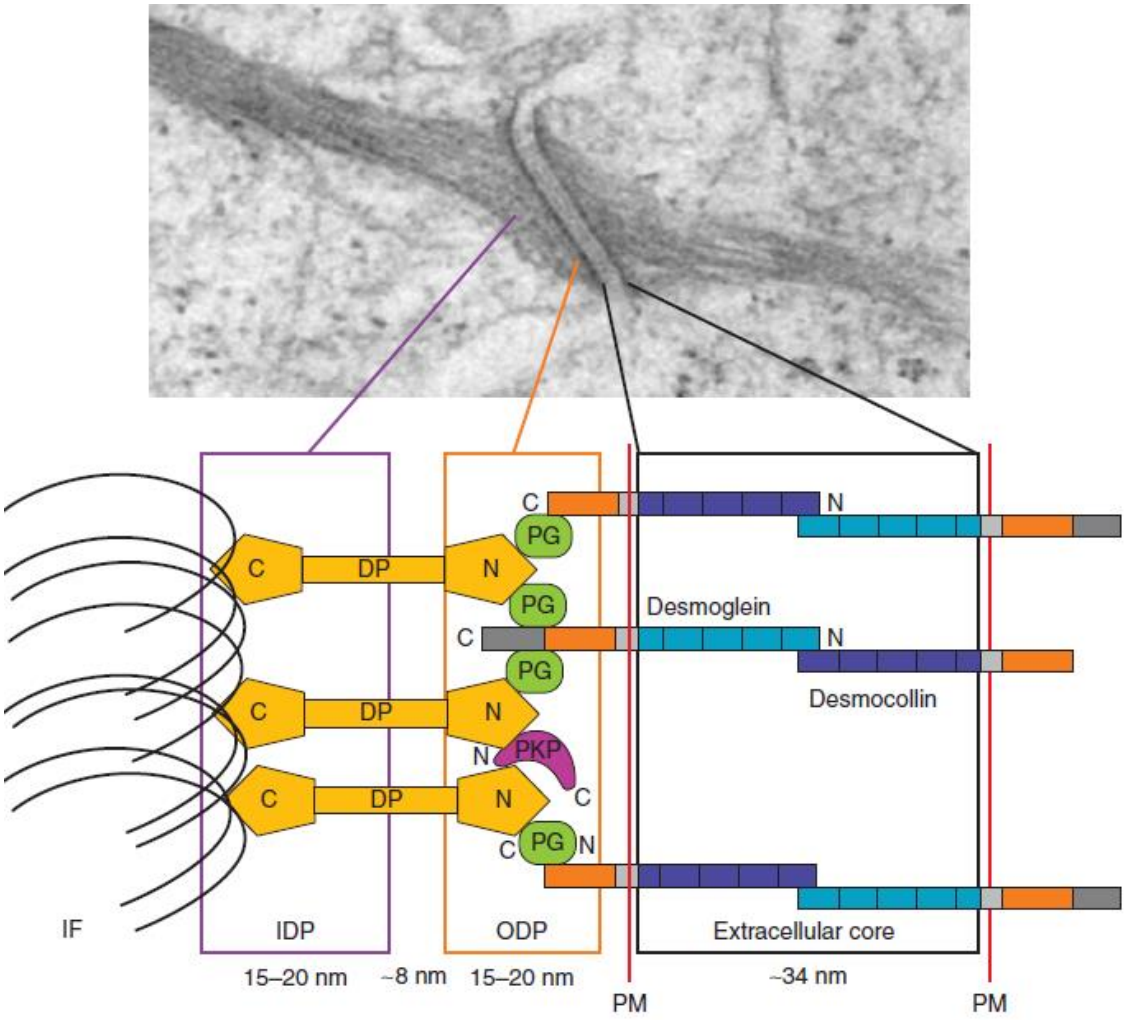


Figure 1.8: Structural model of a desmosome. (A) Electron micrograph of a desmosome. (B) Schematic representation of the structural organization of a desmosome. PM- plasmatic membrane; IDP- inner dense plaque; ODP - outer dense plaque; PG-plakoglobin; DP- desmoplakin; PKP- plakophilins; IF- intermediate filaments; C – c-terminal and N- n-terminal [35].

1.5.2 The desmosomal cadherin family members

Desmogleins (Dsg) and the desmocollins (Dsc) belong to a subfamily of the cadherin superfamily. They are transmembrane proteins that bind heterophically and mediate calcium-dependent cell-cell adhesion [36]. In cultured cells, desmosomal adhesion can be initiated or disrupted by raising or lowering the extracellular calcium concentration, although, *in vivo* the extracellular Ca^{2+} concentration is assumed to be always well above the concentration that is required to regulate desmosomes (+/- 0.1 mM) [36].

Dsc and Dsg structures share four extracellular cadherin repeats (EC1-4) and an extracellular anchor (EA) region, that form Ig-like globular domains with calcium

binding sites between each pair of consecutive repeats (Figure 1.9) [37]. These cadherin sequence repeats are approximately 110 aa in length and the first cadherin repeat (EC1) contains the cell adhesion recognition site (CAR) that possesses a central alanine residue responsible for the adhesive function between desmosomal cadherins. Both Dsc and Dsg also contain an intracellular anchor (IA) and a cadherin-like sequence (ICS), which is also conserved across the other cadherin members. The main difference between Dsc and Dsg is that the latter possesses additional unique sequences at the C-terminal region with yet unknown functions and this includes a proline rich linker region (IPL), a repeat unit domain (RUD) and a Dsg terminal domain (DTD) (Figure 1.9) [37]. In vertebrates, desmosomal cadherins are single transmembrane proteins that are encoded as precursor proteins containing a signal sequence and a prodomain that immediately precede EC1 and are removed by proteolysis [37].

In human, four Dsg (Dsg1-4) and three Dsc genes (Dsc1-3) have been identified and each of the Dsc transcripts exhibit alternative splicing of the cytoplasmic domain giving rise to a longer ‘a’ form and a shorter ‘b’ form (Figure 1.9) [32]. In the “b” form the region encoding the ICS domain is truncated and terminates with an additional 11 aa in Dsc 1 and 2, and eight residues in Dsc 3, not found in the “a” form [38].

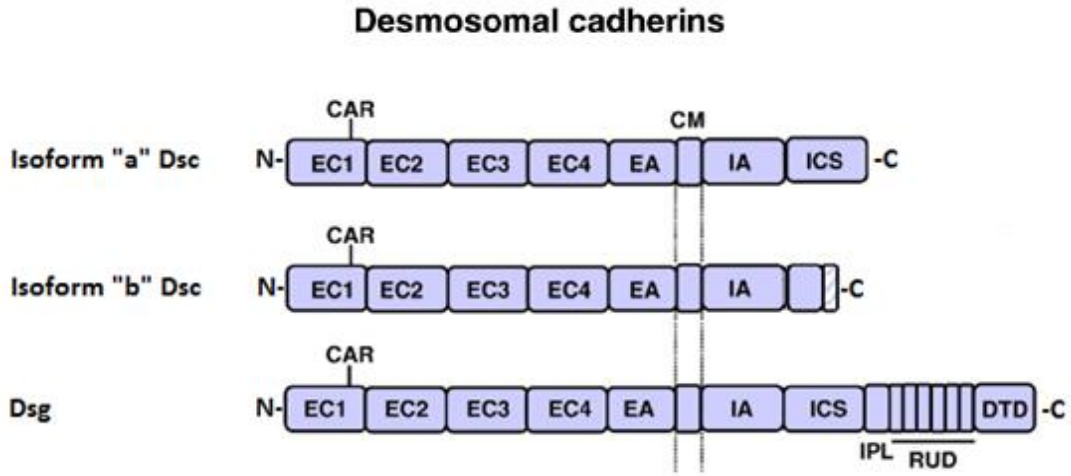


Figure 1.9: Schematic representation of desmosomal cadherins, Dsc and Dsg. In human CAR - cell adhesion recognition; EC’s - extracellular cadherin repeats; EA- Extracellular anchor; CM- cytoplasmic membrane; IA- intracellular anchor; ICS- intracellular cadherin-like sequence; IPL-intracellular proline-rich linker; RUD-repeat unit domain; DTD- desmoglein terminal domain [30].

In human skin, depending on the cell type and cellular layer there are different combinations of Dsg and Dsc and simple epithelia express only the Dsg2/Dsc2, but stratified complex epithelia, such as the epidermis, express primarily Dsc1/3 and Dsg1/3 with low levels of Dsg2/Dsc2 in the basal layers and Dsg4 concentrated in the granular and cornified layers (Figure 1.10) [36].

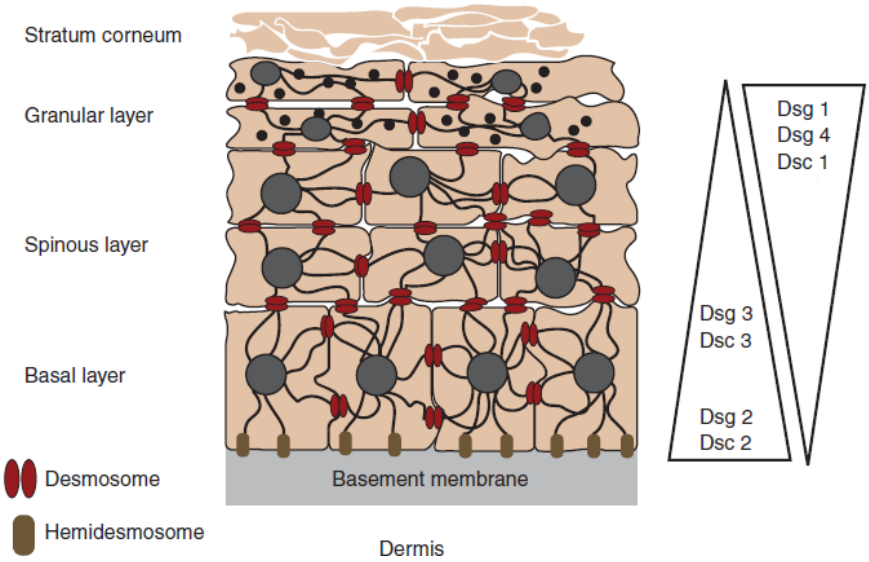


Figure 1.10: Expression patterns of the desmosomal cadherins in the four layers of the epidermis in human [35].

The differentially expressed desmosomal cadherins are not fully understood, but the highly patterned distribution of different adhesion molecules may ensure maintenance of cell relationships during morphogenesis of multilayered tissues [32]. In the desmosome, the desmosomal cadherins bind directly to PG, also known as γ-catenin [30]. Dsc, Dsg and PG together in the absence of other desmosomal components can generate adhesion [36].

A wide range of studies have revealed that desmosomal cadherins are implicated in human diseases, especially in disorders associated with heart and the skin diseases (Table 1.3) [35].

Table 1.3: Tissue distribution of desmosomal cadherins in human and associated diseases [38].

Subfamily	Members	Expression	Disease
Desmogleins	Dsg 1	Stratified epithelia	Pemphigus (vulgaris, foliaceus, paraneoplastic) Striate palmoplantar keratoderma Bullous impetigo Staphylococcal scaled skin syndrome
	Dsg 2	Simple and stratified epithelia, myocardium	Arrythmogenic right ventricular cardiomyopathy Respiratory and urinary adenovirus infection
	Dsg 3	Stratified epithelia	Pemphigus (vulgaris, paraneoplastic)
	Dsg 4	Outermost epidermis, hair follicle cells	Localized recessive hypotrichosis +/- recessive moniletrix
Desmocollins	Dsc 1	Epidermis, hair follicle cells	IgA Pemphigus
	Dsc 2	Simple and stratified epithelia, myocardium	Arrythmogenic right ventricular cardiomyopathy palmoplantar keratoderma and wooly hair
	Dsc 3	Stratified epithelia	Pemphigus vulgaris Hypotrichosis with recurrent skin vesicles

In other vertebrates the existence of dsc and dsg are poorly explored and their role remains unknown. In zebrafish, a teleost model organism, a single *dsc* (*zfdsc*) has been isolated and the transcript has similar size to the mammalian "a" form of Dsc 1 with which it shares 68% of aa homology [18]. Two *dsg* (*zfdsgα* and *zfdsgβ*) have been also identified and they are expressed throughout the development of zebrafish and play an important role in the early embryo morphogenesis [18]. Zfdsgα exhibits 64% homology with the human Dsg 2 but data related to zfdsgβ is not available since the authors only isolated a fragment of the gene. *Zfdsc* and *zfdsgα* are expressed in egg and *zfdsgβ* starts to be expressed at 2.25 hours post fertilization. Knockdown experiments revealed that these proteins have an important role in embryo development and epiboly, gastrulation, convergence-extension movements and structure of desmosomes are affected. Zfdsc protein precursor comprises of a 16 aa signal sequence that is preceded by a 105 aa pre protein and the mature protein has 771 aa, the CAR region in human is YAT but in zebrafish is RAF. ZfDsgα has 40 extra aa, glycine-rich insert in the extracellular domain when compared to zfDsgβ. ZfDsgα comprises of a 16 aa signal sequence followed by a pre-protein composed of 21 aa, the CAR site in the is IAL rather than YAL in the human [18].

1.5.3 The desmosomal armadillo family members

PG and PKP are the armadilo family representatives in desmosomes. These proteins interact to the cytoplasmic tails of desmosomal cadherins and also to DP, which in turn links the desmosome to the IFs network of the cell (Figure 1.8) [29]. PG is a component of both adherens junctions and desmosomes and it has been suggested that cells expressing PG with C-terminal truncations have modified desmosomes. This deletion provokes the formation of large desmosomes, thus the C- terminus limits the size of desmosomes [39]. Beyond the structural function, PG is also found in the cell nucleus, where it plays a role as a gene expression regulator [29]. It was found that PG can control the expression levels of desmosomal genes, revealing an autoregulation mechanism [31]. Knockout studies in mice also identified a critical role for PG in desmosome assembly *in vivo* and PG null animals die because of fragility of the myocardium [40].

Armadillo proteins are characterized by a series of repeat motifs, designated as arm repeats (Figure 1.11) [41]. The central domain of PG comprises a highly conserved series of arms repeats that are involved in its association with Dsc and Dsg [39]. The interaction of armadillo proteins appears to be mediated by overlapping regions of the central arm repeats [39]. Structurally, PG contains 12 arm repeats of 42 aa each [39,43] and repeats (1– 4) are required for Dsg binding, whereas Dsc binding requires both ends of the arm domain [43].

PKP belong to the p120-catenin subfamily that shares a conserved central domain composed of 9 arm repeats, in contrast to the 12 in PG and β-catenin, and are flanked by a N- and C-terminal regions that diverge from one another (Figure 1.11) [44]. It has been proposed that PKP recruit DP to the desmosomal junctions and the lateral interaction between these two proteins extend the size of the desmosome [45].

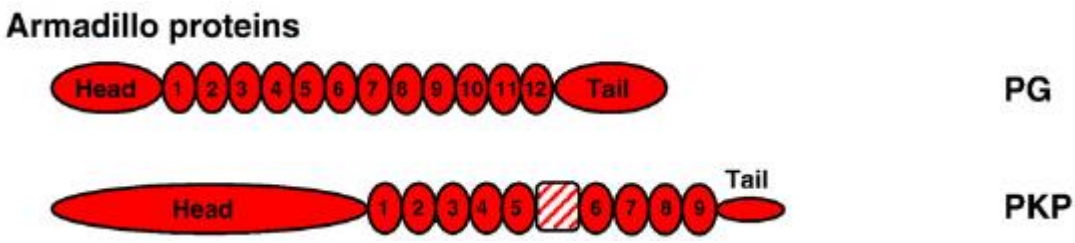


Figure 1.11: Schematic representation of the armadillo proteins (PG and PKP). Hatched box represent the insert between repeats 5 and 6 that is only present in PKP [30].

PKP has a complex expression patterns as desmosomal cadherins and in humans four *PKP* (*PKP1-4*) that are differentially expressed in simple and stratified epithelia, cardiomyocytes, endothelia, and other cell types exist [41]. All PKP have been localized in desmosomes [41]. PKP1 is restricted to stratified and complex epithelia and urothelium and mutations in this gene are related to heart diseases [47,48]. From the PKP, PKP1 is the smallest [48] and PKP2 the largest protein [49]. PKP 2 is present in the basal cells of certain stratified epithelia [47]. Mutations in PKP 2 and PG in human are related to arrhythmogenic right ventricular cardiomyopathy, also known as Naxos disease [50]. In humans, PKP1 and PKP2 each have two isoforms, a shorter "a" form and a longer "b" form generated by alternative splicing [41]. PKP1 enhances predominantly lateral interactions between DP proteins [49] and PKP1a have been reported to be more common in desmosomes than b form [51]. PKP1b have an insertion of 21 aa between arm repeats 3 and 4 [51] and PKP2b have an insertion of 44 aa acids between repeats 2 and 3 [49]. In contrast to PKP1 and PKP2, no alternative isoforms of PKP 3 exist, this protein is present in the desmosomes of all cell layers of stratified epithelia and all simple epithelia, with the exception of hepatocytes [47]. Studies revealed that PKP3 null animals are viable, however defects of morphogenesis and morphology of specific hair follicles were detected, elucidating the role of PKP3 in the development or maintenance of skin appendages [47]. In general, PKPs higher expression have been correlated with tumors [47]. PKP4 is also called p0071 and is a bit different from the other PKP, is more closely related to other members of the p120 subfamily than to the other PKP. Members of p120 subfamily are known to be present in adherent junctions and to interact with classical cadherins. However this PKP has been described as a protein with dual localization depending on the cell type it is present in adherent junctions and desmosomes [52].This protein is involved in recruiting of the proteins to plaques [53].

1.5.4 The desmosomal plakins family members

Desmoplakin (DP) is the only plakin protein present in desmosomes and is the most abundant in this cell junction [35]. This protein interacts with keratin intermediate filaments in epithelial cells, desmin intermediate filaments in cardiomyocytes and vimentin intermediate filaments in arachnoid and follicular

dendritic cells [30]. DP is a tripartite protein with globular head and tail domains flanking a coiled-coil rod region (Figure 1.12). The globular head is a region of protein–protein interactions [54] and the tail consist of three plakin repeat domains (PRDs), designated A, B and C. Each PRD contains 4.5 copies of a 38 aa motif [55], the C-terminus has a glycine–serine–arginine (GSR) rich domain and it was suggested that the phosphorylation of serine residue may regulate the ability of DP to interact with intermediate filaments [56]. In mammals two DP isoforms (I, II), generated by alternative splicing, have been identified. They differ only in the length of the rod domain in the centre and it appears that they are functionally redundant (Figure 1.12) [30].

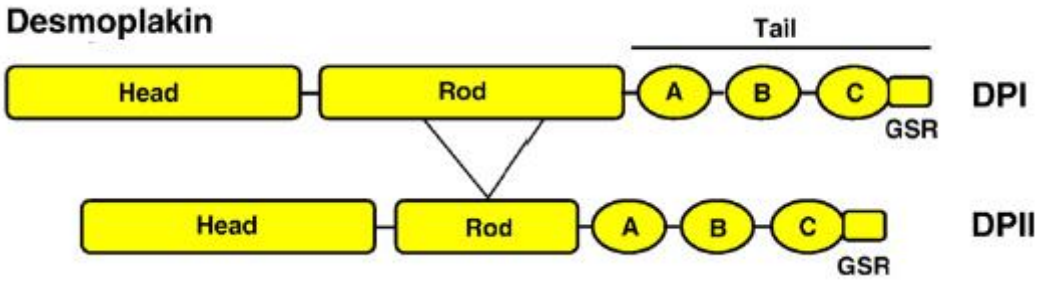


Figure 1.12: Schematic representation of DPI and DPII. A, B, C – plakin repeat domains; GSR- glycine–serine–arginine rich domain [30].

Both isoforms are widely expressed in many tissues, although DPII is absent from the heart and its expression is low in simple epithelia [57]. Mutations in this gene have also been associated with pathologies in humans, including malfunctions of the heart, skin and its appendages. Table 1.4 contain some examples of human diseases associated with mutations in DP.

Table 1.4: Disease associated with DP mutations in humans.

Mutation	Disease
Amino-terminal deletion. Autosomal dominant mutation resulting in DP haploinsufficiency	Striate palmoplantar keratoderma. No heart defects [59,60].
Frame shift mutation in carboxyl terminus. Autosomal recessive	Striate palmoplantar keratoderma, dilated left ventricular cardiomyopathy, woolly hair [60].
Compound heterozygosity for non-sense and missense mutations	Palmoplantar keratoderma more severe than with other mutations, some hair loss, nail defects. No heart defects [61].

1.6 Gene evolution in teleosts

Fishes are an extremely diverse group of vertebrates and they are divided in: jawless fishes (hagfishes, lampreys), cartilaginous fishes (sharks, rays) and bony fishes (coelacanth, lungfishes and ray-finned fishes) [62]. Teleosts are ray-finned fishes and they represent half of the vertebrate species, thus they are by far the most diverse vertebrate clade. This clade contains more than 23000 species, which populate a wide range of habitats around the world. Teleost fishes exhibit a huge biodiversity that is evident in their morphology, ecology and behaviour (Figure 1.13) [63].

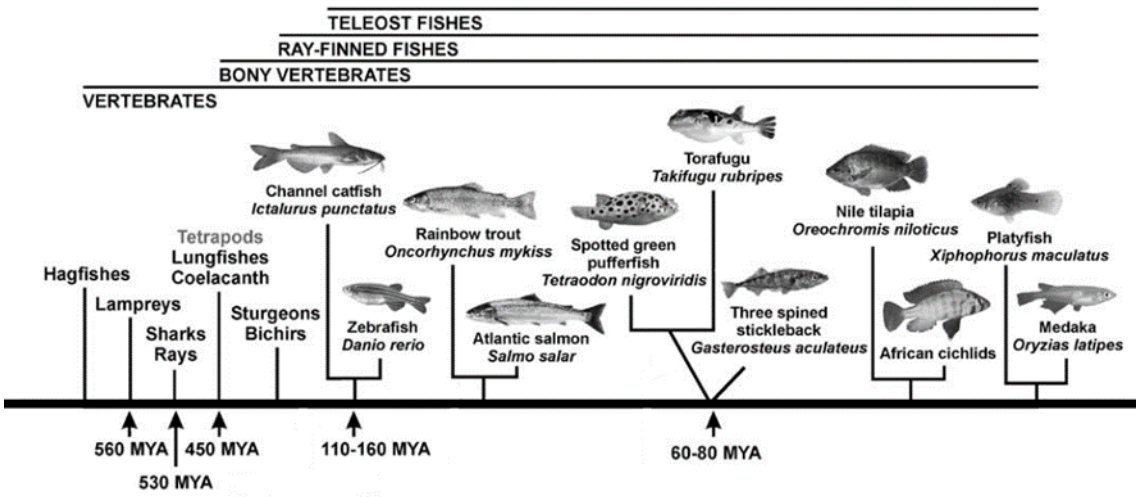


Figure 1.13: Evolution of fishes over the time, from hagfishes to teleosts radiation. MYA- million years ago [63].

The genomes of several species of fish have been sequenced and it has been suggested that genome duplication contributed to teleost diversity [63] . Genome specific gene duplications occurred in the teleosts relative to other vertebrates and the study of gene family evolution reveals teleost-specific duplicate genes [64]. Gene duplication events generate two genes, denominated paralogues that immediately after duplication are highly identical. However, during evolution duplicate genes modify and they may become non-functional (one paralogue disappears), subfunctional (both paralogues remain in the genome and share the ancestral gene functions), neofunctional (both paralogues remain in the genome and one acquires new functions) and dosage selection (both paralogues remain in the genome to not disrupt the relative gene dosage) (Figure 1.14) [65]. The teleost-specific genome duplication

(TSWGD) or 3R is known to have played an important role in physiological, morphological and behavioural diversification of this highly specious group [64].

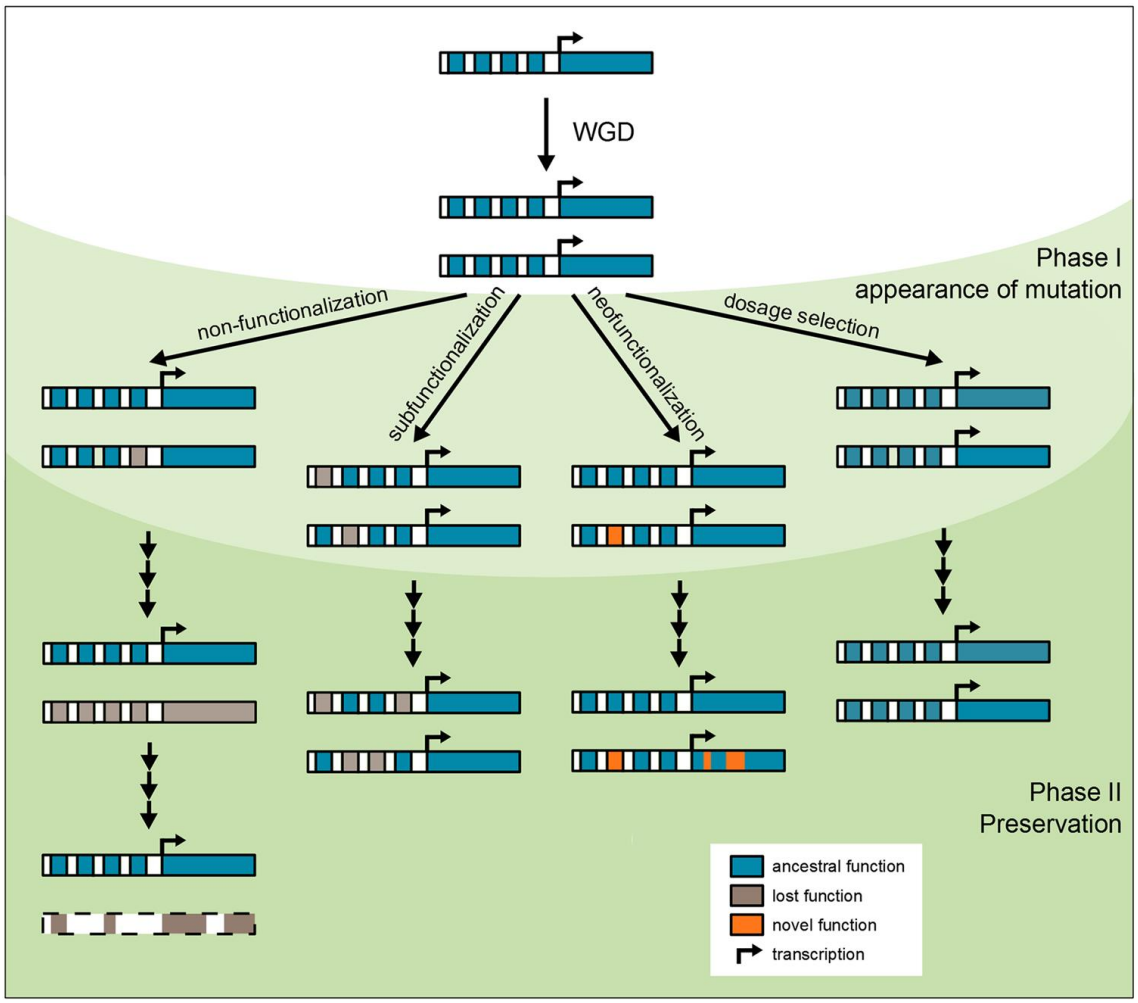


Figure 1.14: Fates of duplicated genes after WGD event. Non-functionalization: it's a very frequent fate in one of the paralogs since immediately after WGD paralogues are functionally redundant, suggesting that the selective constraint of maintaining both is low and that one of them is, therefore, free to disappear. Deleterious mutations occur in one of the paralogues, eventually leading to its silencing (pseudogenization). Mutations continue to accumulate until the structural features of the gene have totally disappeared. Subfunctionalization: genes usually have more than only one function and in this case complementary degenerative mutations in paralogous genes lead to preservation of both. Neofunctionalization: due to the lack of selective constraint on maintaining both paralogues, one of them is free to acquire mutations, eventually generating a new function. Dosage selection: after a WGD genes are doubled and therefore duplicate genes pairs are all expressed at a higher level than the corresponding ancestral gene, however the relative dosage is not disrupted. Maintain gene relative dosage balance is crucial to some genes and in one paralogue can lead to negative developmental or physiological consequences [66].

Prior to the TSWGD, metazoan evolution is proposed to have involved two other events of whole-genome duplication (WGD) (1R and 2R). The WGD is proposed to have occurred in the common ancestor of vertebrates prior to the divergence of jawless fish (< 500 MYA) when most of the gene families that are present in vertebrates are suggested to have emerged (Figure 1.15) [67].

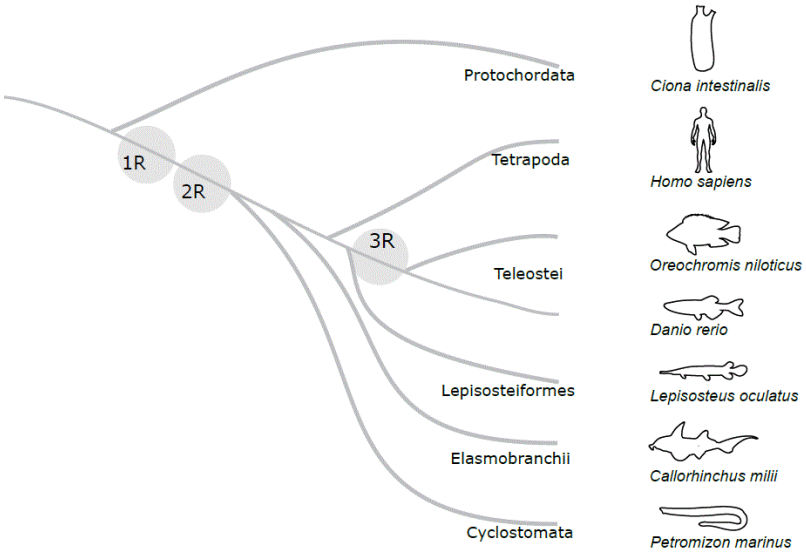


Figure 1.15**:** Proposed rounds of WGD during the vertebrate radiation: 1R and 2R occurred prior or at the emergence of the vertebrates and the 3R or teleost specific genome duplication that occurred only in the teleosts. Adapted from [68].

1.7 Context in biotechnology

The term “biotechnology” derives from the Greek: *bios* (βίος) - life; *technos* (τεχνης) – technology and *logos* (λόγος) – thinking [69] and aims to produce goods/services that are needed and can be provided with safety and at reasonable cost [70]. Biotechnology provides powerful methods for the sustainable development of agriculture, fisheries, forestry and a range of other industries to sustain growth and development [71]. Blue biotechnology is biotechnology based on marine and aquatic environments and it aims to apply molecular biological methods to marine and freshwater organisms to exploit unique molecules, unique biosynthetic processes and unique characteristics of biomaterials and to identify and combine traits in fish and shellfish to increase productivity and improve quality of aquaculture products. The recent investment in blue biotechnology makes aquaculture a growing field of animal research. There is a growing demand for aquaculture and biotechnology can help to meet this demand. Some examples of blue biotechnology are: transgenic aquatic

organisms, molecular diagnostic methods in aquatic organisms, cryopreservation of gametes and gene banking [72].

This project focuses in blue biotechnology and aims to explore the desmosomal genes in teleosts and characterise the desmosomal cadherins expression during larval development, in larvae reared in the two main aquaculture systems (intensive and mesocosm) and in larvae with heterogeneous growth. In mammals, desmosomal proteins are major cell structural and cell integrity proteins and are involved in the regulation of cell growth but they remain largely unknown in fish. This study will provide a deep comprehension of desmosomal genes evolution in fish and their potential involvement in larval development and physiology. The results obtained will enlarge our current knowledge on this family of proteins and on how they have evolved and adapted in fish and their potential association with fish development and growth.

1.8 Objectives

Gilthead sea bream (*Sparus aurata*) is one of the major commercial farmed fish species in the Mediterranean region with an increasing need for methods for monitoring quality and growth performance at all stages of production [15]. The objective of this thesis is to unveil the evolution of desmosomal genes in teleosts and characterise the desmosomal cadherins role in gilthead sea bream larvae, revealing their eventual potential as molecular markers in aquaculture.

The specific objectives were:

1. Identify desmosomal genes in teleosts and compare with other metazoans;
2. Characterize the desmosomal cadherins tissue distribution in gilthead sea bream;
3. Perform an ontogenic expression of desmosomal cadherins in gilthead sea bream larvae;
4. Investigate differences in the expression of desmosomal cadherins in gilthead sea bream larvae with heterogeneous growth;
5. Assess the impact of the aquaculture systems (mesocosm and intensive systems) in desmosomal cadherin expression in gilthead sea bream larvae.

2 Materials and methods

The work carried out in this study was divided in two main procedures: a) bioinformatics analysis and b) molecular biology techniques. The bioinformatics analysis aimed to identify *in silico* the desmosomal genes in fish and to perform a comparative analysis with other vertebrates. The molecular biology techniques were performed to characterize the expression of desmosomal cadherins in larvae of the marine teleost fish the gilthead sea bream at different developmental stages, to assess how different aquaculture systems affected their expression and to evaluate if a correlation with their expression exist with larvae exhibiting heterogeneous growth. The work plan carried out is summarized in Figure 2.1.

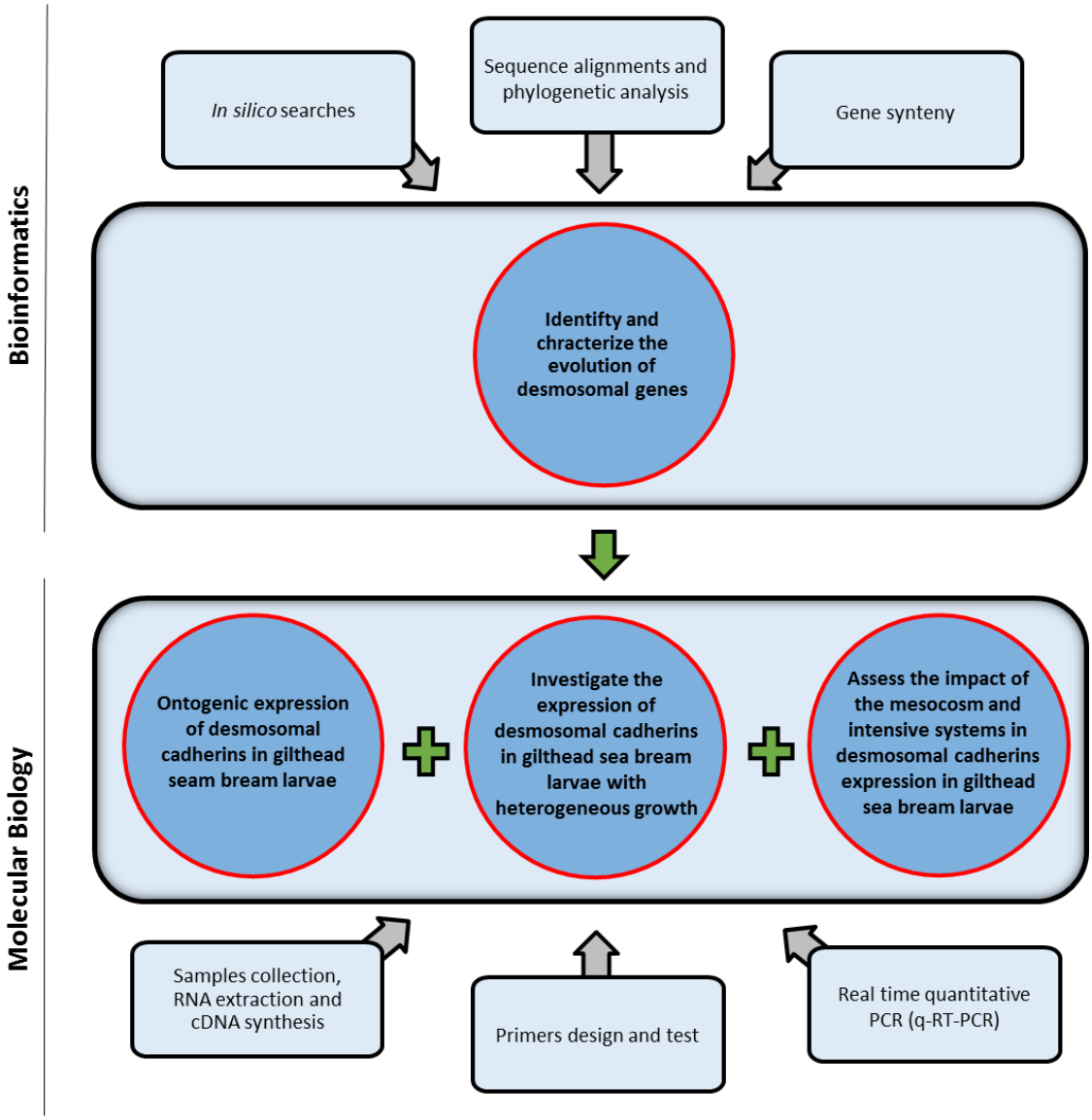


Figure 2.1: Flow chart of the methodology followed in this research project.

2.1 Biological material

2.1.1 Study 1: expression during ontogeny an in larvae with heterogeneous growth

Gilthead sea bream larvae used in the ontogeny study were collected from a commercial hatchery at Maliakos Gulf, Greece, as part of a project co-financed by Greek National Funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II, Investing in Knowledge Society. Gilthead sea bream larvae were sampled on 5, 15, 25, 35, 48 and 60 dph, as described by Georgiou, S., Alami-Durante, H., Power, D.M. et al [25]. Another set of larvae were also collected from the same commercial hatchery reared under the same conditions on and 58 dph. Larvae of 58 dph had been subjected to size sorting prior to sampling and the larger ones were collected separately from the smaller. Larger individuals had an average weight of 0.067g and 1.4-2.1 cm of length. Smaller individuals had an average weight of 0.029g and 1.3-1.8 cm of length. This study aimed to evaluate the expression of desmosomal cadherins during sea bream larvae development and to evaluate if the individual that exhibit different growth sizes possess different expression patterns of these genes.

2.1.2 Study 2: mesocosm and intensive systems

The samples used to compare the expression of desmosomal cadherins in different aquaculture systems (intensive and mesocosm) were obtained from the EU-funded research project “SEACASE — Sustainable extensive and semi-intensive coastal aquaculture in Southern Europe” and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) — Research Funding Program: Heracleitus II. Samples were collected from the Hellenic Center for Marine Research (HCMR, Heraklion, Crete, Greece). The larvae used to compare the effect of the aquaculture on gilthead sea bream desmosomal cadherin expression were cultured and sampled at 4, 15, 25, 45 and 81 dph, as described by Georgiou, S. et al [19]. This study aimed to assess about the impact in tissue integrity of larvae reared in different aquaculture systems.

2.2 Bioinformatic analysis

Teleost databases were interrogated to identify desmosomal genes and to characterize their evolution.

2.2.1 *In silico* searches

In silico searches were performed to identify the desmosomal genes (*dsc*, *dsg*, *pg*, *dp* and *pkp*) in fish by exploring different genome, transcript and expressed sequence tag (EST) databases. The deduced protein sequences of desmosomal genes of human (*Homo sapiens*, *DSC1* ENSG00000134765; *DSC2* ENSG00000134755; *DSC3* ENSG00000134762; *DSG1* ENSG00000134760; *DSG2* ENSG00000046604; *DSG3* ENSG00000134757; *DSG4* ENSG00000175065; *PG* ENSG00000173801; *DP* ENSG00000096696; *PKP1* ENSG00000081277; *PKP2* ENSG00000057294; *PKP3* ENSG00000184363; *PK4* ENSG00000144283) and of zebrafish (Danio renio, *dsc* ENSDARG00000039677; *pg* ENSDARG00000070787; *pkp2* ENSDARG00000023026; *pkp3* ENSDARG00000051861) were used as queries. The Actinopterygii genomes of the tilapia (*Oreochromis niloticus*); amazon moly (*Poecilia Formosa*); platyfish (*Xiphophorus maculatus*); tetraodon (*Tetraodon nigroviridis*); takifugu (*Takifugu rubripes*); stickleback (*Gasterosteus aculeatus*); cod (*Gadus morhua*); cave fish (*Astyanax mexicanus*), sea bass (*Dicentrarchus labrax*) and zebrafish (*Danio renio*) were searched. Searches were also extended to the Sarcopterigi fish genome the coelacanth (*Latimeria chalumnae*), the cartilaginous fish elephant shark (*Callorhinchus milii*) and the jawless fish the japanese lamprey (*Petromyzus marinus*). Other vertebrate genomes the amphibian xenopus (*Xenopus tropicalis*), the reptile anole lizard (*Anolis carolinensis*), the bird chicken (*Gallus gallus*) and the marsupial opossum (*Monodelphis domestica*) were also explored. The majority of the searches were performed in Ensembl (http://www.ensembl.org/) with the exception of the elephant shark (http://esharkgenome.imcb.a-star.edu.sg/) and sea bass (http://seabass.mpipz.de). Databases were accessed in November of 2015.

The searches for target transcripts in the gilthead sea bream were performed using an in house nucleotide database [73] (http://sea.ccmar.ualg.pt, accessed in November of 2015) using as queries the tilapia desmosomal deduced protein

sequences, as this species was found to exhibit a larger number of desmosomal genes than other teleost species. The nucleotide sequences found were translated to aa using ExPASy program (http://web.expasy.org). The identity of the retrieved sequences was confirmed by searching the NCBI protein database (http://www.ncbi.nlm.nih.gov/).

Searches in teleost ESTs databases were performed in order to identify possible new transcripts and identify target tissues where the desmosomal cadherins are expressed. Searches were performed using the NCBI database with desmosomal cadherins from zebrafish (*Danio renio*) as queries. The algorithm used was *tblastn* and the selected results had an expected value smaller than e^{-20}, results were verified by BLAST against all the species in NCBI in order to confirm identity.

2.2.2 Sequence alignments and phylogenetic analysis

The sequence alignments were performed for: a) primers design; b) classification of the *dsg* ESTs and the new *dsg* found in sea bream (nucleotide alignment); and c) for the phylogenetic trees construction (protein alignment). The alignments were executed in the Clustal W online program (http://www.genome.jp/tools/clustalw/).

The teleost ESTs were aligned with the three different *dsg* members of stickleback: *dsg a* (ENSGACG00000002160), *dsg b* (ENSGACG00000013161) and *dsg c* (ENSGACG00000002084) to investigate the expression and abundance of the different gene members. A similar strategy was used to classify the new *dsg* found in sea bream.

The deduced aa sequences of Dsc and Dsg, PG, DP and PKP from fish and other vertebrate were aligned and four phylogenetic trees were constructed. The Dsc and Dsg are highly related in sequence and were the candidate genes selected for the project. The sequence alignments were edited using the Aliview program to remove gaps. For the construction of the phylogenetic tree, the edited sequence alignment was subjected to an online statistical analysis to select the most suitable matrix for tree construction according to the statistical model Akaike Information Criterion (AIC) [74] ProtTest server 2.4 (http://darwin.uvigo.es/software/prottest2_server.html). Trees were built using the maximum likelihood method and branch statistical support was obtained using a bootstrap analysis of 100 replicates and constructed using the

PhyML available from the ATGC platform (http://www.atgc-montpellier.fr/phyml/). The Dsc and Dsg tree was built using the model WAG+I+G with gamma shape (4 rate categories) of 1.464 and 0.022 of proportion of invariable sites; PG was constructed with the JTT+I+G+F with 0.683 of gamma shape (4 rate categories and 0.226 of proportion of invariable sites and the human β-catenin 1 protein (CTNNB1) (ENSP00000344456.5) was used to root the tree; for the DP tree the model JTT+I+G+F with 1.545 of gamma shape (4 rate categories) and 0.083 of proportion of invariable sites was used and tree was rooted with the human plectin (PLEC) (ENSP00000434583); and the PKP tree was built according to the JTT+I+G+F with 1.596 of gamma shape (4 rate categories) and 0.014 of proportion of invariable sites and rooted with the human δ-catenin 1 protein (CTNND1) (ENSP00000436543).

2.2.3 Gene synteny analysis

The vertebrate desmosomal gene synteny was characterized according to the annotation of the Genomicus website (http://www.genomicus.biologie.ens.fr) and the elephant shark and sea bass gene environment were characterized blasting their genomes with the conserved neighbouring gene and synteny maps of the vertebrate *dsg* and *dcs* were constructed.

2.3 Molecular biology techniques

Molecular biology techniques were applied to isolate desmosomal genes, to follow their expression in gilthead sea bream development, in larvae with heterogeneous growth and to identify how different hatchery systems (intensive and mesocosm) may affect their expression.

2.3.1 Primers design

In order to isolate and characterize the expression of desmosomal cadherins genes in gilthead sea bream, specific primers were designed to amplify the transcripts based on the sequences retrieved *in silico* (Table 3.1).

Due to the likelihood of the presence of more *dsg* transcripts than those identified in the sea bream database by comparison with evolutionary related teleost

species (Figure 3.2), attempts to amplify the missing transcripts were performed. The procedure consisted of designing primers in the conserved regions of nucleotide alignments of the *dsg* transcripts of sea bream, sea bass and stickleback. Nucleotide alignments were analysed in the GeneDoc program to highlight the conserved regions to select the target regions for primer design. Primers of approximate 20 bp in length were designed with a melting temperature of approximately 60°C to amplify a product of 150-200 bp. Primers were chemically synthesized (Sigma-Aldrich, UK) and their sequences are described in Table 2.1.

2.3.2 Polymerase chain reaction (PCR)

The PCR is a technique invented in 1983 by Kary Mullis to amplify a DNA sequence. This consists of a thermal, enzymatic and cyclic reaction where the products of a cycle are the substrates of the next cycle and the number of DNA copies increase exponentially (2^n, n=number of cycles). This process is based on three steps, the denaturation of the DNA, primers annealing and the extension of the DNA chain. Denaturation of the template DNA molecule is caused by the heat that breaks the hydrogen bonds between the nitrogenated bases. This step usually occurs at 92-96 °C for 30 sec to 1.5 min. The primers are specific oligonucleotides that bind to the end of the denaturated DNA chains, to allow the enzyme to start the extension. The primers annealing step is performed at 48-65 °C during 30 sec to 1.5 min. In the extension phase the enzyme *Taq* DNA polymerase, that is a thermostable polymerase obtained from a thermophilic bacteria (*Thermus aquaticus*), binds to the 3' end of the primers and polymerizes a new DNA chain complementary to the template using deoxyribonucleotides (dNTPs) that are provided in the master mix. This step is usually done at 72 °C. The reaction is performed in a thermocycler that allows the variations of temperature required for the success of the reaction, the steps are repeated multiples times in order to generate numerous copies of the DNA sequence [75].

PCR was carried out to test the primers described in Table 2.1 and to perform a tissue expression analyses of desmosomal cadherins in several tissues. The PCR mixture to test the specific primers for *dsc* and *dsg b* transcripts retrieved from the sea bream data base contained 0.3U/µl of the enzyme *Taq* polymerase (Dream *Taq* DNA polymerase); 1X of a specific buffer for the enzyme (Dream *Taq* Buffer), 0.2µM of a

mixture of the 4 dNTPs, 0.2µM of the specific forward primer, 0.2µM of the specific reverse primer, 0.1 µl of cDNA from gilthead sea bream larva and MiliQ purified water to make up the mixture to a final volume of 15µl. To evaluate the *dsc* and *dsg b* expression in several sea bream tissues (skin, kidney, muscle, gills, liver, duodenum, brain and stomach), the mixture was the same but with 0.1 µl of cDNA of each tissue. To test the primers designed on conserved regions of the nucleotide alignments between sea bream, sea bass and stickleback (described in 2.3.1) a pool of larva, duodenum, skin, kidney, muscle, gill, liver and stomach cDNAs was used in the reaction. Reactions were placed in a BioRad thermocycler using the following thermal program: 3 min at 95 °C, followed by a cyclic phase of 95 °C during 20 sec, X °C during 20 sec (X for each set of primers) and 72 °C during 30 sec repeated 35 times and a final extension phase of 5 min during at 72 °C. For each set of primers (Table 2.1) two temperatures were tested: *dsc* at 60 and 64 °C; *dsg b* at 58 and 60 °C (Figure 3.7); *dsg c* (primers designed in 2.3.1) at 54 and 58 °C (data not shown).

2.3.3 Agarose gel electrophoresis

The PCR results were observed in agarose gel electrophoresis. Agarose gel electrophoresis is the most effective way of separating nucleic acids. This technique is based on the application of an electric field in an agarose gel, promoting the movement of the molecules according to their mass and size. Agarose is isolated from the seaweed genera *Gelidium* and *Gracilaria*, its polymers associate non-covalently and form a network of bundles whose pore sizes determine a gel's molecular sieving properties. Agarose powder was dissolved in buffer Tris-Acetate-EDTA (TAE) and warmed up until it completely melts. The solution was cooled down under running water and ethidium bromide (BrEt) was added. BrEt allows the visualization of the nucleic acids, since this compound is an intercalating agent that exhibits fluorescence at UV light (λ =300nm). DNA and RNA are negatively charged, thus in an electric field they migrate to the positive pole, the molecules that migrate more have a higher charge density and tend to be smaller, the results were observed using a transiluminator by comparing the position of the band containing the nucleic acid with the bands in the DNA ladder. The agarose concentration in the gel is selected according to the size of the amplification product. The agarose concentration defines

the porosity of the gel, thus affecting the migration of the DNA fragments. The higher is the agarose concentration in the gel the lower is the speed of migration of the DNA fragments, which improves the separation of the smaller molecules. Lowering the agarose concentration in the gel favours the separation of the larger molecules [76]. The concentration of agarose in the gel used in this project was 2% to better visualize the fragments lower than 200 bp with 0.5µg/mL of EtBr. 5 µL of each sample was loaded with 5 µL of loading buffer. The DNA ladder used was 1kb from Thermofisher.

2.3.4 Gene cloning

Gene cloning involves the insertion of the gene of interest into a vector, which will ensure its replication and maintenance in the new host cells, and *Eschericha coli* (*E. coli*) is the most commonly used bacterial host. The cloned genes in a vector is inserted into a host cell and replication of the gene occurs when the bacteria divide [75].

2.3.4.1 Ligation reaction

The PCR products derived from the PCR ligated in the cloning vector pGEM® - T Easy vector (Promega) (cloning vector) using the T4 DNA ligase enzyme, that catalyzes formation of phosphodiester bonds between the 3 ' -OH and 5'-P of DNA and requires ATP and Mg^{2+} as cofactors. The reactions performed were: 2.5U/µL of T4 DNA ligase enzyme, 0.2X of T4 ligase buffer, 0.3 µl of the vector pGEM® - T Easy and 4.2 µl of the PCR product. The reaction mixture was incubated at 4 °C overnight.

2.3.4.2 Bacterial transformation

To 100 µl of competent cells of *E. coli*, previously prepared in the laboratory and stored at -80 °C, 5 µl of the ligation reaction was added and incubated for 30 min on ice. Then the mixture was subjected to a thermal shock at 42 °C during 2 min and then placed again on ice for 5 min. Bacteria were plated on LB agar with ampicillin (50 mg/ml), IPTG (0.5 M) and X-gal (80 mg/ml) and were placed at 37 °C and incubated overnight. The pGEM-T easy vector has an ampicillin resistance gene that allows only the drug-resistant bacteria to grow.

2.3.4.3 Positive clone selection

Grown positive clones have a white colour as the vector possesses a lac- Z gene segment for the *E. coli* β-galactosidase at the multiple cloning site (MCS) and when there is insertion of foreign DNA the production of the enzyme B-galactosidase is inhibited. When there is no insertion of the fragment into the MCS the enzyme β-galactosidase is synthesized and degrades the X-gal substrate (added to the plates) and clones become blue. Colony PCR was performed using vector specific primers to confirm positive clones. For that white colonies were picked from the plates and grown for 2 hours at 37 °C in 70 µl of LB liquid medium and the culture used for colony PCR. One µl of the bacterial culture was used directly as template in the PCR reaction with the primers M13F and M13R (Table 2.1). The thermal program was the following: 95 °C during 10 min (to ensure disruption of the bacteria by breaking the cell wall); 95 °C during 30 sec, 60 °C during 30 sec, 72 °C during 30 sec repeated 35 times; and 72 °C during 5 min. The results were observed on agarose gel and positive colonies show a DNA band which size combines the DNA fragment of interest plus the cloning vector (180 bp). The negative ones only produce bands of 180bp that corresponds to the empty cloning vector. After confirmation they were subsequently grown in 5 ml of Liquid LB + ampicillin for isolation of plasmid DNA (2.3.5) with agitation at 37 °C overnight.

2.3.5 Plasmid DNA extraction – MiniPrep

This procedure is based on the alkaline lysis method. Approximately 1.5 ml of bacterial culture in liquid LB with ampicillin was transferred to a microcentifuge tube and centrifuged for 1 min at 13,000 rpm at room temperature. With this a cell pellet was obtained and the supernatant discarded, this step was repeated twice. The cell pellet was resuspended in 300 µl of P1 solution (50 mM glucose, 25 mM Tris-HCl, 10mM EDTA, pH 8 with RNAse A), then the cell suspension was incubated for 5 min at room temperature. The next step was to add 300 µl of P2 (0.2 M NaOH, 1% SDS), mix by inversion and place the mixture on ice for 10 min. To the lysate 300 µl of P3 (3 M potassium acetate, that precipitates the DNA) was added. Mixed by inversion and incubated 10 min on ice. Then was centrifuged 15 min at 13,000 rpm and the supernatant was removed to a new microcentrifuge tube, where 700 µl of cold 100%

ethanol was added to promote DNA precipitation. This was mixed by inversion and centrifuged for 15 min at 13000 rpm. The DNA pellet was washed twice with 200 µl of 70% ethanol using centrifugations of 5 min at room temperature at 13000 rpm. The pellet was air dried and the DNA resuspended in 40 µl of MilliQ water. The isolated plasmid DNA was sent for sequencing to confirm its identity in the Sequencing facilities of the Center of Marine Sciences from the University of Algarve. The results of *dsc* and *dsg* were aligned with the original sequence retrieved from the *Sparus aurata* database using ClustalW.

2.3.6 RNA extraction and cDNA synthesis

To perform a tissue expression analysis, total RNA was extracted from skin, kidney, muscle, gills, liver, duodenum, brain and stomach from gilthead sea bream, for the synthesis of cDNA and qualitative analysis of expression by PCR. The Maxwell® 16 Total RNA Purification Kit (Promega) was used to extract and purify total RNA from tissues. This method comprises four essential steps: lysis of tissues and cells, denaturation of nucleoprotein complexes, inactivation of endogenous ribonucleases and contaminants removal (proteins and DNA). The success of this kit is based on disruptive and protective properties of guanidine thiocyanate that can lysate samples, denature nucleoprotein complexes and inactivate ribonucleases. The protocol was followed as described by the kit manufacturer. Approximately 50 mg from each tissue were collect and homogenized in lysis solution and β-mercaptoethanol (BME), then were mechanically disrupted using the Ultra-Turrax ®. After this, several steps were performed including the addition of Clearing Agent to remove genomic DNA from the sample lysate. To finish the process, total RNA is captured from the cleared sample lysate using magnetic beads (MagneSil® PMPs). Total RNA were further purified automatically from contaminating salts, proteins and cellular impurities by ethanol washes. The RNA obtained was diluted in 30µL of MiliQ water, the RNA was separated on an agarose 2% gel in to verify their integrity. The total RNA previously extracted was treated with TURBO DNA-free™ Kit (Ambion ® by life technologies) according to manufacturer's protocol. This kit is designed to remove contaminating DNA from RNA preparations and to subsequently remove the DNase and divalent cations from the samples. To the total RNA (up to a maximum of 10 µg) were added 5µL of buffer, 0.75

µL DNAse (DNAse Turbo ™, Ambion, life technology, 2 U / µl) and water to obtain a final volume of 50 µl. The reaction mixture was incubated for 30 min at 37°C and then 5 µL of DNase inactivation reagent were added and incubated during 5 min at room temperature, mixing occasionally. Finally, the reaction mixtures were centrifuge at 10,000 × g for 1.5min and the supernatant (RNA) was transferred to a fresh tube. Thermo Scientific RevertAid RT Kit was used to cDNA synthesis from total RNA previously extracted, purified and treated. The kit uses RevertAid Reverse Transcriptase and recombinant Thermo Scientific RiboLock RNase Inhibitor to effectively protect RNA from degradation at temperatures up to 55°C. The experimental procedure was done according to the manufacturer protocol. 500ng of total RNA were used to cDNA synthesis, to a maximum volume of 12.5 µL.

Larvae used to perform study 1 and 2 were anesthetized in 2-phenoxyethanol (1:5000, Sigma-Aldrich, P1126) and placed in RNAlater Reagent (Sigma-Aldrich, R0901) and stored at -20 °C until use. Total RNA was extracted from the individual using TRI Reagent (Sigma,T9424) according to the manufacturer's instructions. Total RNA was subsequently subjected to DNAse treatment with DNA-free (Ambion, AM1906) to remove traces of contaminating genomic DNA and was stored at -80 °C until further use. cDNA synthesis was performed simultaneously for all samples to ensure the same reaction efficiency. cDNA was generated from 1 µg total RNA by using 200U/µl SuperScript II reverse transcriptase (Invitrogen, 18064-014), 3 µg random primers (Invitrogen) and 40 U/µl recombinant RNaseOUT ribonuclease inhibitor (Invitrogen) for a total reaction volume of 27 µl.

2.3.7 Real time quantitative PCR (q-RT-PCR)

The q-RT-PCR is a variant of the classical PCR with the same basis where the amplification and quantification are accomplished in a single step. This is possible since a fluorescent reporter molecule is present in the master mix to monitor the progress of the amplification. Thus, the amount of amplified product can be correlated to the fluorescence intensity. In this work SYBR Green I was the reporter molecule used, this compound exhibits a low fluorescence when is free in a solution but when it bounds to a double-stranded DNA its fluorescence increases by over 1000-fold. A threshold level of fluorescence must be set above the background level fluorescence of the free SYBR

Green I, but still within the linear phase of amplification for all the plots. The cycle number at which an amplification plot crosses the threshold level of fluorescence is defined as Ct. The Ct is the indicator of the initial DNA template in the sample, the lower the Ct the greater the amount of DNA template in the sample [77].

All of the q-RT-PCR reactions were carried out in duplicates (< 5 %variation between duplicates) with SYBR® Select Master Mix (Applied Biosystems) in a Mx 3005P ™ qPCR System (Stratagene®) using the MxPro ™ QPCR software. All reactions were set up using 100–300 nmol/L of each primer and 0.17 µg/µL of cDNA sample (1:5 diluted) in a reaction with 20 µL of final volume. The thermal cycle used in q-RT-PCR were 2 min at 50 °C, then 2 min at 95 °C followed by 40 cycles of 15 sec at 95 °C, 1 min at 61 °C and 1 min at 72 °C. The previous procedure was followed by the dissociation curve step (1 min at 95 °C, 30 sec at 55 °C, 30 sec at 95 °C) to verify that the amplification of a single product occurred to exclude the possibility of contaminations.

Standard curves were constructed to determine the efficiency of each primer pair at different combinations of primer concentration. Calculations were performed according to [77] and considering a perfect doubling of the template: an efficiency of 100% or 1, a 10-fold amplification should take 3.32 cycles ($2^{3.32}$=10). Thus, in a plot of Ct values against the log of the dilution factor, the slope must be close to -3.32. The efficiency (Table 2.1) was calculated directly from the slope using the equation below:

$$Efficiency = 10^{(-1/slope)}\text{-}1$$

When measuring RNA expression it is essential to control the error between samples. A standard procedure is to normalize transcript expression using internal reference or housekeeping genes (HKG). A good HKG should have a constant expression in all the tissues and under different experimental conditions [78]. In study 1, three HKG were used, elongation factor 1-alpha (*ef1a*), *β-actin* and ribosomal protein L13a (*rpl13a*) to calculate the normalization factor. In study 2, *rpl13a* and ribosomal protein S18 (*rps18*) were the genes selected to normalize the genes of interest (Table 2.1).

Table 2.1: Set of primers used in the PCR and q-RT-PCR and their respective efficiency. Fw – forward and Rv – reverse."*"indicates primers already designed in previous studies [75,26]

Gene	Primers	Efficiency (%)
dsc	Fw - ATGGATCAACTTATCCTCCGAAT Rv - TACAGAAGCCCCTCATTCG	96.00
dsg b	Fw - ATGAGAGTGATAGCCACTGAT Rv - CATGACCTCTCCAGTCTGAG	93.30
dsg c	Fw - GGACTACAGTTATGAAAGTAAC Rv- TGCCTCGTTGCCTTTGACA	-
rpl13a*	Fw - TCTGGAGGACTGTCAGGGGCATGC Rv - AGACGCACAATCTTAAGAGCAG	102.89
β-actin*	Fw – CGACATCCGTAAGGACCTGT Rv - ACATCTGCTGGAAGGTGGAC	98.69
ef1a*	Fw - TCAAGGGATGGAAGGTTGAG Rv - AGTTCCAATACCGCCGAT	100.76
rps18*	Fw - AGGGTGTTGGCAGACGTTAC Rv - GAGGACCTGGCTGTATTTGC	101.11
M13	Fw - GTAAAACGACGGCCAGT Rv - AACAGCTATGACCATG	-

The starting fluorescence (R_0) for all genes, housekeeping and genes of interest, were calculated in all samples and subsequently, the normalization factor (NF) was calculated using the geometric mean of the R_0 of the HKGs for each sample [78]. The R_0 is proportional to the starting template quantity and it can be calculated using the following equation:

$$R_0 = \frac{Treshold}{1+Efficiency^{ct}}$$

The R_0 of the genes of interest were normalized dividing with NF. Many biological variables are not normally distributed and/or the standard deviations are not homogeneous. Thus, to use parametric statistical tests, transforming data will allow it to fit the assumptions better. In this experiment, due the reasons previously described, the final results were square-root transformed to meet assumptions of normality and/or homogeneity.

2.4 Statistical analyses

In study 1, two-way ANOVA was performed to evaluate differences in genes expression levels (*dsc* and *dsg b*) at different dph. T- test was used to compare larvae of 58 dph, that had been subjected to size-sorting. The desmosomal cadherins expression levels were correlated to other genes of interest (*mlc2a*; *mlc2b*; myogenin (*myog*), myostatin (*mstn*); collagen type I alpha 1 (*col1a1*); insulin-like growth factor 2 (*igf-2*); follistatin (*fst*) and myogenic regulatory factor 4 (*mrf4*)), previously studied in the same set of samples and for which the results are published [79]. The test used was the Pearson correlation.

In study 2, two-way ANOVA was used to determine intergroup differences with dph and aquaculture system (intensive vs mesocosm) as factors.

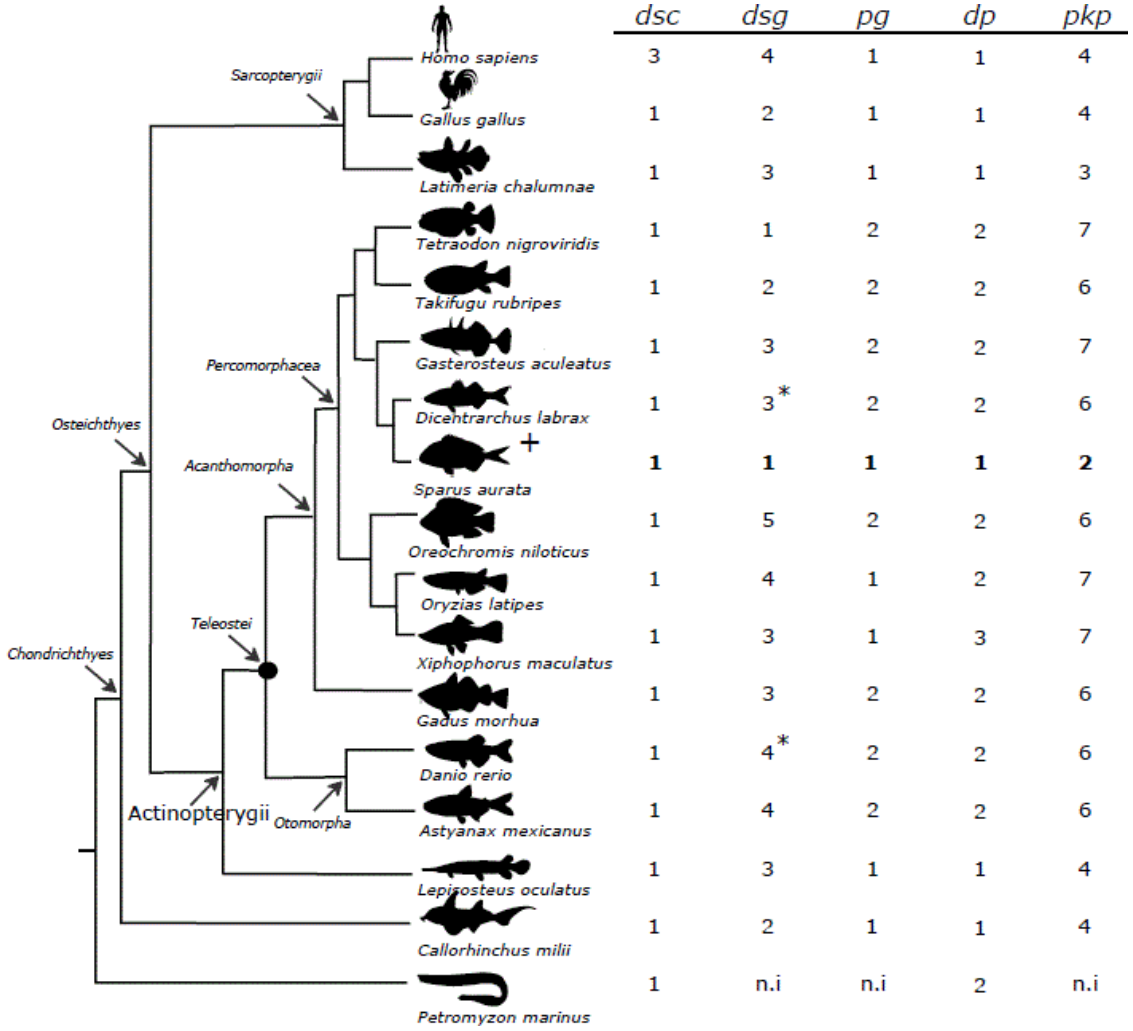
Gene expression values were expressed as means ± S.E.M. and for all statistical tests, the level of significance was set at $P < 0.05$.

3 Results

3.1 Identification of desmosomal genes from fish

The desmosomal genes were identified based on their sequence similarity with the human and zebrafish homologues by *in silico* searches in the genomes, nucleotides and ESTs databases of several fish. The search included a wide range of species evolutionarily distinct, such as the *Agnatha* class*,* the lamprey (*Petromyzus marinus)*; the cartilaginous fish elephant shark (*Callorhinchus millii*); several teleost fish and the coelacanth (*Latimeria chalumnae*), a lobefin fish sister of tetrapods.

Figure 3.1 summarizes the number of genes and transcripts (for the sea bream) of the different desmosomal families that were identified during this study in fish and other vertebrates. All the genes and respective proteins accession numbers retrieved in the genome databases are supplied in Annex Tables I.1; I.2; I.3, I.4 and I.5; the accession number of the ESTs retrieved are supplied in Annex Tables II.1 and II.2; and the desmosomal protein transcripts that were retrieved from the sea bream database are described in Table 3.1.



	dsc	*dsg*	*pg*	*dp*	*pkp*
Homo sapiens	3	4	1	1	4
Gallus gallus	1	2	1	1	4
Latimeria chalumnae	1	3	1	1	3
Tetraodon nigroviridis	1	1	2	2	7
Takifugu rubripes	1	2	2	2	6
Gasterosteus aculeatus	1	3	2	2	7
Dicentrarchus labrax	1	3*	2	2	6
Sparus aurata +	**1**	**1**	**1**	**1**	**2**
Oreochromis niloticus	1	5	2	2	6
Oryzias latipes	1	4	1	2	7
Xiphophorus maculatus	1	3	1	3	7
Gadus morhua	1	3	2	2	6
Danio rerio	1	4*	2	2	6
Astyanax mexicanus	1	4	2	2	6
Lepisosteus oculatus	1	3	1	1	4
Callorhinchus milii	1	2	1	1	4
Petromyzon marinus	1	n.i	n.i	2	n.i

Figure 3.1**:** Cladogram describing the number of desmosomal genes (*dsc*, *dsg*, *pg*, *dp* and *pkp*) identified in fish and other vertebrates. The TSWGD is represented by the closed circle "•". "+" indicates that for sea bream the members correspond to transcripts retrieved from its specific assembly. "*" indicates members not predicted in the genome and retrieved from EST; n.i. - not identified. Accession numbers are available in Annex Tables I.1, I.2, I.3, I.4, I.5, II.1, II.2 and Table 3.1.

Homologues of all five families of desmosomal genes that have been identified in human were found in fish, suggesting that members of these families aroused early in the vertebrate evolution. However, the gene number between the families and across species of the same family group was distinct. In all teleost and other non-mammalian vertebrates searched in this study a single *dsc* gene was found in the genomes (Figure 3.1). This includes key evolutionary important species such as the coelacanth, a lobefin fish that diverged prior to tetrapods, the spotted gar, a ray-fin fish that diverged prior to the teleost, the cartilaginous fish elephant shark and the

jawless fish lamprey. This contrasts with the gene family expansion that occurred in the Human genome.

Across teleost fish the number of *dsg* is highly variable and a different number of genes ranging from one in tetraodon to five in tilapia were retrieved (Figure 3.1). In the spotted gar and coelacanth three *dsg* were found, two in the elephant shark but in lamprey no putative *dsg* were identified and this may be the consequence of its incomplete genome annotation. In chicken two *dsg* were collected but in mammals gene expansion seems to have occurred and four *dsg* were found in human and three in opossum.

In the majority of the teleost, two *pg* exist but a single gene was retrieved from medaka and platyfish (Figure 3.1). A single *pg* was also retrieved from tetrapods, coelacanth, spotted gar and elephant shark. In the lamprey genome, *pg* were not identified. The *dp* retrieved revealed a pattern similar to the *pg* (Figure 3.1). In general teleosts have two *dp* and a single *dp* was found in all tetrapods, in the coelacanth, in spotted gar and in the elephant shark genomes. In contrast, the lamprey genome revealed the existence of two *dp*.

The *pkp* family was the most diverse and a highly variable number of genes were found across vertebrates (Figure 3.1). In general, six to seven *pkp* were retrieved from teleost fish and in the spotted gar and elephant shark genomes, four *pkp* were found. In lamprey no putative *pkp* were found. Also the tetrapods human, opossum and chicken possess four *pkp* while in coelacanth only three *pkp* were mined.

Genomes databases searches were complemented with ESTs (Annex Tables II.1 and II.2). Exploitation of ESTs databases revealed the existence of another *dsg* (that were not retrieved from the genome assembly) in sea bass (FM020296.1) and in zebrafish (CB363329.1), which matched for no-annotated genome regions in these species. These results are not totally reliable, since the EST sequences were too short and using the procedure described in 2.2.2 the classification can be misleading.

Searches in the transcript database of sea bream, identified: a single *dsc* and *dsg* and two *pg*, *dp* and *pkp* putative transcripts (Table 3.1). However analysis of the sequence alignment of the retrieved transcripts in the sea bream with the homologues from other teleost revealed that the two *pg* transcripts (lcl|Contig14336 and lcl|Contig7275) correspond to different regions of the same transcript: the

lcl|Contig14336 maps to the N-terminus and the lcl|Contig7275 to the C-terminus. Similarly the two *pg* transcripts (SRR278741_isotig2745 and Contig8069) correspond to different regions of the same transcript. The transcripts were concatenated and used in the phylogenetic analysis.

Table 3.1: Accession numbers of the desmosomal transcripts retrieved from the sea bream nucleotide database.

Member	Accession number
dsc	SRR278741_isotig25820
dsg	SRR278741_isotig19972
pg	lcl|Contig14336/ lcl|Contig7275
dp	lcl|SRR278741_isotig2745 lcl|Contig8069
pkp	lcl|SRR278741_isotig43378 lcl|SRR278741_isotig28089

3.2 Phylogenetic analysis

The fish and other vertebrate desmosomal deduced proteins were used to construct phylogenetic trees to characterize the evolution of this large protein family in vertebrates. Trees for the different family members were constructed using the conserved domains and with an outgroup sequence to root. To build the trees some sequences were excluded: the *dsg* sequences found in the EST analysis for the sea bass (FM020296.1) and zebrafish (CB363329.1) as they were very incomplete; the new *dsg* from sea bream found using the procedure described in 2.3.1, since the sequence was also very incomplete; the *pg* sequences from medaka (ENSORLP00000015702) and coelacanth (ENSLACP00000006244) as they share very low sequence similarity with the other vertebrate *pg* and may be wrongly deduced; and the lamprey *dp* (ENSPMAP00000001819 and ENSPMAP00000002255) since the first one was too short and the second one was very different from the rest of the others *dp*.

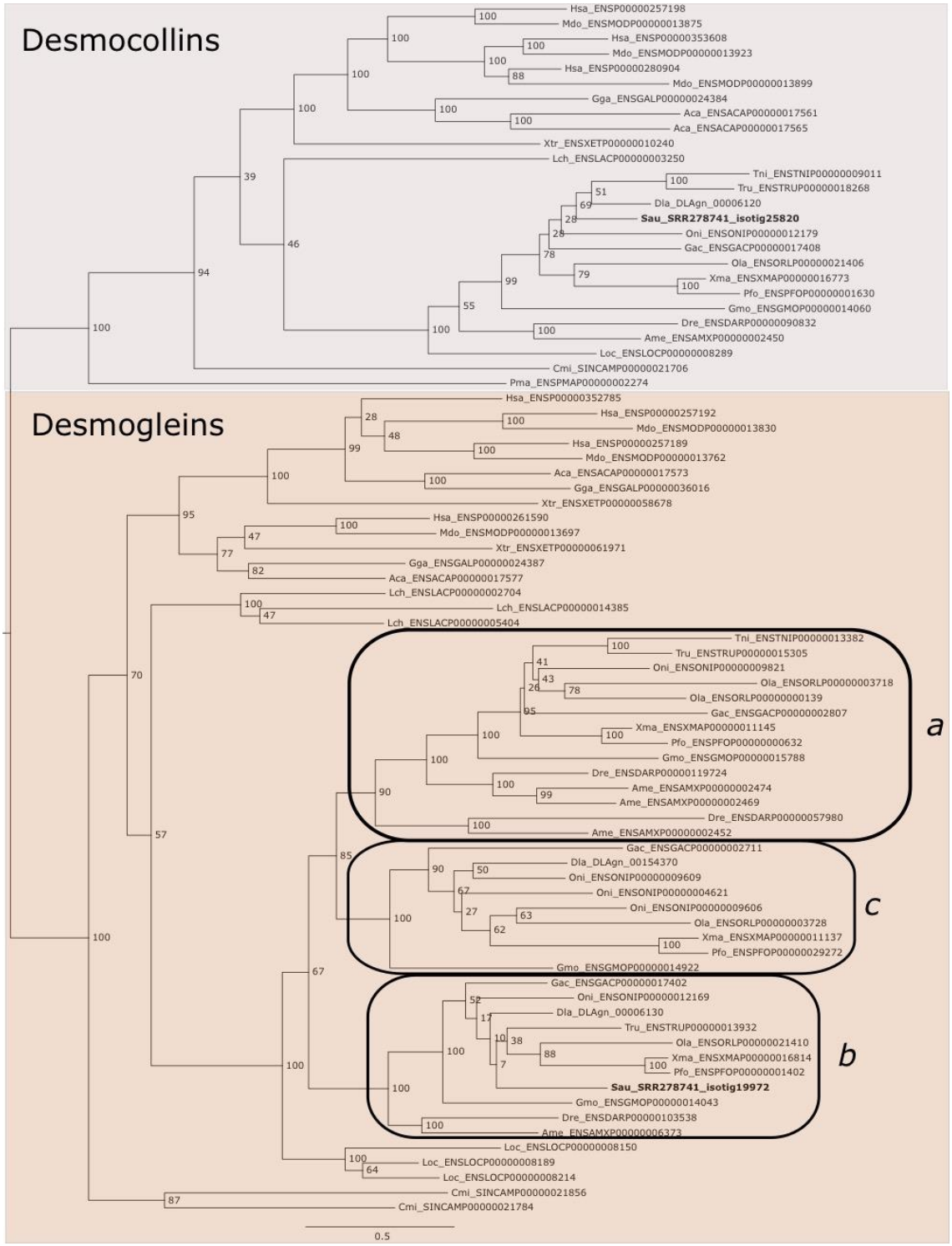


Figure 3.2: Phylogenetic tree of *dsc* and *dsg* members of fish and other vertebrates. Tree was obtained with the model WAG+I+G using the method maximum likelihood with a bootstrap analysis of 100 replicates. *a*, *b* and c represent the different *dsg* duplicates identified by the clustering organization. Accession numbers in Annex Table I.1, I.2 and Table 3.1.

Dsc and *dsg* sequences were compared in the same tree as they correspond to different members of the same superfamily and they are the candidate genes for the study. Analysis of the fish and other vertebrate *dsc* confirmed the existence in the majority of the vertebrates of a single gene. However, in human and opossum three genes were identified and clustering of the sequences in the tree suggest that they are the result of two lineage specific gene duplication that only occurred in mammals. The evolution of the *dsg* is much more complex and the different gene numbers in the species analysed seem to have resulted from lineage specific or species-specific gene duplication events. Phylogenetic analysis clustered the teleost *dsg* in three different clades, *dsg a*, *dsg b* and *dsg c*, with members of the first two groups previously described in zebrafish [80], whereas the *dsg c* group was new and discovered for the first time in this study. The teleosts, tilapia, stickleback, medaka, amazon moly and platyfish genomes contain members of the three different *dsg* clades and according to the tree topology they seem to have been originated during the gene/genome tetraploidization that occurred early at the teleost radiation and followed by a subsequent duplication event. *Dsg a* and *dsg c* share the same ancestral origin and emerged subsequent to *dsg b*. The maintenance of the gene duplicates in the teleost genomes was species specific. Several teleosts contain a different number of *dsg* that seemed to be the result of species-specific gene duplications. The tree classified the *dsg* transcript from the gilthead sea bream as *dsg b*. *Dsg a* was absent although verification will require access to the full genome. In other fishes the members of *dsg* also duplicated as a consequence of species-specific events and three genes were found in coelacanth, spotted gar and two in the elephant shark. In the tetrapods, chicken, xenopus and lizard a specific gene duplication occurred. In the mammals human and opossum more rounds of duplication of *dsg* occurred leading to three gene duplicates in opossum and four genes in the human.

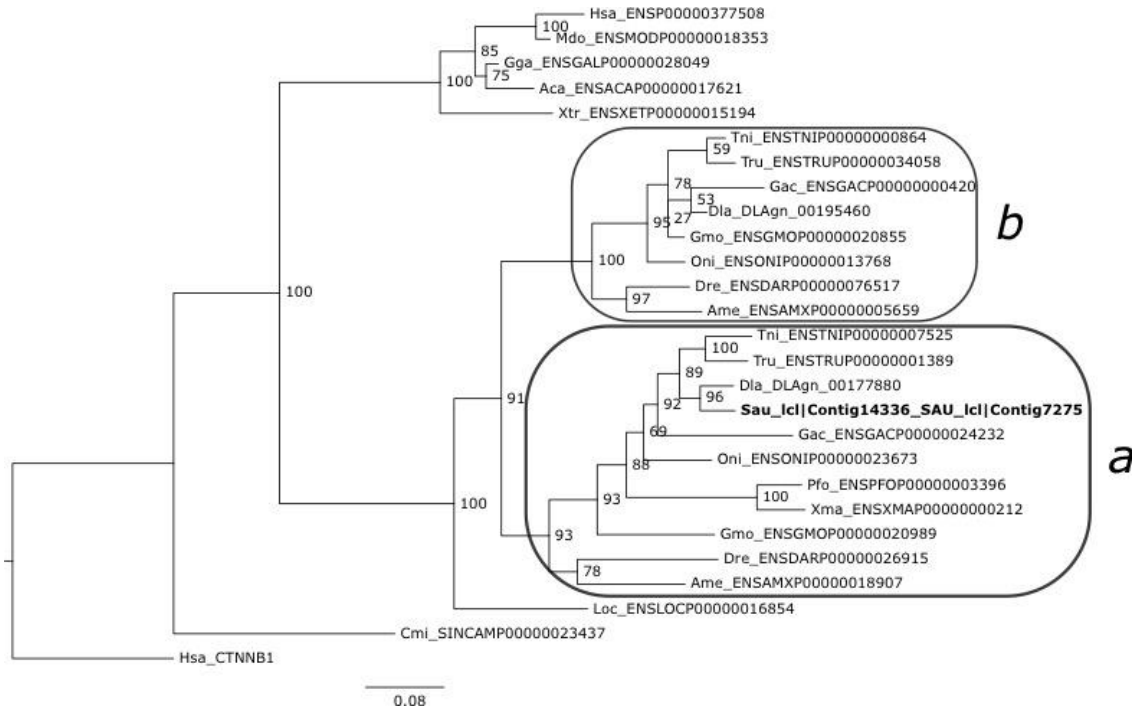


Figure 3.3: Phylogenetic tree of the *pg* members of fish and other vertebrates. The tree was constructed with the model JTT+I+G+F using the method maximum likelihood with a bootstrap analysis of 100 replicates. The tree was rooted with the human β-catenin 1 protein. *a* and *b* represent the different *pg* duplicates identified by the clustering organization. Accession numbers in Annex Table I.3 and Table 3.1.

The phylogenetic tree exhibits two clusters in teleosts, revelling the existence of two *pg* duplicates (Figure 3.3). The duplicates were classified according to the previous study where the authors considered ENSDARP00000026915 sequence of zebrafish as *pg a* [81], thus the other duplicate from now on will be denominated as *pg b*. The tree shows that the sequence of sea bream obtained by the concatenation of the two transcripts retrieved from the database was a *pg a* (Figure 3.3). Several species contain the two *pg* duplicates (tetraodon, fugu, sea bass, stickleback, tilapia, zebrafish, cod and cavefish) while sea bream, platyfish and amazon moly only exhibit the *a* duplicate. Specific duplications were not found, all the species contain only a single form of each *pg* duplicate. In tetrapods, spotted gar and elephant shark a single *pg* exists (Figure 3.3).

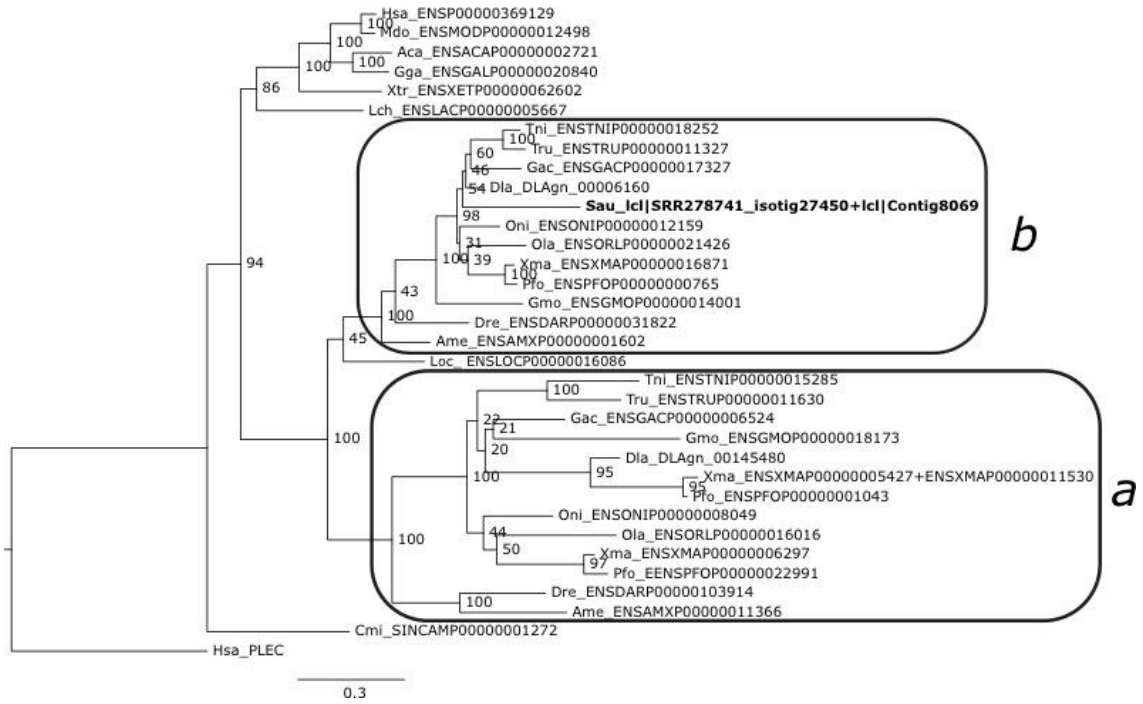


Figure 3.4: Phylogenetic tree of *dp* members of fish and other vertebrates, obtained with the model JTT+I+G+F using the method maximum likelihood with a bootstrap analysis of 100 replicates. The tree was rooted with the human plectin protein. *a* and *b* represent the different *dp* duplicates identified by the clustering organization. Accession numbers in Annex Table I.4 and Table 3.1.

Dp tree shows a similar organization as the *pg* tree. The tree exhibits two perfect clusters in the teleosts revealing the existence of two duplicates. However previous studies on *dp* in this infraclass were not found, thus the genes were denominated as *a* and *b* and the duplicate *a* is the most similar to human *DP* (Figure 3.4) The teleosts analysed and from which *dp* were retrieved and used to construct the phylogenetic tree had *dp* duplicates, with the exception of the sea bream that contained a single *dp*, which clustered with the *dp a*, the absence of a duplicate in sea bream is most likely a result of the incomplete genome assembly. Platyfish and amazon moly are the only species that have specific duplications, in *dp b* both exhibit two forms. The tetrapods and the fish spotted gar and elephant shark only contain one *dp* (Figure 3.4).

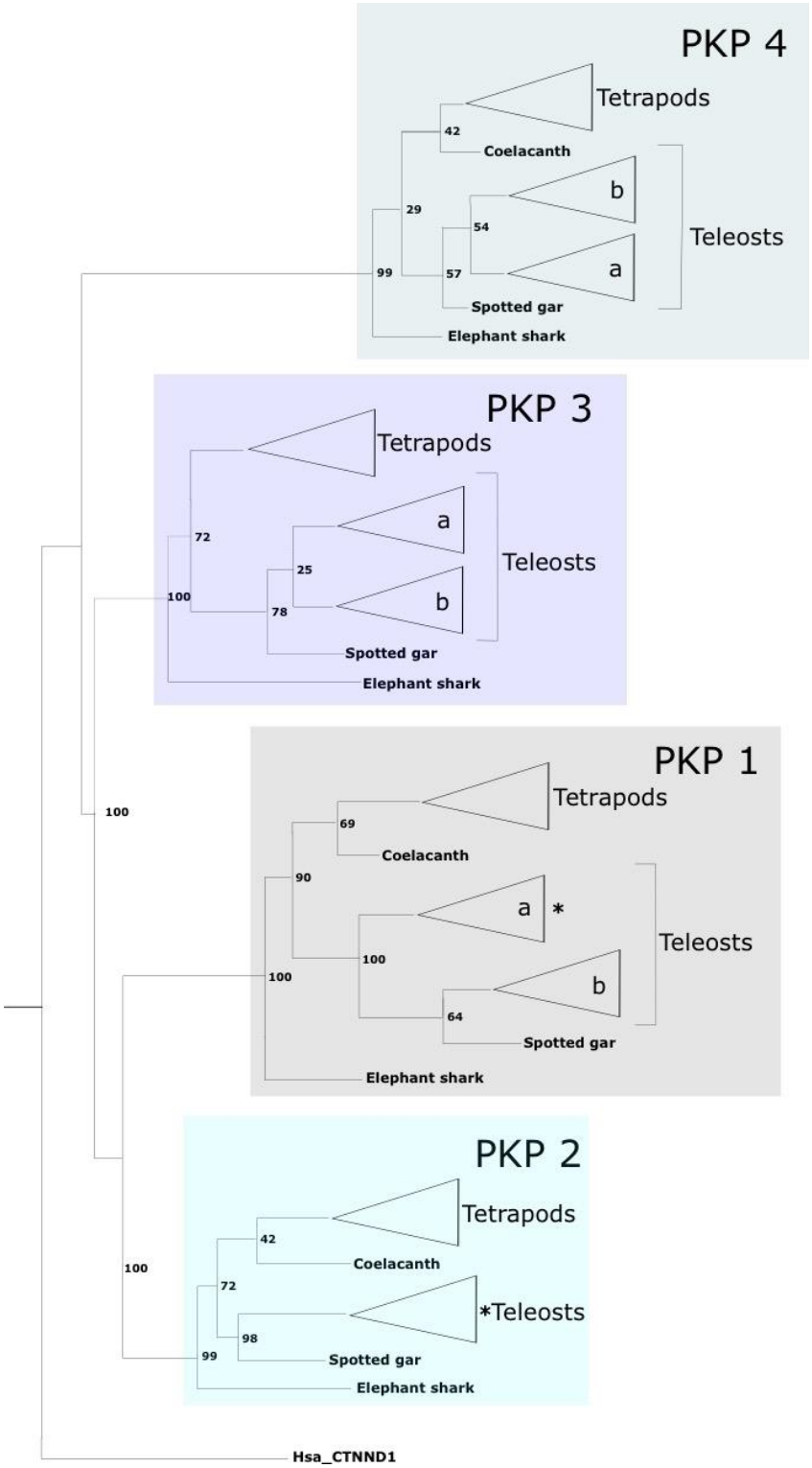


Figure 3.5: Phylogenetic tree of *pkp* members of fish and other vertebrates, obtained with the model JTT+I+G+F using the method maximum likelihood with a bootstrap analysis of 100 replicates. The tetrapod and teleost branches were collapsed to facilitate interpretation. The complete tree is available as Figure Annex III.1. The tree was rooted with the human δ-catenin 1. * indicates the presence of a gilthead sea

bream transcript in the clade. *a* and *b* represent the different *pkp* duplicates. Accession numbers in Annex Table I.5 and Table 3.1

The *pkp* were classified according to the human *PKP*. Four different *pkp* were identified and *pkp1*, *pkp3,* and *pkp4* exhibit two duplicates in teleosts denominated as *a* and *b*. A previous study on zebrafish, considered as *pkp3a* the sequence ENSDARG00000051861 [82], thus this nomenclature was followed and the other duplicate was denominated as *pkp3 b*. For *pkp4* and *pkp1* previous studies were not found in teleosts, therefore the duplicates more similar to the human were considered as *a* and the other as *b*. According to tree the sequences retrieved from sea bream database were a *pkp2* (lcl|SRR278741_isotig43378) and a *pkp1 a* (lcl|SRR278741_isotig28089). All *pkp* emerged from a common ancestral *pkp* molecule and the tree topology suggests that *pkp4* was the first to have diverged. According to the tree the three *pkp* found in the coelacanth correspond to *pkp4*, *pkp2* and *pkp1* (Figure 3.5).

A phylogenetic analysis was also performed with nucleotide sequences to characterize the *dsg* ESTs and the new *dsg* amplified in sea bream with the primers designed on the conserved regions of the alignment between sea bream, sea bass and stickleback (data not shown). This tree shows that the new *dsg* found in sea bream clusters with the *dsg c* from stickleback, which allows conclude that the new transcript found may be a *dsg c.* The classification results of the ESTs are described on the Annex Table II.1 and II.2.

3.3 Gene synteny of the desmosomal cadherins

The gene synteny was performed for desmosomal cadherins, as they were the main focus of this study due to their importance in the formation of the desmosomes and the fact that evolution was quite complex (Figure 3.6).

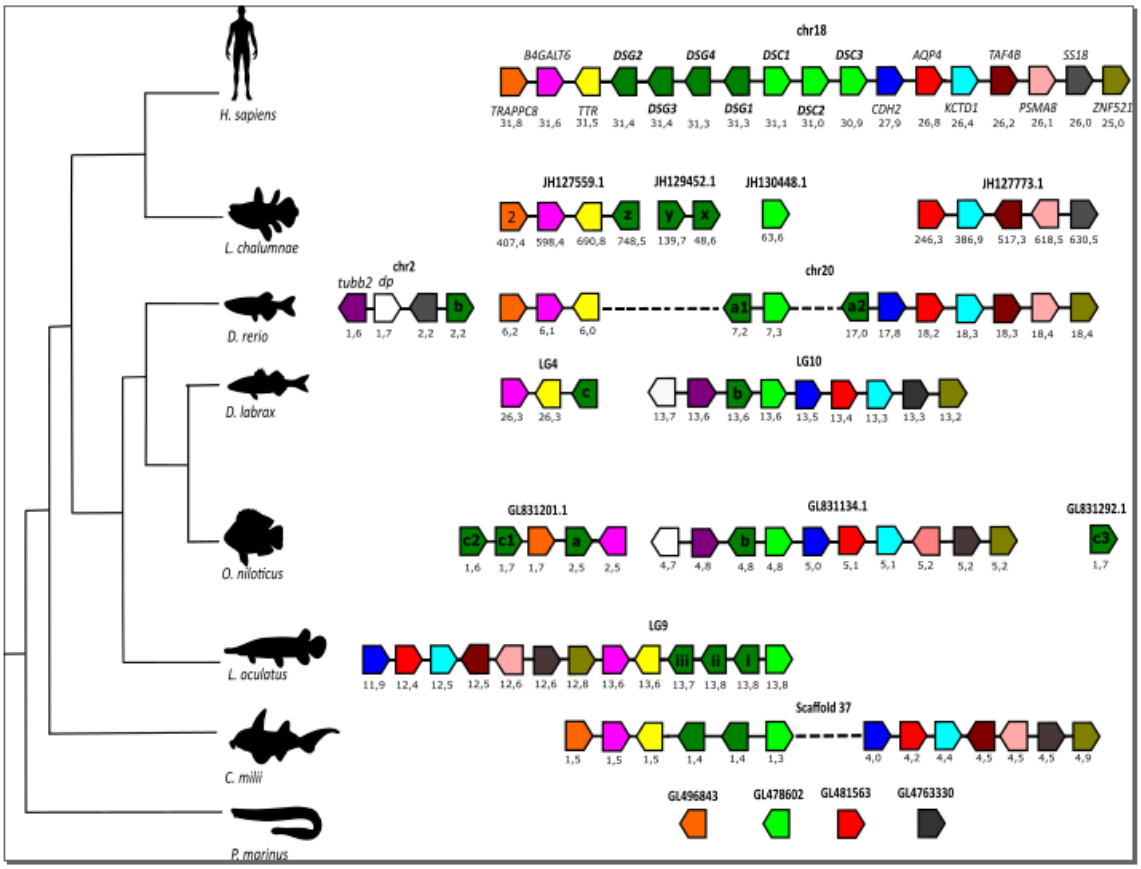


Figure 3.6: Characterization of the neighbouring gene environment of the desmosomal cadherins in some fishes and in human. Each box represents a gene and each gene is defined by a colour, the position of the gene is indicated below and the arrow defines gene orientation. The genes represented in the figure contain their official abbreviations: *DSC*-desmocollin; DSG-desmoglein; *TTR*-transthyretin; *BGALT6*-β-1,4-galactosyltransferase 6; *TRAPPC8*- trafficking protein particle complex 8; *CDH2*-cadherin 2; *AQP4*-aquaporin 4; *KCTD1*- potassium channel tetramerization domain containing 1; *taf4b* - TATA-box binding protein associated factor 4b; *PSMA8*-proteasome subunit alpha 8; *SS18*- nBAF chromatin remodeling complex subunit; *ZNF521* - zinc finger protein 521; *tubb2*- tubulin β-2A class IIa and *dp*-desmoplakin. The dashed line represents the existence of a big distance between genes on the same chromosome or scaffold that is described in the upper part of the genes. Single numbers inside the forms represent that "x" number of that gene exits one after the other. a, b and c letters inside *dsg* forms describes which duplicate it is according to the phylogenetic tree (Figure 3.2). For coelacanth and spotted gar x,y,z and i, ii, iii represent respectively the *dsg* duplicates of these two species.

Analysis of the *dsc* and *dsg* gene environment revealed that in general in the vertebrates both genes map to the same chromosomes, suggesting that they may have shared common origin and that they emerged from a gene tandem duplication event prior or very early to the vertebrate radiation (Figure 3.6).

In human, both *DSG* and *DSC* are localized in chromosome 18 and the disposition of the four *DSG* and of the three *DSC* genes suggests that the extra family members that are present in mammals are the result of tandem gene duplications that occurred specifically in their lineage (Figure 3.6).

In fish, the evolution of the family members was complex as revealed by the characterization of the homologue gene environment in the teleost. However, in the coelacanth (the lobe-fined fish), in the ray finfish the spotted gar and in the cartilaginous fish the elephant shark, despite the different number of *dsg* and *dsc* a similar gene environment to human was found. The existence of multiple *dsg* in the genomes of these species confirm phylogenetic rearrangements and suggests that they also resulted from tandem gene duplication events that specifically occurred in each fish lineage (Figure 3.6). In lamprey no conserved gene synteny was found as this is probably a consequence of the existence of short genome scaffolds.

In teleosts as the result of the lineage specific gene duplication and species divergence members of this family duplicated and gene retention was distinct (Figure 3.6). Analysis of the teleost gene environment revealed an interesting pattern and the paralogue *dgs b* gene is always near *tubb2* and the *dsc* gene localised near the genes *cdh2*, *aqp4*, *kcdtd1*, *taf4b ss18 psma8* and *znf521*. In zebrafish *dsc* maps to chromosome 20 and is in this chromosome that most of the genes that share sequence homology for the human *DSG/DSC* gene environment are localized. The zebrafish exhibits a specific duplication in *dsg a* (*dsg a1* and *dsg a2*), both forms plus *dsc* gene are present in chromosome 20. *Dsg*a2 form is closer to the *dsc* gene and the other form share several neighbour genes with other species. *Dsg b* is present in chromosome 2 and has as neighbour genes *ss18*, *dp* and *tubb2*. The sea bass *dsg c* gene environment is very similar to the *dsg* from other species. *Dsg b* maps to the same chromosome as *dsc* and these genes also sit next to each other and resemble the gene environment of *dsg b* in the zebrafish. In tilapia duplicates of *dsg c* (*dsg c1, dsg c2, dsg c3*) exist and *dsg c1* and *dsg c3* are in the same chromosome as *dsg b*, while *dsgc2* sits in a completely different chromosome. *Dsg b* sits next to *dsc* and their gene environment is identical to the other teleost species.

3.4 Isolation of the gilthead sea bream *dsc* and *dsg*

Specific primers were designed to amplify the sea bream *dsc* and *dsg b* based on the nucleotide sequences of the partial transcripts retrieved from this species-specific database (Table 2.1). The primers for each transcript were tested by PCR with sea bream larvae cDNA at several temperatures.

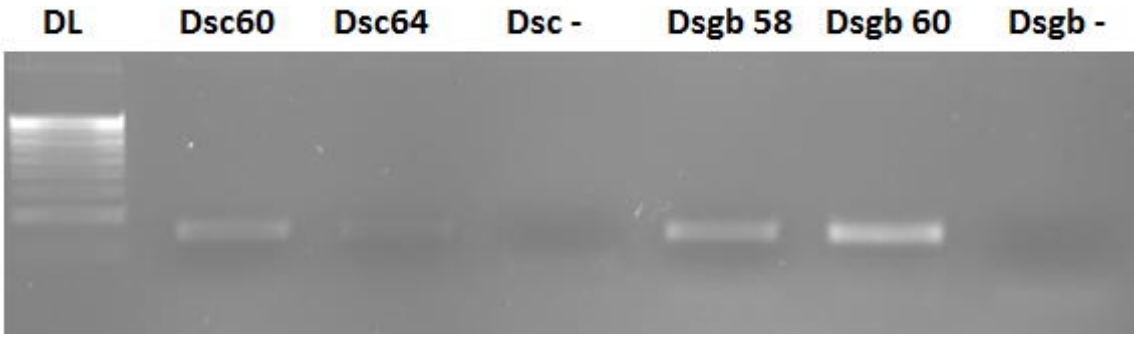


Figure 3.7**:** Agarose gel electrophoresis of the *dsc* and *dsg b* PCRs . DL – 1 kb DNA ladder; Dsc60 – *dsc* primers tested with Tm° of 60; Dsc64 – *dsc* primers tested with Tm° of 64; Dsc- – negative control; Dsgb 58 – *dsg b* primers tested with Tm° of 58; Dsgb 60 – *dsg b* primers tested with Tm° of 60; Dsgb - – negative control.

Analysis of the *dsg* from other teleost suggested that at least three *dsg* may potentially exist in sea bream however only a single transcript was found. In order to amplify the putative missing *dsg* transcripts primers were designed based on conserved regions of a nucleotide alignment of *dsg* transcripts from evolutionarily related species (Figure 3.8).

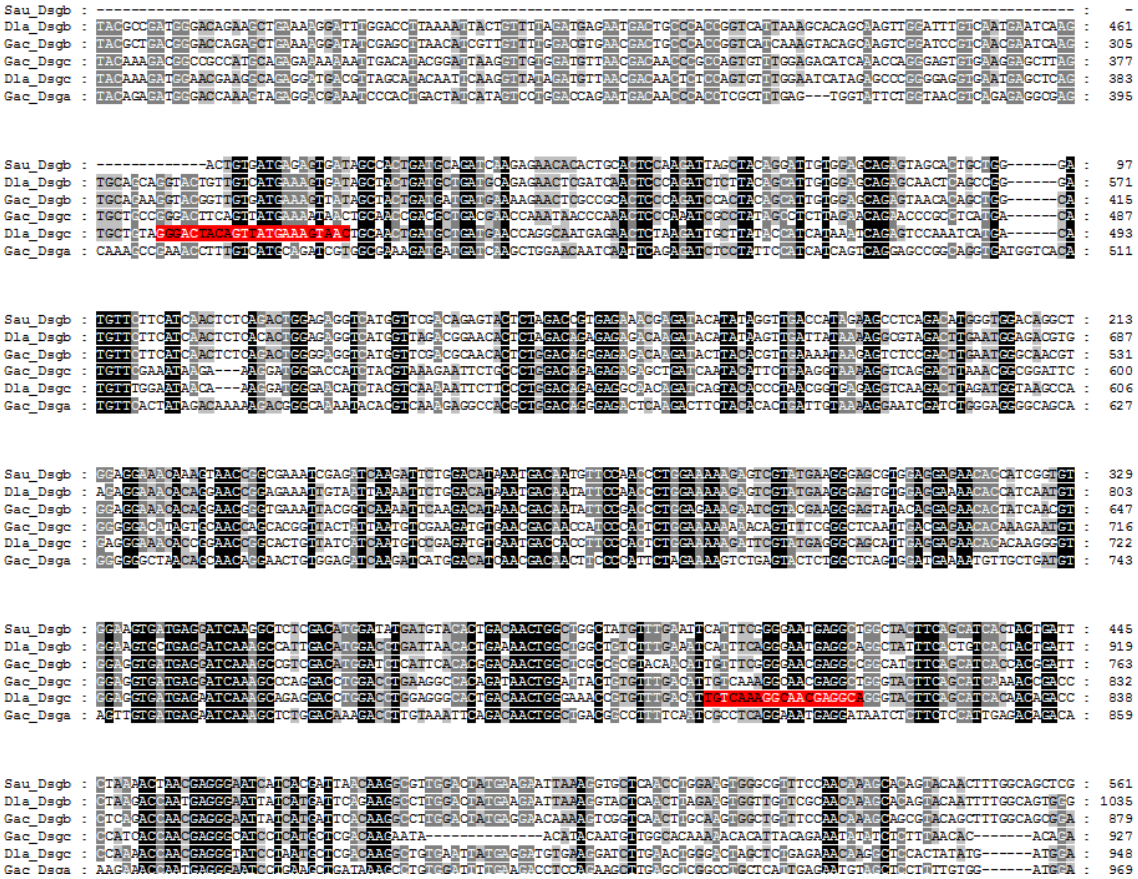


Figure 3.8**:** Nucleotide sequence alignment between *dsg a*, *dsg b* and *dsg c* from stickleback; *dsg b* and *dsg c* from sea bass and *dsg b* from sea bream. The conservation of nucleotides among the genes are represented with a colour gradient where total conservation is black and zero conservation is white. The primers that generated an amplification product are highlighted with red.

After several attempts, a set of primers (Table 3.1) generated an amplified product from a pool of sea bream cDNAs (duodenum, skin, kidney, muscle, gill, liver and stomach) at 54 °C. Sequence analysis revealed that this product may corresponded to the *dsg c* from other teleosts (data not shown). However this gene is not or is poorly expressed in sea bream larvae and thus was excluded from the q-RT-PCR analysis in the subsequent studies.

3.5 Tissue distribution of the *dsc* and *dsg* in gilthead sea bream

The EST analysis aimed to identify possible new transcripts but also to have an idea about the expression and distribution of the desmosomal transcripts in teleosts in order to identify the tissues in which they may play a key physiological role. Searches were performed for the two desmosomal cadherins family members (*dsc* and *dsg*) in

teleost ESTs database. The Annex Tables II.1 and II.2 contain the hits obtained and the respective local of expression of each EST. Overall, many ESTs were found for both transcripts and in general the members of this family seem to have a widespread distribution as ESTs were retrieved from a variety of tissues. Digital expression analysis revealed that a larger number of *dsc* ESTs (92 sequences) were identified in relation to *dsg* (23 sequences), suggesting that members of *dsc* may be more abundantly expressed.

Dsc exhibits an extensive expression in teleost fish, from larvae to several other tissues from spleen, intestines, testis, ovary, eye, skin, head, bone, fin, jaw, thymus among others. This wide expression was expected since a single *dsc* is present in teleosts and this gene encodes for an essential desmosome structural protein [36].

All the three *dsg* duplicates were found in the EST database in teleosts (Annex Table II.2). Fewer ESTs were found for dsg *b*, suggesting that this transcript is low abundance. *Dsg a* was the only duplicate with an EST in larva. Tissue expression in several sea bream tissues confirmed the digital expression profile given by ESTs and revealed that *dsc* has a widespread tissue distribution and was most abundant in liver, skin, kidney and gills In contrast *dsg b* was low abundance and weak amplification occurred in the gills, liver and duodenum (Figure 3.9).

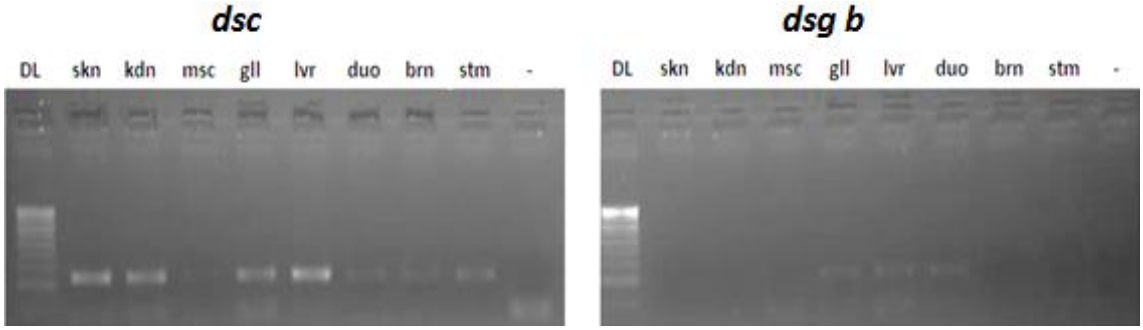


Figure 3.9**:** Tissue distribution by PCR of *dsc* and *dsg b* in sea bream. DL–DNA ladder; skn–skin; msc–muscle; gll–gills; lvr–liver; duo–duodenum; stm–stomach and negative control (–).

3.6 Ontogenic expression of *dsc* and *dsg b* in sea bream larvae

Only the expression of *dsc* and *dsg b* were studied in sea bream larvae (study 1, 2.1.1), as *dsg c* was found to be absent from larvae (data not shown) and the existence of *dsg a* still remains to be established. Q-RT-PCR was performed in larvae from 5, 15,

25, 35, 48, 60 dph and also at day 58 dph, when individuals were subjected to size sorting: larger (LA) and smaller (SM). *Ef1a*, *β-actin* and *rpl13a* expression were used to normalize the abundance of *dsc* and *dsg b*.

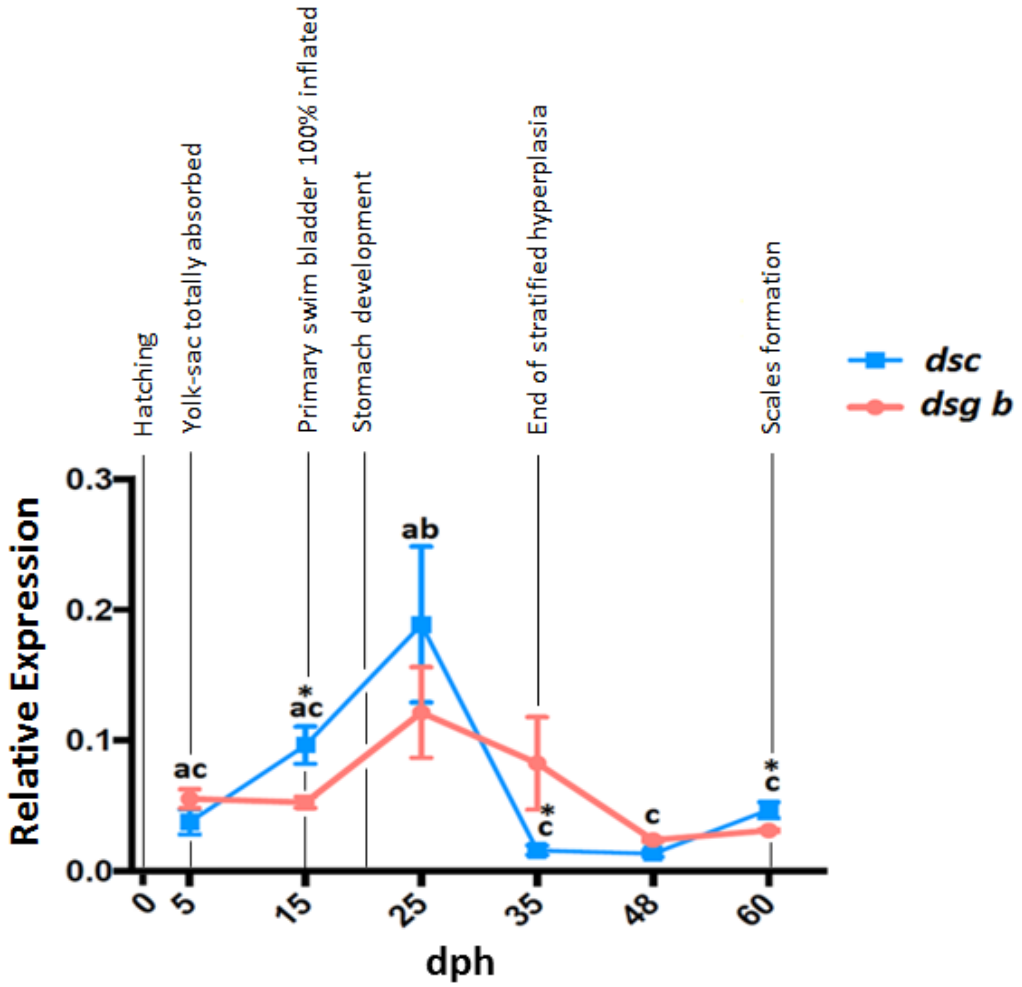


Figure 3.10: Relative expression of *dsc* and *dsg b* in gilthead sea bream larvae at 5, 15, 25, 35, 48 and 60 dph. 5 dph: *dsg b* n= 9; *dsc* n= 4. 15 dph: *dsg b* n= 9; *dsc* n= 12. 25 dph: *dsg b* n= 13; *dsc* n= 13. 35 dph: *dsg b* n= 15; *dsc* n= 15. 48 dph: *dsg b* n= 9; *dsc* n= 9. 60 dph: *dsg b* n= 7; *dsc* n= 7. The main morphological events on these dph are mapped. Statistically significant differences were assessed using two-way ANOVA. "*" indicates statistically significant differences ($p < 0.05$) between the expression of *dsc* and *dsg b*. Same lower case letters indicate *dsc* expression without significant differences ($p > 0.05$) between the dph, while different lower case letters indicate statistically significant *dsg b* expression differences on the dph ($p < 0.05$). *Dsg b* did not reveal significant ($p > 0.05$) differences in its expression during the period studied

Dsc and dsg b exhibit a similar expression pattern, and both genes are in general mostly expressed during the initial stages of development and their expression peaked on 25 dph (*dsc*, 0.189; *dsg b* 0.121). After this phase, both genes are down-

regulated and *dsc* exhibit the lowest levels of expression on 35 dph, whereas *dsg b* reached the lowest levels on 48 dph. The complete primary swim bladder inflation (15 dph), the end of stratified muscle hyperplasia and the formation of the scales (60 dph), were the morphological events that registered significant differences ($p < 0.05$) of expression between *dsc* and *dsg b*.

Expression of both genes was also verified at 58 dph when fish were sized graded and q-RT-PCR was performed in smaller and larger larvae to characterize desmosomal cadherin expression with larval growth.

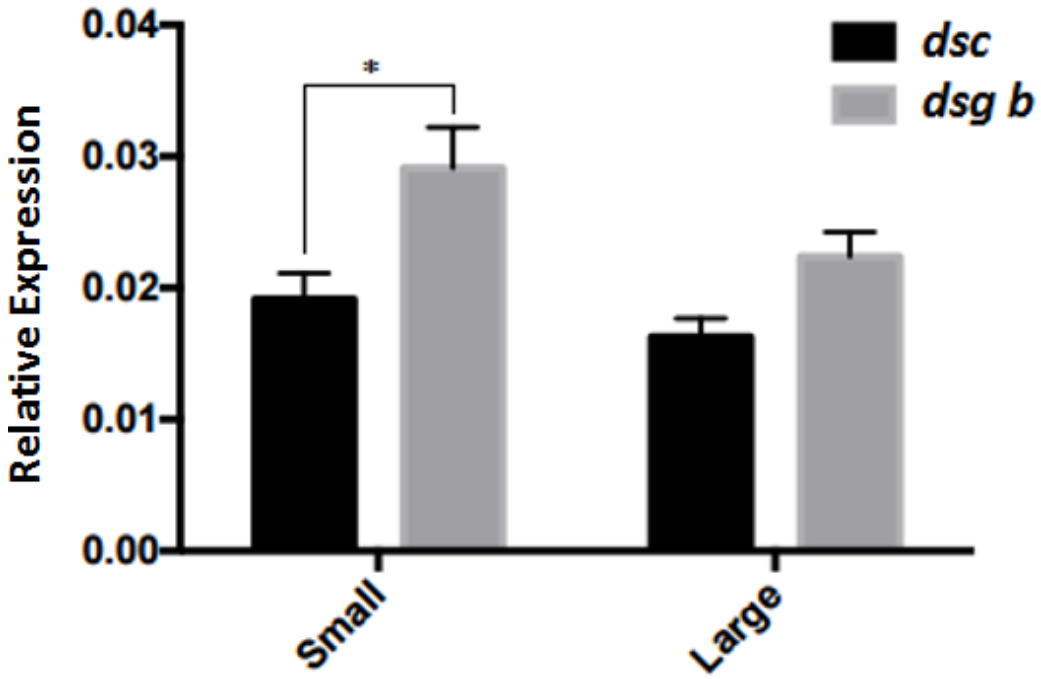


Figure 3.11: Relative expression of *dsc* and *dsg b* in sea bream larvae at 58 dph in individuals of different sizes. Statistical significances we assessed using t-test; "*" indicates $p < 0.05$. Large=0.0067g average weight and length=1.4-2.1 cm; Small=0.029g average weight and length=1.3-1.8 cm. Small: *dsg b* n= 10; *dsc* n= 10. Large: *dsg b* n= 10; *dsc* n= 10.

In Figure 3.11 it is evident that on 58 dph smaller fish larvae express higher levels of desmosomal cadherins (*dsc* and *dsg b*) when compared with larger individuals. In smaller fish, the difference of expression of *dsg b* and *dsc* was statistically significant and *dsg b* was the most expressed transcript ($p < 0.05$). No statistical differences were observed between the expression levels of the same gene in different sized larvae.

Correlations analysis of *dsc* and *dsg b* expression during larval development revealed that the expression of the two genes are correlated, which is expected since they are part of the same complex of proteins (Table 3.2). Correlation analysis was also performed with structural genes, myogenic and hormonal factors (*mlc2a*, *mlc2b*, *myog*, *mstn*, *col1a1*, *igf-2*, *fst* and *mrf4*) involved in muscle growth (Table 3.2).

Table 3.3: Correlation analysis of *dsc* and *dsg b* expression with other transcripts involved in teleost muscle growth *mlc2a*, *mlc2b*, *myog*, *mstn*, *col1a1*, *igf-2*, *fst* during the same time period (5, 15, 25, 35, 48 and 60 dph). The statistical test used was Pearson correlation and "*" indicates statistically significant correlation at a level of $p < 0.05$.

	mlc2a	***mlc2b***	***myog***	***mstn***	***col1a1***	***igf-2***	***fst***	***mrf4***	***dsg b***
dsc	-0,305	0,159	0,821*	0,058	0,162	0,714*	0,860*	-0,108	0,560*
dsg b	-0,189	0,096	0,729*	0,027	0,091	0,810*	0,762*	0,217	

The expression of *dsc* and *dsg b a*re tightly correlated with *myog, igf-2* and *fst* and when expression of *dsc* and *dsg b* increases/decreases, expression of *myog, igf-2* and *fst* follows the same trend suggesting that the desmosomal cadherins are regulated by similar processes with these three genes.

3.7 Expression of *dsc* and *dsg b* in sea bream larvae reared in mesocosm and intensive aquaculture systems

Larvae of 4, 15, 25, 45 and 81 dph were used to compare the expression of desmosomal cadherins in the different aquaculture systems (intensive vs mesocosm) and data was normalized using *rps18* and *rpl13a* transcripts.

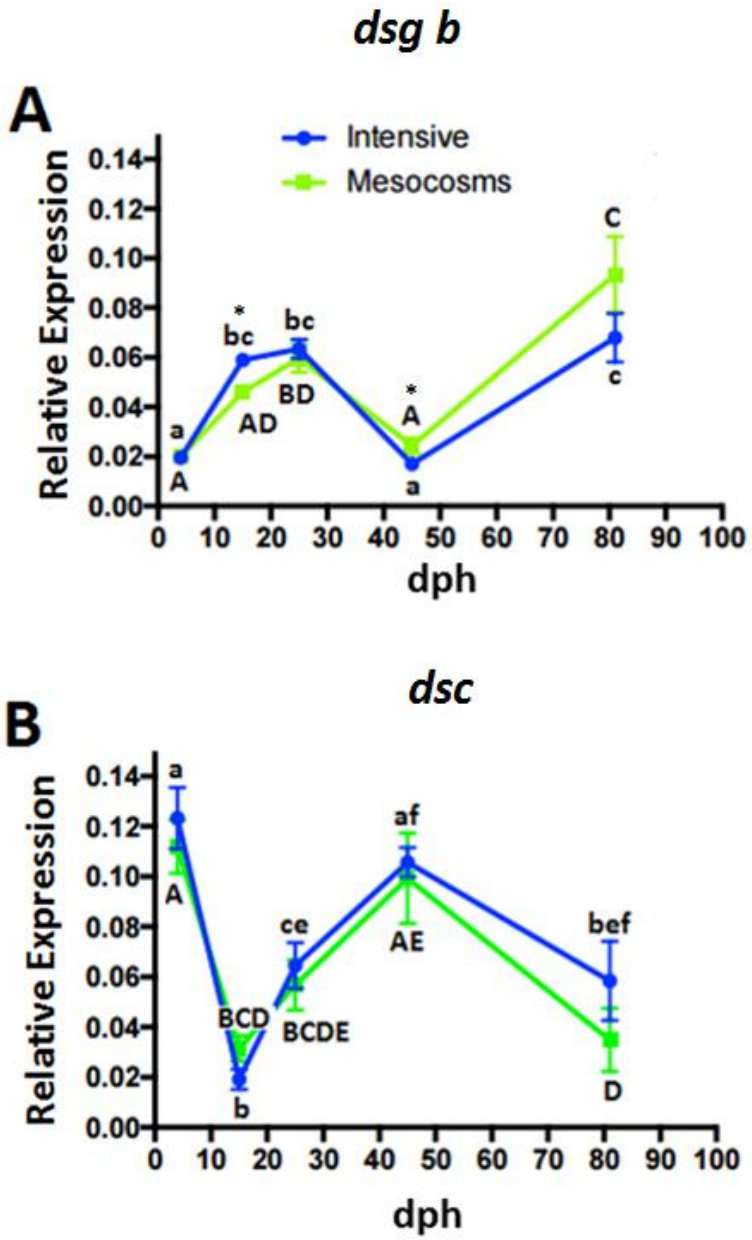


Figure 3.12: Relative expression analysis of the desmosomal cadherins (*dsc and dsg b*) in gilthead sea bream larvae at 4, 15, 25, 35, 45 and 81 dph reared in different aquaculture systems (intensive (I) and mesocosm (M)). Two-way ANOVA was performed to assess statistically significant differences. A: Expression of *dsg b*. 4 dph: I n=10; M n= 12. 15 dph: I n=12; M n= 9. 25 dph: I n=12; M n= 10. 45 dph: I n=12; M n= 10. 81 dph: I n=12; M n= 10. B: Expression of *dsc*. 4 dph: I n=12; M n= 10 .15 dph: I n= 9; M n= 10. 25 dph: I n=12; M n= 10. 45 dph: I n=12; M n= 11. 81 dph: I n= 7; M n= 8.Same lower case letters indicate gene expression without significant differences ($p > 0.05$) between the dph in the intensive system, while different lower case letters indicate statistically significant gene expression differences on the dph ($p < 0.05$) in the intensive system. Same upper case letters indicates gene expression without significate differences ($p > 0.05$) between the dph in the mesocosm system, while different upper case letters indicates statistically significate gene expression differences on the dph ($p < 0.05$) in the intensive system. "*" Indicates statically differences of *dsg b* expression between intensive and mesocosm systems ($p < 0.05$).

In both aquaculture systems (intensive and mesocosm) an up-regulation of *dsg b* was observed from 4 to 25 dph ($p < 0.05$), followed by a down-regulation to 45 dph and a subsequent up-regulation until 81 dph. On 15 dph and 45 dph *dsg b* expression between intensive and mesocosm systems was statistically significantly different. On 15 dph, *dsg b* levels were higher in intensive than in mesocosm but on 45dph the trend was reversed ($p < 0.05$) (Figure 3.12 A).

Dsc expression was down regulated on 15 dph ($p < 0.05$) compared with day 4 dph and subsequently increased ($p < 0.05$) until day 45 dph when it reached levels similar to day 4 dph to decrease again by 81 dph (Figure 3.12 B). No significant differences were observed on the expression of *dsc* between the two aquaculture systems suggesting that the type of hatchery system does not affect the expression of this gene.

4 Discussion

In this study, the homologues of the human desmosomal genes (*DSC*, *DSG*, *PG*, *DP* and *PKP*) were characterised in fish and the expression of desmosomal cadherins was investigated in sea bream.

Members of the desmosomal family have a distinct evolution trajectory in fish and in the teleost species-specific gene duplications were identified suggesting that they play a specific functional role in each species potentially associated with their adaptation.

Desmosomal cadherins were selected as candidate genes and *dsc*, *dsg b* and *dsg c* were isolated from the gilthead sea bream, however *dsg c* seem to be absent from the larval phase and *dsg a* remains to be found. Thus expression analysis studies in larvae were performed only for *dsc* and *dsg b*. The results revealed that *dsc* is expressed in all tissues analyzed while *dsg b* has a more limited expression, this distribution pattern matches with the teleost desmosomal cadherins EST search performed. Ontogenic expression revealed that *dsc* and *dsg b* exhibit a similar profile of expression. Expression of desmosomal cadherins in larvae was not affected in general by the rearing aquaculture systems. However, in larvae exhibiting heterogeneous growth, *dsg b* expression in smaller individual is significantly up-regulated ($p < 0.05$) when compared to *dsc* expression. Comparison of the *dsc* and *dsg b,* expression profile between the ontogenetic study and the aquaculture rearing studies revealed that while *dsg b* is similar, *dsc* is distinct and this may be a consequence of the genetic background of the larvae.

4.1 Fish desmosomal genes and evolution

In silico searches in different databases were performed and homologues of the human genes were found in all the fish species analyzed, this suggests that they have appeared early in the vertebrate evolution. The exception was lamprey where *dsg*, *pg* and *pkp* were not identified and this is likely to do with its incomplete genome annotation. *Dsc* seems to be the desmosomal gene with the simplest evolution, a single *dsc* was identified in the majority of the vertebrates. In mammals, three *dsc* genes were identified and they are the result of lineage specific gene duplication and

gene synteny analyses showed they resulted from tandem gene duplication. Dsc is a transmembrane protein that binds to Dsg [29] and is known to be one of the main proteins to generate adhesion between cells [36]. In contrast to *dsc, dsg* exhibit a more complex evolution as revealed by gene number identified, phylogeny and comparative gene synteny analysis. A highly variable number of genes were retrieved among the species and phylogenetic tree revealed that three main *dsg* clades exist in teleosts (*dsg a*, *dsg b* and *dsg c*). These results are surprising since elephant shark only possess a gene derived from the 1R and 2R, thus after the 3R occurred it was expected only two duplicates in teleosts. According to the tree topology the *dsg b* was the first to diverged and *dsg a* and *dsg c* were the descendent of a subsequent gene duplication event. Members of the *dsg c* family were only retrieved from Acanthomorpha fish (stickleback, tilapia, cod, medaka, amazon moly and platyfish). This is very intriguing as this comprises a large taxon of teleost fishes characterised by possessing spiny-rays and if this gene duplication is associated with this phenotypic characteristic in tissue organization remains to be further studied [83]. Acanthomorpha fish are the most successful group of vertebrates when looking to the evolution. A similar case was found in vitellogenin genes that suffered a post-R3 lineage-specific gene duplication, according to the authors the new duplicate suffered a neo-functionalization that was a key to the evolution and success of the teleosts in the oceanic environment, since it permitted the fish eggs to float and pre-adapt to the marine environment [84]. Somehow this new *dsg* duplicate may have played a similar role in the adaptation of Acanthomorpha fish and this remains to be further investigated. In addition the cave fish, amazon molly, zebrafish exhibit species-specific duplications, these organisms are all freshwater fish thus it is hypothesized that these specific duplication may be related to different adaptations to the environment. *Dsg* gene duplicates as a consequence of species-specific events were also identified in the spotted gar, cartilaginous fish and coelacanth. Similar, in tetrapods species-specific duplications events also occurred in chicken, xenopus and lizard and human. The reason why the evolution of *dsg* genes is very flexible and dependent of the lineage and specie is unknown and may be related to specific needs of each species according to their physiology and/or environment.

PG and DP belong to different families, armadillo and plakins respectively, but structurally they bind to each other in the ODP of the desmosome [29]. *Pg* and *dp* have

a similar evolution and both gene precursors duplicated in the teleosts (*a* and *b*). To date only a single *pg* has been described in the zebrafish and in this study *in silico* searches identified for the first time a novel gene for this family in the teleost that was designated by *pg b* [85]. β-catenin and N- and E-cadherins (proteins from adherent junctions) and *pg b* are abundantly synthesized and stored during oogenesis in zebra fish [85], PG is a component of both the adherents junction and the desmosomes [39], thus it is hypothesised that each may be specific of each cell junctions. In vertebrates, cells expressing PG with C-terminal truncations causes morphologic alterations in desmosomes [39], thus a higher number of proteins in teleost can assure a better stability of the desmosomal structure in case of abnormal proteins being produced. Moreover, PG has been found to be a major regulator of the expression of other desmosomal proteins [31], thus the higher number of *pg* in teleost may also ensure a better regulation of the expression of these proteins.

Similarly to *pg*, *dp a* and *b* have not been reported from fish and while in the majority of the teleost two genes were found, in the platyfish and amazon moly an extra gene copy of *dp a was* retrieved and the persistence of this gene in the genome of these species remains to be clarified. If it was the result of species-specific duplication or the homologue emerged during TSWGD and was subsequently deleted from the genomes of other teleosts.

Pkp is the family with more member and four different *pkp* were identified and members of *pkp4* and *pkp1* were reported for the first time in this study. All *pkp* emerged from a common ancestral *pkp* molecule and the tree topology suggests that *pkp4* was the first to have diverged. This is not surprising since literature describe this protein as a dual protein and different from the rest of the other PKP [53]. The existence of two duplicates in *pkp4* may be related to its dual function and each duplicate may be specific of each cell junctions. *Pkp2* in zebra fish was associated to dilated cardiomyopathy, the same gene in human is involved in the same pathology this reveled that conservations of the gene functions seem to have occurred throughout evolution process [86].

4.2 Expression of desmosomal cadherins in gilthead sea bream

In human, combination of desmosomal cadherins is distinct according to the cell type and cellular layer, a major example of this is the human skin (Figure 1.10) [36]. These types of combinations seem also to occur in teleosts and EST searches revealed that the *dsg* duplicates are potentially expressed in different tissues whereas the single *dsc* is present in all the tissues. Tissue distribution of *dsc* confirmed this and this gene was found to be expressed in all the tissues tested, whereas *dsg b* was only expressed in the gills, liver and duodenum. These results supported the information found in the ESTs database and thus to form the desmosomal structure in teleosts *dsc* must be always present and binds to *dsg a*, *dsg b* or *dsg c* eventually according to cell type and cellular layer or even developmental stage.

Previous experiments in zebrafish, revealed that *dsc* and *dsg a* are expressed since time zero while *dsg b* starts to be expressed at 2.25 hours post-fertilization [18]. Knockdown experiments of *dsc* and *dsg a* lead to severe phenotypes including: shortened body axis, severely reduced or absent head and or tail, absence of clearly defined somites and sometimes blebbing of the epidermis [80]. This was associated to desmosomal adhesion reduction [18]. In the gilthead sea bream *dsc* and *dsg b* were also found to be expressed in larvae since early stages of development however *dsg c* was not amplified and may be potentially not associated with larval growth.

During larval development, 25 dph was the time period that registered the highest expression level of both *dsc* and *dsg b*. During the q-RT-PCR experiments the amplification curves of the samples tested for this dph, revealed two distinct groups based on the expression level of desmosomal cadherins (data not shown) suggesting that 25 dph may be a crucial period in the morphological development when a peak of expression of *dsc* and *dsg b* occurs, but some individuals may have a delay exhibiting lower levels of these desmosomal cadherins. This peak of expression may be associated to the stomach development [18] that occur at 20 dph, which is an organ where desmosomes play a crucial role [29].

Moreover when larvae of difference sizes were compared on 58dph, smaller larvae tended to express higher levels of *dsg b* compared with larger individuals. Also, in smaller individuals the expression of *dsg b* was statistically significantly higher than

dsc. In contrast, the level of expression of *dsc* was not different between sizes. This result is not surprising since a single *dsc* exits in teleosts and its presence is crucial for the formation of the desmosome complex. On the other hand, variation in the expression of *dsg* are likely to occur during development. These results may also be promising to validate the desmosomal cadherins as molecular markers of growth performance.

Correlation analysis suggests that *dsc* and *dsg b* expression revealed that these genes are correlated confirming their importance in the formation of the desmosomal complex [36]. Expression of *dsc* and *dsg b a*re tightly correlated, with *myog, igf-2* and *fst,* the first regulatory genes that are implicated in the regulation of cadherins in fish. *Myog* plays a regulatory role in myogenesis and muscle cell differentiation [87]. Igf-2 is a protein with a critical role in growth and development that promotes cell proliferation in many different tissues [88]. *Fst* is also a gene regulatory of the muscle growth being an antagonist to myostatin, thus inhibits the excessive muscle growth [89]. This reveals an interesting association between desmosomal cadherins and fish muscle development.

After hatching fish larvae are characterized by an intensive hyperplastic period and it was observed that gilthead sea bream larvae exhibits an axial growth constant from 5 to 60 dph and hyperplasia dominates between 15-25 dph when hypertrophy also initiates. This observation interestingly matches with the period where the expression of the desmosomal cadherins was, in general, higher [25]. Thus, desmosomal cadherins seem to be associated with the recruitment of new muscle fibers. This would explain the higher levels of these genes in smaller individuals than the larger ones and opens the possibility to combine expression of both genes for the identification of future juveniles with smaller sizes.

The most surprising results were the divergence in the expression profiles of *dsc* in ontogenic expression (study 1) and in the aquaculture systems (study 2). While *dsc* expression in study 1 peaked on 25 dph and there is an increasing expression of the gene since day 5 dph, in study 2 and in both mesocosm and intensive system there is a massive down-regulation from day 4 dph to 10 dph. This may be explained having in consideration the distinct genetic background of the larvae stock utilized for each experiment.

5 Conclusion

Homologues of the human desmosomal genes (*DSC*, *DSG*, *PG*, *DP* and *PKP*) were characterised in several fish species and teleosts possess a single *dsc*; three *dsg* (*a*,*b* and *c*); two *pg* (*a* and *b*); two *dp* (*a* and *b*); seven *pkp* (*1a*, *1b*, *2*, *3a*, *3b*, *4a* and *4b*). The persistence of multiple genes is the result of teleost specie-specific gene duplications suggesting that they may play a specific functional role in each species potentially associated with their adaptation to different environments. *Dsc* and *dsg b* are expressed since early larval stages and both persist in to adult stage and while *dsc* was expressed in all tissue, *dsg b* was restricted to certain tissues suggesting that distinct desmosomal cadherins complex may exist and they are specific to each tissue. *Dsg b* seem to be associated with growth and in smaller individual was significantly up-regulated ($p < 0.05$) when compared with *dsc*. This opens the opportunity to validate these genes as new molecular markers of growth performance in aquaculture. Comparison of the *dsc* expression profile between the ontogenetic study and the aquaculture rearing studies revealed that genetic background of the larvae may influence the expression of this gene. Expression of desmosomal cadherins in larvae reared in different aquaculture systems was not affected suggesting that cell integrity was not compromised.

6 Future work

To complete this work it would be essential to isolate *dsg a,* in order to have the complete set of desmosomal cadherins of gilthead sea bream. Due the inexistence of transcripts of this gene in the *Sparus aurata* database, it would be necessary to follow the traditional strategy to isolate this gene. Also a histological analysis would enrich the results to observe if morphological alterations occur associated to the different genes expression. To better understand the full morphologic role of these proteins it would be interesting to perform expression analysis in several tissues of larvae. To assess the real potential of desmosomal cadherins as molecular markers of growth performance it would be necessary to expand the number of samples and to perform the same experiment in more dph. Despite the fact that in vivo extracellular Ca^{2+} concentration is assumed to be always well above to the concentration that is required to regulate desmosomes (+/- 0.1 mM) [37], a calcium supplementation could promote a better stability of the desmosomal structure and consequently of the larva health, thus it would be an interesting experiment to try after this study. In a long term, isolation of all the desmosomal members would fully complete this project and give the entire vision how important these genes are in teleosts.

7 Bibliography

[1] M. Landrau, *Introduction to Aquaculture*. New York: John Wiley & Sons, 1992.

[2] A. Dunham, *Aquaculture and Fisheries Biotechnology. Genetic Approaches*. Alabama: CABI Publishing, 2004.

[3] J. Muir, "Managing to harvest? Perspectives on the potential of aquaculture.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 360, no. 1453, pp. 191–218, 2005.

[4] C. Clemmesen, "The effect of food availability, age or size on the RNA/DNA ratio of individually measured herring larvae: laboratory calibration," *Mar. Biol.*, vol. 118, no. 3, pp. 377–382, 1994.

[5] Food and Agriculture Organization (FAO), "Global aquaculture production.," *FAO Fish. Stat. Collect.*, no. October, 2015.

[6] Food and Agriculture Organization (FAO), *Fishery and Aquaculture Statistics*. 2014.

[7] J. Bostock, B. McAndrew, R. Richards, K. Jauncey, T. Telfer, K. Lorenzen, D. Little, L. Ross, N. Handisyde, I. Gatward, and R. Corner, "Aquaculture: global status and trends.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 365, no. 1554, pp. 2897–2912, 2010.

[8] Food and Agriculture Organization (FAO), "FishStat fishery statistical collections: aquaculture production (1950 –2008; released March 2010). Rome, Italy: Food and Agriculture Organization of the United Nations.," 2010.

[9] P. Divanach and M. Kentouri, "Hatchery techniques for specific diversification in Mediterranean finfish larviculture," *Recent Adv. Mediterr. Aquac. finfish species Diversif.*, vol. 47, pp. 75–87, 2000.

[10] Food and Agriculture Organization of the United Nations, "Cultured Aquatic Species Information Programme Sparus aurata (Linnaeus, 1758)," *Food Agric. Organ. United Nations - Fish. Aquac. Dep.*, p. 722, 2009.

[11] D. D. Benetti, "Mesocosm Systems For Semi-Intensive Larval Rearing Of Marine Fish," *The Advocate*, pp. 17–18, 2001.

[12] R. J. Shields, "Larviculture of marine finfish in Europe," *Aquaculture*, vol. 200, no. 1–2, pp. 55–88, 2001.

[13] M. E. Cunha, H. Quental-Ferreira, C. Boglione, E. Palamara, P. Gavaia, and P. Pousão-Ferreira, "Rearing fish larvae for extensive the ' natural ' mesocosms system," vol. 35, no. June, pp. 23–26, 2010.

[14] “Fisheries and Aquaculture Statistics,” *FAO - Food Agric. Organ. United Nations*, 2013.

[15] AquaMaps, “Reviewed distribution maps for Sparus aurata (Gilthead seabream), with modelled year 2100 native range map based on IPCC A2 emissions scenario. www.aquamaps.org, version of Aug. 2013. Web. Accessed 8 Feb. 2016.,” 2013.

[16] FishDatabase, “Fish Data Base - Sparus aurata,” 2016. .

[17] M. P. Moretti, Alessandro; Fernandez-Criado and R. Vetillart, *Manual on Hatchery Production of Seabass and Gilthead Seabream.* Rome, 1991.

[18] Jensen G., “Sorting and grading warmwater fish,” *South. Rgeinal Aquac. Cent.*, no. 391, p. 8, 1990.

[19] S. Georgiou, P. Makridis, D. Dimopoulos, D. M. Power, Z. Mamuris, and K. a. Moutou, “Myosin light chain 2 isoforms in gilthead sea bream (Sparus aurata L.): Molecular growth markers at early stages,” *Aquaculture*, vol. 432, pp. 434–442, 2014.

[20] F. a Huntingford, N. B. Metcalfe, J. E. Thorpe, W. D. Graham, and C. E. Adams, “Social dominance and body size in Atlantic salmon parr, Salmo solar L.,” *J. Fish Biol.*, vol. 36, no. 6, pp. 877–881, 1990.

[21] S. Zhang, H. Yang, and L. Singh, “Increased information leakage from text,” *CEUR Workshop Proc.*, vol. 1225, no. January 2003, pp. 41–42, 2014.

[22] G. K. Wallat, L. G. Tiu, H. P. Wang, D. Rapp, and A. C. Leighfield, “The Effects of Size Grading on Production Efficiency and Growth Performance of Yellow Perch in Earthen Ponds,” *N. Am. J. Aquac.*, vol. 67, pp. 34–41, 2005.

[23] L. M. P. Valente, K. a. Moutou, L. E. C. Conceição, S. Engrola, J. M. O. Fernandes, and I. a. Johnston, “What determines growth potential and juvenile quality of farmed fish species?,” *Rev. Aquac.*, vol. 5, no. SUPPL.1, 2013.

[24] A. Mangoula, “Application of molecular markers to aquaculture and broodstock management with special emphasis on microsatellite DNA,” *CIHEAM Options Mediterrannes*, vol. 168, pp. 153–168, 1998.

[25] S. Georgiou, H. Alami-Durante, D. M. Power, E. Sarropoulou, Z. Mamuris, and K. a. Moutou, “Transient up- and down-regulation of expression of myosin light chain 2 and myostatin mRNA mark the changes from stratified hyperplasia to muscle fiber hypertrophy in larvae of gilthead sea bream (Sparus aurata L.),” *Cell Tissue Res.*, vol. 363, no. 2, pp. 541–554, 2016.

[26] A. P. Kowalczyk and K. J. Green, “Structure, Function and Regulation of Desmosomes,” vol. 18, no. 9, pp. 1199–1216, 2013.

[27] G. R. Owen and D. L. Stokes, “Exploring the Nature of Desmosomal Cadherin Associations in 3D,” *Dermatol. Res. Pract.*, vol. 2010, pp. 1–12, 2010.

[28] X. Cheng, Z. Den, and P. J. Koch, “Desmosomal cell adhesion in mammalian development,” *Eur. J. Cell Biol.*, vol. 84, pp. 215–223, 2005.

[29] X. Cheng and P. J. Koch, “In vivo function of desmosomes.,” *J. Dermatol.*, vol. 31, pp. 171–187, 2004.

[30] D. Garrod and M. Chidgey, “Desmosome structure, composition and function,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1778, no. 3, pp. 572–587, 2008.

[31] A. Schmidt and P. J. Koch, “Desmosomes Just Cell Adhesion Or Is There More?,” *Cell Adh. Migr.*, vol. 1, no. 1, pp. 28–32, 2007.

[32] K. J. Green and C. L. Simpson, “Desmosomes: New Perspectives on a Classic,” *J. Invest. Dermatol.*, vol. 127, no. 11, pp. 2499–2515, 2007.

[33] J. North, W. G. Bardsley, J. Hyam, E. a Bornslaeger, H. C. Cordingley, B. Trinnaman, M. Hatzfeld, K. J. Green, a I. Magee, and D. R. Garrod, “Molecular map of the desmosomal plaque.,” *J. Cell Sci.*, vol. 112 (Pt 2, pp. 4325–4336, 1999.

[34] D. Garrod, “Desmosomes In Vivo,” *Dermatol. Res. Pract.*, vol. 2010, pp. 1–17, 2010.

[35] E. Delva, D. K. Tucker, and A. P. Kowalczyk, “The Desmosome,” vol. 1, no. 2, p. 17, 2009.

[36] D. R. Garrod, A. J. Merritt, and Z. Nie, “Desmosomal cadherins,” *Curr. Opin. Cell Biol.*, vol. 14, no. 5, pp. 537–545, 2002.

[37] J. Brasch, O. J. Harrison, B. Honing, and L. Shapiro, “Thinking outside the cell: how cadherins drive adhesion,” *Trends Cell Biol.*, vol. 18, no. 9, pp. 1199–1216, 2013.

[38] M. Saito, D. K. Tucker, D. Kohlhorst, C. M. Niessen, and a. P. Kowalczyk, “Classical and desmosomal cadherins at a glance,” *J. Cell Sci.*, vol. 125, no. 11, pp. 2547–2552, 2012.

[39] H. L. Palka and K. J. Green, “Roles of plakoglobin end domains in desmosome assembly.,” *J. Cell Sci.*, vol. 110 (Pt 1, pp. 2359–71, 1997.

[40] C. Bierkamp, K. J. Mclaughlin, H. Schwarz, O. Huber, and R. Kemler, “Embryonic heart and skin defects in mice lacking plakoglobin.,” *Dev. Biol.*, vol. 180, no. 2, pp. 780–5, Dec. 1996.

[41] M. Hatzfeld, “Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion?,” *Biochim. Biophys. Acta*, vol. 1773, no. 1, pp. 69–77, 2007.

[42] M. Peifer, P. D. McCrea, K. J.Green, E. Wieschaus, and B. M.Gumbiner, “The Vertebrate Adhesive Junction Proteins -catenin and Plakoglobin and the Drosophila Segment Polarity Gene armadillo Form a Multigene Family with Similar Properties,” *J. Cell Biol.*, vol. 121, no. 5, pp. 1133–1140, 1993.

[43] H.-J. Choi, J. C. Gross, S. Pokutta, and W. I. Weis, “Interactions of plakoglobin and beta-catenin with desmosomal cadherins: basis of selective exclusion of alpha- and beta-catenin from desmosomes,” *J. Biol. Chem.*, vol. 284, no. 46, pp. 31776–31788, 2009.

[44] R. H. Carnahan, A. Rokas, E. a Gaucher, and A. B. Reynolds, “The Molecular Evolution of the p120-Catenin Subfamily and Its Functional Associations.,” *PLoS One*, vol. 5, no. 12, p. e15747, 2010.

[45] A. P. Kowalczyk, M. Hatzfeld, E. a. Bornslaeger, D. S. Kopp, J. E. Borgwardt, C. M. Corcoran, A. Settler, and K. J. Green, “The head domain of plakophilin-1 binds to desmoplakin and enhances its recruitment to desmosomes. Implications for cutaneous disease,” *J. Biol. Chem.*, vol. 274, no. 26, pp. 18145–18148, 1999.

[46] S. Breuninger, S. Reidenbach, C. G. Sauer, P. Ströbel, J. Pfitzenmaier, L. Trojan, and I. Hofmann, “Desmosomal plakophilins in the prostate and prostatic adenocarcinomas: implications for diagnosis and tumor progression.,” *Am. J. Pathol.*, vol. 176, no. 5, pp. 2509–2519, 2010.

[47] S. Neuber, M. Mühmer, D. Wratten, P. J. Koch, R. Moll, and A. Schmidt, “The desmosomal plaque proteins of the plakophilin family,” *Dermatol. Res. Pract.*, vol. 2010, no. 1, 2010.

[48] A. P. South, H. Wan, M. G. Stone, P. J. C. Dopping-Hepenstal, P. E. Purkis, J. F. Marshall, I. M. Leigh, R. a J. Eady, I. R. Hart, and J. a McGrath, “Lack of plakophilin 1 increases keratinocyte migration and reduces desmosome stability.,” *J. Cell Sci.*, vol. 116, pp. 3303–3314, 2003.

[49] C. Mertens, C. Kuhn, and W. W. Franke, “Plakophilins 2a and 2b: constitutive proteins of dual location in the karyoplasm and the desmosomal plaque.,” *J. Cell Biol.*, vol. 135, no. 4, pp. 1009–25, Nov. 1996.

[50] L. Antoniades, A. Tsatsopoulou, A. Anastasakis, P. Syrris, A. Asimaki, D. Panagiotakos, C. Zambartas, C. Stefanadis, W. J. McKenna, and N. Protonotarios, “Arrhythmogenic right ventricular cardiomyopathy caused by deletions in plakophilin-2 and plakoglobin (Naxos disease) in families from Greece and Cyprus: Genotype-phenotype relations, diagnostic features and prognosis,” *Eur. Heart J.*, vol. 27, no. 18, pp. 2208–2216, 2006.

[51] A. Schmidt, L. Langbein, M. Rode, S. Prätzel, R. Zimbelmann, and W. W. Franke, “Plakophilins 1a and 1b: widespread nuclear proteins recruited in specific epithelial cells as desmosomal plaque components.,” *Cell Tissue Res.*, vol. 290, no. 3, pp. 481–99, Dec. 1997.

[52] M. Hatzfeld, K. J. Green, and H. Sauter, “Targeting of p0071 to desmosomes and adherens junctions is mediated by different protein domains.,” *J. Cell Sci.*, vol. 116, no. Pt 7, pp. 1219–33, 2003.

[53] I. Hofmann, T. Schlechter, C. Kuhn, M. Hergt, and W. W. Franke, “Protein p0071 - an armadillo plaque protein that characterizes a specific subtype of adherens junctions.,” *J. Cell Sci.*, vol. 122, pp. 21–24, 2009.

[54] J. J. Jefferson, C. Ciatto, L. Shapiro, and R. K. H. Liem, “Structural Analysis of the Plakin Domain of Bullous Pemphigoid Antigen1 (BPAG1) Suggests that Plakins Are Members of the Spectrin Superfamily,” *J. Mol. Biol.*, vol. 366, no. 1, pp. 244–257, 2007.

[55] H.-J. Choi, S. Park-Snyder, L. T. Pascoe, K. J. Green, and W. I. Weis, “Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure.,” *Nat. Struct. Biol.*, vol. 9, no. 8, pp. 612–620, 2002.

[56] T. S. Stappenbeck, J. a Lamb, C. M. Corcoran, and K. J. Green, “Phosphorylation of the desmoplakin COOH terminus negatively regulates its interaction with keratin intermediate filament networks.,” *J. Biol. Chem.*, vol. 269, no. 47, pp. 29351–4, 1994.

[57] B. D. Angst, L. a Nilles, and K. J. Green, “Desmoplakin II expression is not restricted to stratified epithelia.,” *J. Cell Sci.*, vol. 97 (Pt 2), no. 1986, pp. 247–57, 1990.

[58] D. Keith, B. Armstrong, K. E. McKenna, P. E. Purkis, K. J. Green, R. a J. Eady, I. M. Leigh, and A. E. Hughes, “Haploinsufficiency of desmoplakin causes a striate subtype of palmoplantar keratoderma,” *Hum. Mol. Genet.*, vol. 8, no. 1, pp. 143–148, 1999.

[59] N. V. Whittock, G. H. S. Ashton, P. J. C. Dopping-Hepenstal, M. J. Gratian, F. M. Keane, R. a J. Eady, and J. a. McGrath, “Striate palmoplantar keratoderma resulting from desmoplakin haploinsufficiency,” *J. Invest. Dermatol.*, vol. 113, no. 6, pp. 940–946, 1999.

[60] E. E. Norgett, S. J. Hatsell, L. Carvajal-Huerta, J.-C. R. Cabezas, J. Common, P. E. Purkis, N. Whittock, I. M. Leigh, H. P. Stevens, and D. P. Kelsell, “Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma,” *Hum. Mol. Genet.*, vol. 9, no. 18, pp. 2761–2766, 2000.

[61] N. V. Whittock, H. Wan, S. M. Morley, M. C. Garzon, L. Kristal, P. Hyde, W. H. I. McLean, L. Pulkkinen, J. Uitto, A. M. Christiano, R. a J. Eady, and J. a. McGrath, "Compound heterozygosity for non-sense and mis-sense mutations in desmoplakin underlies skin fragility/woolly hair syndrome," *J. Invest. Dermatol.*, vol. 118, no. 2, pp. 232–238, 2002.

[62] J. S. Nelson, *Fishes of the World*, 3rd ed. New York: John Wiley and Sons, 1994.

[63] J.-N. Volff, "Genome evolution and biodiversity in teleost fish.," *Heredity (Edinb).*, vol. 94, no. 3, pp. 280–294, 2005.

[64] J. C. Opazo, G. T. Butts, M. F. Nery, J. F. Storz, and F. G. Hoffmann, "Whole-genome duplication and the functional diversification of teleost fish hemoglobins," *Mol. Biol. Evol.*, vol. 30, no. 1, pp. 140–153, 2013.

[65] D. Steinke, S. Hoegg, H. Brinkmann, and A. Meyer, "Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates.," *BMC Biol.*, vol. 4, p. 16, 2006.

[66] S. M. K. Glasauer and S. C. F. Neuhaus, "Whole-genome duplication in teleost fishes and its evolutionary consequences," *Mol. Genet. Genomics*, 2014.

[67] S. Kuraku, A. Meyer, and S. Kuratani, "Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after?," *Mol. Biol. Evol.*, vol. 26, no. 1, pp. 47–59, 2009.

[68] J. C. R. Cardoso, R. C. Félix, C. a. Bergqvist, and D. Larhammar, "New insights into the evolution of vertebrate CRH (corticotropin-releasing hormone) and invertebrate DH44 (diuretic hormone 44) receptors in metazoans," *Gen. Comp. Endocrinol.*, vol. 209, pp. 162–170, 2014.

[69] P. Kafarski, "Rainbow code of biotechnology," *Chemik*, vol. 66, no. 8, pp. 811–816, 2012.

[70] M. Eroglu S., Toprak S., Urgan O, MD, Ozge E. Onur, MD, Arzu Denizbasi, MD, Haldun Akoglu, MD, Cigdem Ozpolat, MD, Ebru Akoglu, *Basic Biotechnology*, 2nd ed., vol. 33. Cambridge: Cambridge University Press, 2001.

[71] K. Sambamurthy and K. Ahutosh, *Pharmaceutical Biotechnology*. New Delhi: New Age International Publishers, 2006.

[72] T. S. Mayekar, A. A. Salgaonkar, J. M. Koli, P. R. Patil, C. Ajit, N. Pawar, S. Kamble, A. Giri, G. G. Phandke, and P. Kapse, "Biotechnology and its aplications in aquaculture and fisheries," *Aquafind - Aquat. Fsh Database*.

[73] B. Louro, J. Marques, D. Power, and A. Canário, "Having a BLAST: Searchable transcriptome resources for the gilthead sea bream and the European sea bass," *Mar. Genomics*, 2016.

[74] F. Abascal, R. Zardoya, and D. Posada, “ProtTest: Selection of best-fit models of protein evolution,” *Bioinformatics*, vol. 21, no. 9, pp. 2104–2105, 2005.

[75] A. Massey and K. Hellen, *Genetic Engineering and Biotechnology*, 2nd ed. Porto Alegre: Artemed, 2002.

[76] P. Y. Lee, J. Costumbrado, C.-Y. Hsu, and Y. H. Kim, “Agarose gel electrophoresis for the separation of DNA fragments.,” *J. Vis. Exp.*, no. 62, pp. 1–5, 2012.

[77] Stratagene, *Introduction to Quantitative PCR - Methods and Applications Guide*. California, 2007.

[78] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman, “Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.,” *Genome Biol.*, vol. 3, no. 7, 2002.

[79] K. A. Moutou, M. Codina, S. Georgiou, J. Gutiérrez, and Z. Mamuris, “Myosin light chain 2 in gilthead sea bream (Sparus aurata): a molecular marker of muscle development and growth,” no. 2001, p. 3018, 2008.

[80] A. Goonesinghe, X.-M. Luan, A. Hurlstone, and D. Garrod, “Desmosomal cadherins in zebrafish epiboly and gastrulation,” *BMC Dev. Biol.*, vol. 12, no. 1, p. 1, 2012.

[81] M. Takamiya, B. D. Weger, S. Schindler, T. Beil, L. Yang, O. Armant, M. Ferg, G. Schlunck, T. Reinhard, T. Dickmeis, S. Rastegar, U. Strähle, and Y. F. Leung, “Molecular description of eye defects in the zebrafish pax6b mutant, sunrise, reveals a pax6b-dependent genetic network in the developing anterior chamber,” *PLoS One*, vol. 10, no. 2, pp. 1–23, 2015.

[82] J. I. Dermatol, “HHS Public Access,” vol. 133, no. 1, pp. 68–77, 2013.

[83] W. J. Chen, C. Bonillo, and G. Lecointre, “Repeatability of clades as a criterion of reliability: A case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa,” *Mol. Phylogenet. Evol.*, vol. 26, no. 2, pp. 262–288, 2003.

[84] R. N. Finn and B. a. Kristoffersen, “Vertebrate vitellogenin gene duplication in relation to the ‘3R hypothesis’: Correlation to the pelagic egg and the oceanic radiation of teleosts,” *PLoS One*, vol. 2, no. 1, 2007.

[85] J. Cerdà, S. Reidenbach, S. Prätzel, and W. W. Franke, “Cadherin-catenin complexes during zebrafish oogenesis: heterotypic junctions between oocytes and follicle cells.,” *Biol. Reprod.*, vol. 61, pp. 692–704, 1999.

[86] Y. H. Shih, Y. Zhang, Y. Ding, C. a. Ross, H. Li, T. M. Olson, and X. Xu, “Cardiac Transcriptome and Dilated Cardiomyopathy Genes in Zebrafish,” *Circ. Cardiovasc. Genet.*, vol. 8, no. 2, pp. 261–269, 2015.

[87] J. P. Brockes and A. Kumar, “Appendage regeneration in adult vertebrates and implications for regenerative medicine.,” *Science*, vol. 310, no. 5756, pp. 1919–23, 2005.

[88] “igf 2 gene,” *Natl. Libr. Med.*, pp. 1–6.

[89] H. Ogino, S. Yano, S. Kakiuchi, H. Muguruma, K. Ikuta, M. Hanibuchi, H. Uehara, K. Tsuchida, H. Sugino, and S. Sone, “Follistatin suppresses the production of experimental multiple-organ metastasis by small cell lung cancer cells in natural killer cell-depleted SCID mice,” *Clin. Cancer Res.*, vol. 14, no. 3, pp. 660–667, 2008.

8 Annexes

Annex I

Tables containing the gene and respective protein accession numbers of the desmosomal members found in the *in silico* searches in the genomes databases. All the species contain an abbreviation in order to facilitate their representation in the phylogenetic trees.

Table Annex I.1**:** Genes and respective proteins accession numbers of the *dsc* retrieved from the genomes databases searches.

Species	*Abbreviation*	*Gene accession number*	*Protein accession number*
Human	Hsa	ENSG00000134765	ENSP00000257198
		ENSG00000134755	ENSP00000280904
		ENSG00000134762	ENSP00000353608
Opossum	Mdo	ENSMODG00000011082	ENSMODP00000013875
		ENSMODG00000011103	ENSMODP00000013899
		ENSMODG00000011124	ENSMODP00000013923
Chicken	Gga	ENSGALG00000015140	ENSGALP00000024384
Anole lizard	Aca	ENSACAG00000017830	ENSACAP00000017561
		ENSACAG00000017838	ENSACAP00000017565
Xenopus	Xtr	ENSXETG00000004721	ENSXETP00000010240
Coelacanth	Lch	ENSLACG00000002907	ENSLACP00000003250
Tetraodon	Tni	ENSTNIG00000006258	ENSTNIP00000009011
Fugu	Tru	ENSTRUG00000007388	ENSTRUP00000018268
Stickleback	Gac	ENSGACG00000013163	ENSGACP00000017408
Sea bass	Dla	DLAgn_00006120	DLAgn_00006120
Tilapia	Oni	ENSONIG00000009697	ENSONIP00000012179
Medaka	Ola	ENSORLG00000017106	ENSORLP00000021406
Platyfish	Xma	ENSXMAG00000016712	ENSXMAP00000016773
Amazon moly	Pfo	ENSPFOG00000001588	ENSPFOP00000001630
Cod	Cod	ENSGMOG00000013140	ENSGMOP00000014060
Zebra fish	Dre	ENSDARG00000039677	ENSDARP00000090832
Cave fish	Ame	ENSAMXG00000002361	ENSAMXP00000002450
Spotted gar	Loc	ENSLOCG00000006830	ENSLOCP00000008289
Elephant shark	Cmi	SINCAMG00000014073	SINCAMP00000021706
Lamprey	Pma	ENSPMAG00000002057	ENSPMAP00000002274

Table Annex I.2: Genes and respective proteins accession numbers of the *dsg* retrieved from the genomes databases searches.

Species	*Abbreviation*	*Gene accession number*	*Protein accession number*
Human	Hsa	ENSG00000134760	ENSP00000257192
		ENSG00000046604	ENSP00000261590
		ENSG00000134757	ENSP00000257189
		ENSG00000175065	ENSP00000352785
Opossum	Mdo	ENSMODG00000011042	ENSMODP00000013830
		ENSMODG00000010945	ENSMODP00000013697
		ENSMODG00000011018	ENSMODP00000013762
Chicken	Gga	ENSGALG00000015142	ENSGALP00000024387
		ENSGALG00000017398	ENSGALP00000036016
Anole lizard	Aca	ENSACAG00000017850	ENSACAP00000017577
		ENSACAG00000017842	ENSACAP00000017573
Xenopus	Xtr	ENSXETG00000034243	ENSXETP00000061971
		ENSXETG00000031232	ENSXETP00000058678
Coelacanth	Lch	ENSLACG00000002419	ENSLACP00000002704
		ENSLACG00000004808	ENSLACP00000005404
		ENSLACG00000012662	ENSLACP00000014385
Tetraodon	Tni	ENSTNIG00000010474	ENSTNIP00000013382
Fugu	Tru	ENSTRUG00000006266	ENSTRUP00000015305
		ENSTRUG00000005736	ENSTRUP00000013932
Stickleback	Gac	ENSGACG00000002084	ENSGACP00000002711
		ENSGACG00000013161	ENSGACP00000017402
		ENSGACG00000002160	ENSGACP00000002807
Sea bass	Dla	DLAgn_00006130	DLAgn_00006130
		DLAgn_00154370	DLAgn_00154370
Tilapia	Oni	ENSONIG00000007624	ENSONIP00000009609
		ENSONIG00000003673	ENSONIP00000004621
		ENSONIG00000009689	ENSONIP00000012169
		ENSONIG00000007798	ENSONIP00000009821
		ENSONIG00000007622	ENSONIP00000009606
Medaka	Ola	ENSORLG00000002986	ENSORLP00000003728
		ENSORLG00000017110	ENSORLP00000021410
		ENSORLG00000000110	ENSORLP00000000139
		ENSORLG00000002978	ENSORLP00000003718

Platyfish	Xma	ENSXMAG00000011111	ENSXMAP00000011137
		ENSXMAG00000016745	ENSXMAP00000016814
		ENSXMAG00000011118	ENSXMAP00000011145
Amazon Moly	Pfo	ENSPFOG00000001104	ENSPFOP00000001402
		ENSPFOG00000000749	ENSPFOP00000029272
		ENSPFOG00000000645	ENSPFOP00000000632
Cod	Gmo	ENSGMOG0000001393	ENSGMOP00000014922
		ENSGMOG0000001317	ENSGMOP00000014043
		ENSGMOG0000001475	ENSGMOP00000015788
Zebra fish	Dre	ENSDARG00000076426	ENSDARP00000103538
		ENSDARG00000062750	ENSDARP00000119724
		ENSDARG00000039665	ENSDARP00000057980
Cave fish	Ame	ENSAMXG00000002403	ENSAMXP00000002452
		ENSAMXG00000002421	ENSAMXP00000002474
		ENSAMXG00000002407	ENSAMXP00000002469
		ENSAMXG00000006213	ENSAMXP00000006373
Spotted gar	Loc	ENSLOCG00000006763	ENSLOCP00000008189
		ENSLOCG00000006796	ENSLOCP00000008214
		ENSLOCG00000006734	ENSLOCP00000008150
Elephant shark	Cmi	SINCAMG00000014170	SINCAMP00000021784
		SINCAMG00000014205	SINCAMP00000021856

Table Annex I.3: Genes and respective proteins accession numbers of the *pg* retrieved from the genomes databases searches.

Species	*Abbreviation*	*Gene accession number*	*Protein accession number*
Human	Hsa	ENSG00000173801	ENSP00000377508
Opossum	Mdo	ENSMODG00000014681	ENSMODP00000018353
Chicken	Gga	ENSGALG00000017414	ENSGALP00000028049
Anole lizard	Aca	ENSACAG00000017889	ENSACAP00000017621
Xenopus	Xtr	ENSXETG00000006970	ENSXETP00000015194
Coelacanth	Lch	ENSLACG00000005538	ENSLACP00000006244
Tetraodon	Tni	ENSTNIG00000004863	ENSTNIP00000007525
		ENSTNIG00000010216	ENSTNIP00000000864
Fugu	Tru	ENSTRUG00000000583	ENSTRUP00000001389
		ENSTRUG00000013363	ENSTRUP00000034058
Stickleback	Gac	ENSGACG00000018328	ENSGACP00000024232

		ENSGACG00000000333	ENSGACP00000000420
Sea bass	Dla	DLAgn_00177880	DLAgn_00177880
		DLAgn_00195460	DLAgn_00195460
Tilapia	Oni	ENSONIG00000018803	ENSONIP00000023673
		ENSONIG00000010944	ENSONIP00000013768
Medaka	Ola	ENSORLG00000012542	ENSORLP00000015702
Platyfish	Xma	ENSXMAG00000000209	ENSXMAP00000000212
Amazon moly	Pfo	ENSPFOG00000003394	ENSPFOP00000003396
Cod	Gmo	ENSGMOG00000019511	ENSGMOP00000020989
		ENSGMOG00000019390	ENSGMOP00000020855
Zebra fish	Dre	ENSDARG00000070787	ENSDARP00000026915
		ENSDARG00000059067	ENSDARP00000076517
Cave fish	Ame	ENSAMXG00000018337	ENSAMXP00000018907
		ENSAMXG00000005521	ENSAMXP00000005659
Spotted gar	Loc	ENSLOCG00000013664	ENSLOCP00000016854
Elephant shark	Cmi	SINCAMG00000015235	SINCAMP00000023437

Table Annex I.4: Genes and respective proteins accession numbers of the *dp* retrieved from the genomes databases searches.

Species	*Abbreviation*	*Gene accession number*	*Protein accession number*
Human	Hsa	ENSG00000096696	ENSP00000369129
Opossum	Mdo	ENSMODG00000009982	ENSMODP00000012498
Chicken	Gga	ENSGALG00000012790	ENSGALP00000020840
Anole lizard	Aca	ENSACAG00000002761	ENSACAP00000002721
Xenopus	Xtr	ENSXETG00000020444	ENSXETP00000062602
Coelacanth	Lch	ENSLACG00000005036	ENSLACP00000005667
Tetraodon	Tni	ENSTNIG00000015195	ENSTNIP00000018252
		ENSTNIG00000012317	ENSTNIP00000015285
Fugu	Tru	ENSTRUG00000004758	ENSTRUP00000011327
		ENSTRUG00000004870	ENSTRUP00000011630
Stickleback	Gac	ENSGACG00000013099	ENSGACP00000017327
		ENSGACG00000004923	ENSGACP00000006524
Sea bass	Dla	DLAgn_00006160	DLAgn_00006160
		DLAgn_00145480	DLAgn_00145480
Tilapia	Oni	ENSONIG00000009681	ENSONIP00000012159
		ENSONIG00000006386	ENSONIP00000008049

Medaka	Ola	ENSORLG00000017125	ENSORLP00000021426
		ENSORLG00000012791	ENSORLP00000016016
Platyfish	Xma	ENSXMAG00000016816	ENSXMAP00000016871
		ENSXMAG00000005421	ENSXMAP00000005427
		ENSXMAG00000011514	ENSXMAP00000011530
		ENSXMAG00000006281	ENSXMAP00000006297
Amazon molly	Pfo	ENSPFOG00000000696	ENSPFOP00000000765
		ENSPFOG00000001001	ENSPFOP00000001043
		ENSPFOG00000022504	ENSPFOP00000022991
Cod	Gmo	ENSGMOG00000013073	ENSGMOP00000014001
		ENSGMOG00000016860	ENSGMOP00000018173
Zebrafish	Dre	ENSDARG00000022309	ENSDARP00000031822
		ENSDARG00000076673	ENSDARP00000103914
Cave fish	Ame	ENSAMXG00000011053	ENSAMXP00000011366
		ENSAMXG00000001574	ENSAMXP00000001602
Spotted gar	Loc	ENSLOCG00000013051	ENSLOCP00000016086
Elephant shark	Cmi	SINCAMG00000000826	SINCAMP00000001272
Lamprey	Pma	ENSPMAG00000001658	ENSPMAP00000001819
		ENSPMAG00000002058	ENSPMAP00000002255

Table Annex I.5: Genes and respective proteins accession numbers of the *pkp* retrieved from the genomes databases searches.

Species	*Abbreviation*	*Gene accession number*	*Protein accession number*
Human	Hsa	ENSG00000081277	ENSP00000263946
		ENSG00000057294	ENSP00000070846
		ENSG00000184363	ENSP00000331678
		ENSG00000144283	ENSP00000374409
Opossum	Mdo	ENSMODG00000000329	ENSMODP00000000392
		ENSMODG00000012782	ENSMODP00000015991
		ENSMODG00000004479	ENSMODP00000035942
Chicken	Gga	ENSGALG00000000299	ENSGALP00000030824
		ENSGALG00000012913	ENSGALP00000021029
		ENSGALG00000004260	ENSGALP00000006763
		ENSGALG00000012561	ENSGALP00000020484
Anole lizard	Aca	ENSACAG00000003509	ENSACAP00000003475
		ENSACAG00000012265	ENSACAP00000012092

		ENSACAG00000013428	ENSACAP00000013241
		ENSACAG00000010447	ENSACAP00000010256
Xenopus	Xtr	ENSXETG00000017085	ENSXETP00000037240
		ENSXETG00000031884	ENSXETP00000062574
		ENSXETG00000006706	ENSXETP00000014670
		ENSXETG00000010376	ENSXETP00000022793
Coelacanth	Lch	ENSLACG00000011620	ENSLACP00000013195
		ENSLACG00000008628	ENSLACP00000023417
		ENSLACG00000014090	ENSLACP00000023600
Tetraodon	Tni	ENSTNIG00000015032	ENSTNIP00000018083
		ENSTNIG00000000548	ENSTNIP00000000697
		ENSTNIG00000018475	ENSTNIP00000021649
		ENSTNIG00000000568	ENSTNIP00000002827
		ENSTNIG00000013679	ENSTNIP00000016679
		ENSTNIG00000003425	ENSTNIP00000006012
		ENSTNIG00000016492	ENSTNIP00000019591
Fugu	Tru	ENSTRUG00000011845	ENSTRUP00000029943
		ENSTRUG00000016573	ENSTRUP00000042376
		ENSTRUG00000001418	ENSTRUP00000003262
		ENSTRUG00000012930	ENSTRUP00000032784
		ENSTRUG00000003274	ENSTRUP00000007659
		ENSTRUG00000006426	ENSTRUP00000015732
Stickleback	Gac	ENSGACG00000009752	ENSGACP00000012869
		ENSGACG00000005123	ENSGACP00000006769
		ENSGACG00000018804	ENSGACP00000024861
		ENSGACG00000005842	ENSGACP00000007727
		ENSGACG00000012660	ENSGACP00000016730
		ENSGACG00000005780	ENSGACP00000007634
		ENSGACG00000015057	ENSGACP00000019862
Sea bass	Dla	DLAgn_00130340	DLAgn_00130340
		DLAgn_00050280	DLAgn_00050280
		DLAgn_00096250	DLAgn_00096250
		DLAgn_00136730	DLAgn_00136730
		DLAgn_00154600	DLAgn_00154600
		DLAgn_00169150	DLAgn_00169150
Tilapia	Oni	ENSONIG00000012321	ENSONIP00000015520

		ENSONIG00000014703	ENSONIP00000018508
		ENSONIG00000003463	ENSONIP00000004360
		ENSONIG00000008268	ENSONIP00000010404
		ENSONIG00000008949	ENSONIP00000011248
		ENSONIG00000012299	ENSONIP00000015489
Medaka	Ola	ENSORLG00000008336	ENSORLP00000010468
		ENSORLG00000006317	ENSORLP00000007934
		ENSORLG00000013540	ENSORLP00000016971
		ENSORLG00000005896	ENSORLP00000007413
		ENSORLG00000001869	ENSORLP00000002327
		ENSORLG00000016767	ENSORLP00000020977
		ENSORLG00000005902	ENSORLP00000007420
Platyfish	Xma	ENSXMAG00000008184	ENSXMAP00000008207
		ENSXMAG00000018083	ENSXMAP00000018115
		ENSXMAG00000004067	ENSXMAP00000004071
		ENSXMAG00000001128	ENSXMAP00000001124
		ENSXMAG00000007072	ENSXMAP00000007093
		ENSXMAG00000002174	ENSXMAP00000002189
		ENSXMAG00000011363	ENSXMAP00000011398
Amazon moly	Pfo	ENSPFOG00000003088	ENSPFOP00000002996
		ENSPFOG00000014167	ENSPFOP00000014207
		ENSPFOG00000016568	ENSPFOP00000016630
		ENSPFOG00000011157	ENSPFOP00000011161
		ENSPFOG00000019645	ENSPFOP00000023260
		ENSPFOG00000003023	ENSPFOP00000028259
		ENSPFOG00000017679	ENSPFOP00000022090
		ENSPFOG00000024727	ENSPFOP00000022300
Cod	Gmo	ENSGMOG00000010719	ENSGMOP00000011475
		ENSGMOG00000015210	ENSGMOP00000016293
		ENSGMOG00000007316	ENSGMOP00000007818
		ENSGMOG00000018398	ENSGMOP00000019805
		ENSGMOG00000005922	ENSGMOP00000006294
		ENSGMOG00000010334	ENSGMOP00000011064
Zebra fish	Dre	ENSDARG00000090598	ENSDARP00000124864
		ENSDARG00000052705	ENSDARP00000069093
		ENSDARG00000023026	ENSDARP00000035112

		ENSDARG00000051861	ENSDARP00000106964
		ENSDARG00000079438	ENSDARP00000101802
		ENSDARG00000045331	ENSDARP00000128634
Cave fish	Ame	ENSAMXG00000013604	ENSAMXP00000013984
		ENSAMXG00000007234	ENSAMXP00000007435
		ENSAMXG00000006089	ENSAMXP00000006241
		ENSAMXG00000017863	ENSAMXP00000018405
		ENSAMXG00000004555	ENSAMXP00000004662
		ENSAMXG00000010785	ENSAMXP00000011072
Spotted gar	Loc	ENSLOCG00000009825	ENSLOCP00000012011
		ENSLOCG00000015230	ENSLOCP00000018744
		ENSLOCG00000006681	ENSLOCP00000008066
		ENSLOCG00000005927	ENSLOCP00000007156
Elephant shark	Cmi	SINCAMG00000010966	SINCAMP00000016853
		SINCAMG00000015313	SINCAMP00000023540
		SINCAMG00000000122	SINCAMP00000000177
		SINCAMG00000000314	SINCAMP00000000514

Annex II

Tables containing the accession numbers of the desmosomal cadherin members found in the *in silico* searches in the ESTs databases. Tables describe the species and the local of expression of the respective EST. Dsg ESTs also describe the hypothetic duplicate form based on phylogenetic analysis.

Table Annex II.1: ESTs accession numbers of *dsc* found in the *in silico* search.

Species	*EST Accession number*	*Local of expression*
Nothobranchius furzeri	JZ298797.1	Whole body
	JZ302981.1	
	JZ244505.1	
	JZ268991.1	
	JZ322516.1	
	JZ245803.1	
	JZ240788.1	
	JZ265555.1	
	JZ297770.1	
	JZ229120.1	
	JZ288240.1	
Gadus morhua	GW844560.1	Beard
Gasterosteus aculeatus	DT949369.1	Gills
	DT961130.1	
	DT981497.1	Whole larva
	DT977460.1	
	DT977459.1	
	DN657386.1	Skin
	DN658874.1	
	DN657906.1	
	CD509394.1	Head and internal organs combined
Lates calcarifer	GT221672.1	Spleen
	GT219557.1	
Dicentrarchus labrax	FM023428.1	Intestine
Takifugu rubripes	CA588628.1	Skin
	BU805980.1	Whole fin

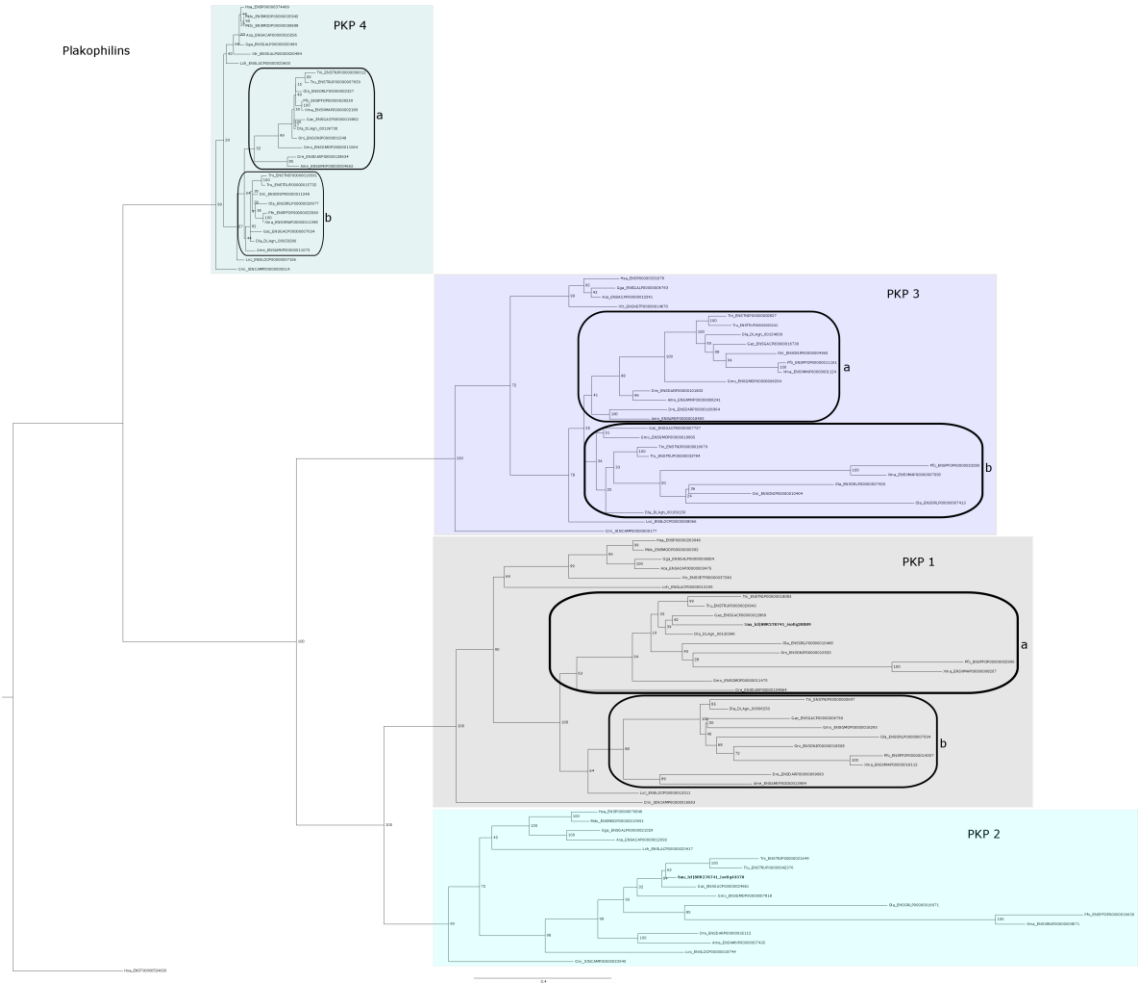
	BU806351.1	
	BU808527.1	
	BU806147.1	
	CA332821.1	Ovary
Oncorhynchus mykiss	BX303194.3	Adipose tissue, blood, brain, gonads in differentiation, gills, interrenal, intestine, kidney, liver, muscle, ovary, pituitary, testis
	BX315141.2	
	CA359520.1	Pooled
	CA358823.1	
	CA345207.1	
	CX718889.1	Pituitary
	CU066760.1	Multi-tissues
Danio rerio	BE201136.1	Fin
	EV756843.1	
	BQ480926.1	Testis (pooled)
	BM316366.1	
	EB980162.1	Bone
	CB358405.1	Embryo
	DN897810.1	Eye
	AI657752.1	26 somite embryos, adult livers, shield stage embryos
	AI588785.1	
	EB882489.1	Skin
	EB882139.1	
	AW594758.1	1 day fin regenerates
	EB904223.1	Gills
	EB902742.1	
	EB904244.1	
	CN512917.1	Whole body
	EE301622.1	
	CF996609.1	
	BI889630.1	Whole embryo
	GW712262.1	
Oreochromis niloticus	GR679856.1	Ovary
	GR623146.1	Whole embryo
Pimephales promelas	DT172230.1	Whole body
	DT304289.1	Testis

Ictalurus furcatus	FD216774.1	Mixed tissues from head, kidney, liver, spleen, gill, skin and intestine
	FD244642.1	
	FD078611.1	Mixed tissues from stomach, muscle, olfactory tissue and trunk kidney
	FD100899.1	
	FD158331.1	Library from stomach, muscle, olfactory tissue and trunk kidney
	FD152869.1	Trunk kidney
Ictalurus punctatus	GH685402.1	Mixed tissues from head, kidney, liver, spleen, gill, skin and intestine
	FD014933.1	Mixed tissues from stomach, muscle, olfactory tissue and trunk kidney
	FD309802.1	Whole fish
Astyanax mexicanus	FO317910.1	Whole embryo and larvae
	FO253889.1	
	FO256492.1	
	FO283587.1	
	FO264078.1	
Poecilia reticulata	ES381797.1	Embryo
	ES372657.1	Brain
Carassius auratus	AM929097.1	Olfactory epithelium
Oryzias latipes	DC267595.1	Regenerated adult caudal fin, 10 days postamputation
	AM140524.1	Gastrula
	AM139048.1	
	AM140524.1	
Paralichthys olivaceus	CX284884.1	Stomach
Fundulus heteroclitus	GT099825.1	Embryo
Misgurnus anguillicaudatus	BJ827040.1	Olfactory epithelium
Cyprinus carpio	EC392278.1	Skin
	EC394563.1	
Salmo salar	DY734559.1	Brain, kidney, spleen
	EG809013.1	Thymus
Sebastes caurinus	GE812493.1	Brain, kidney, spleen
Haplochromis chilotes	BJ687165.1	Jaw
	BJ686629.1	

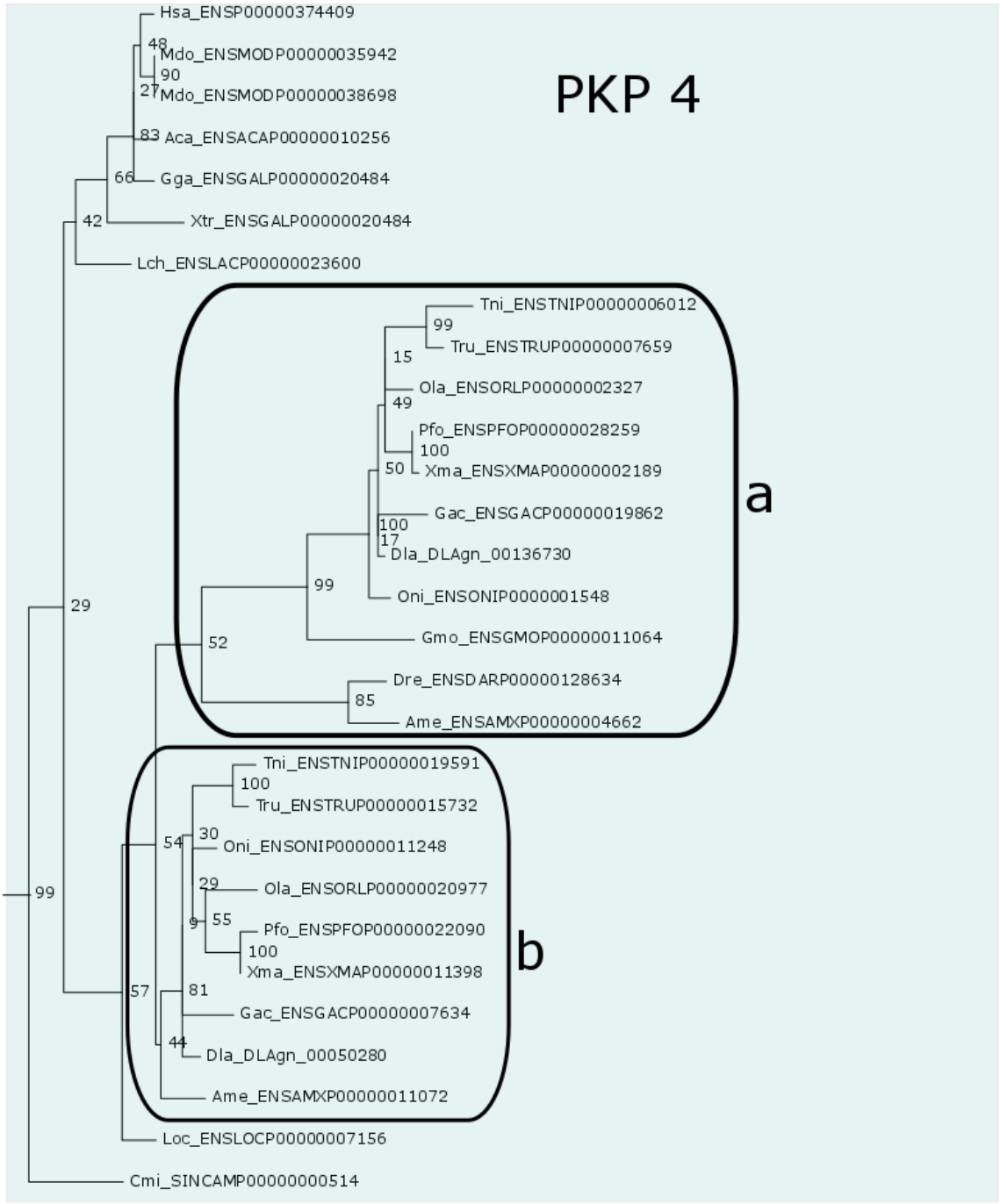
Table Annex II.2: ESTs accession numbers of *dsg* found in the *in silico* search.

Form	*Species*	*EST accession number*	*Local of expression*
a	*Gadus morhua*	GW853954.1	Skin
	Dicentrarchus labrax	FM020296.1	Gills
	Astyanax mexicanus	FO269165.1	Whole embryo and larvae
	Danio rerio	AL919813.1	Whole embryo or fish
		EB849701.1	Day 1 embryo
		DN897895.1	Eye
		EB935465.1	Testis
		EB931403.1	
b	*Pimephales promelas*	DT183833.1	Whole body
	Danio rerio	EB937168.1	Testis
		EB935830.1	
		EB937081.1	
	Oryzias latipes	DC266248.1	Regenerated adult caudal fin, 10 days postamputation
	Monopterus albus	GW584512.1	Intestinal tract
c	*Gasterosteus aculeatus*	DN676604.1	Gills
		DT954212.1	
		DN666077.1	
		DN656654.1	Skin
	Gadus morhua	GW844497.1	Beard
	Ictalurus furcatus	FD250479.1	Mixed tissues from head kidney, liver, spleen, gill, skin and intestine
	Oryzias latipes	DC267541.1	Regenerated adult caudal fin, 10 days postamputation
	Danio rerio	CB363329.1	Embryo
	Pimephales promelas	DT138703.1	Whole body

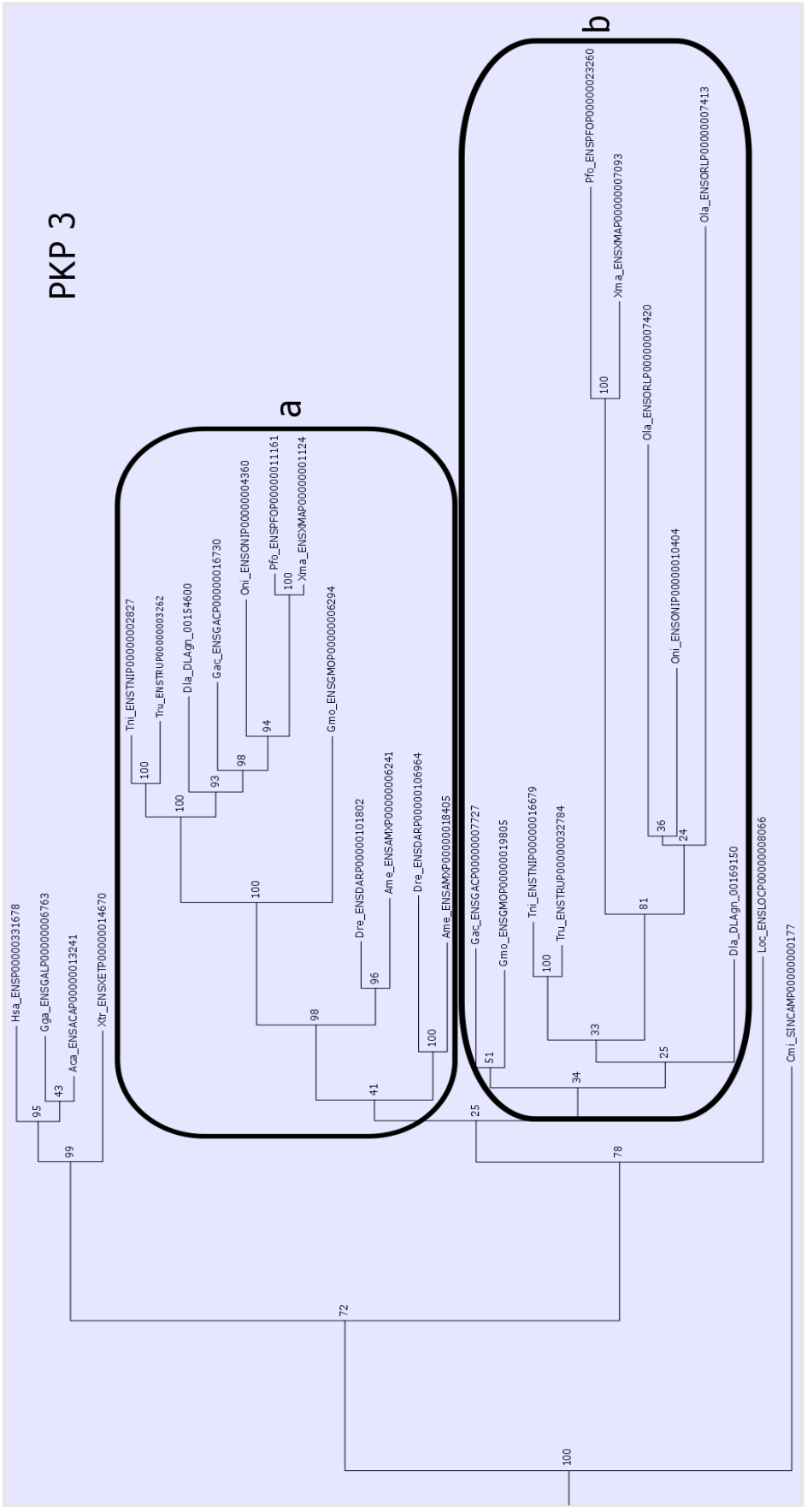
Annex III



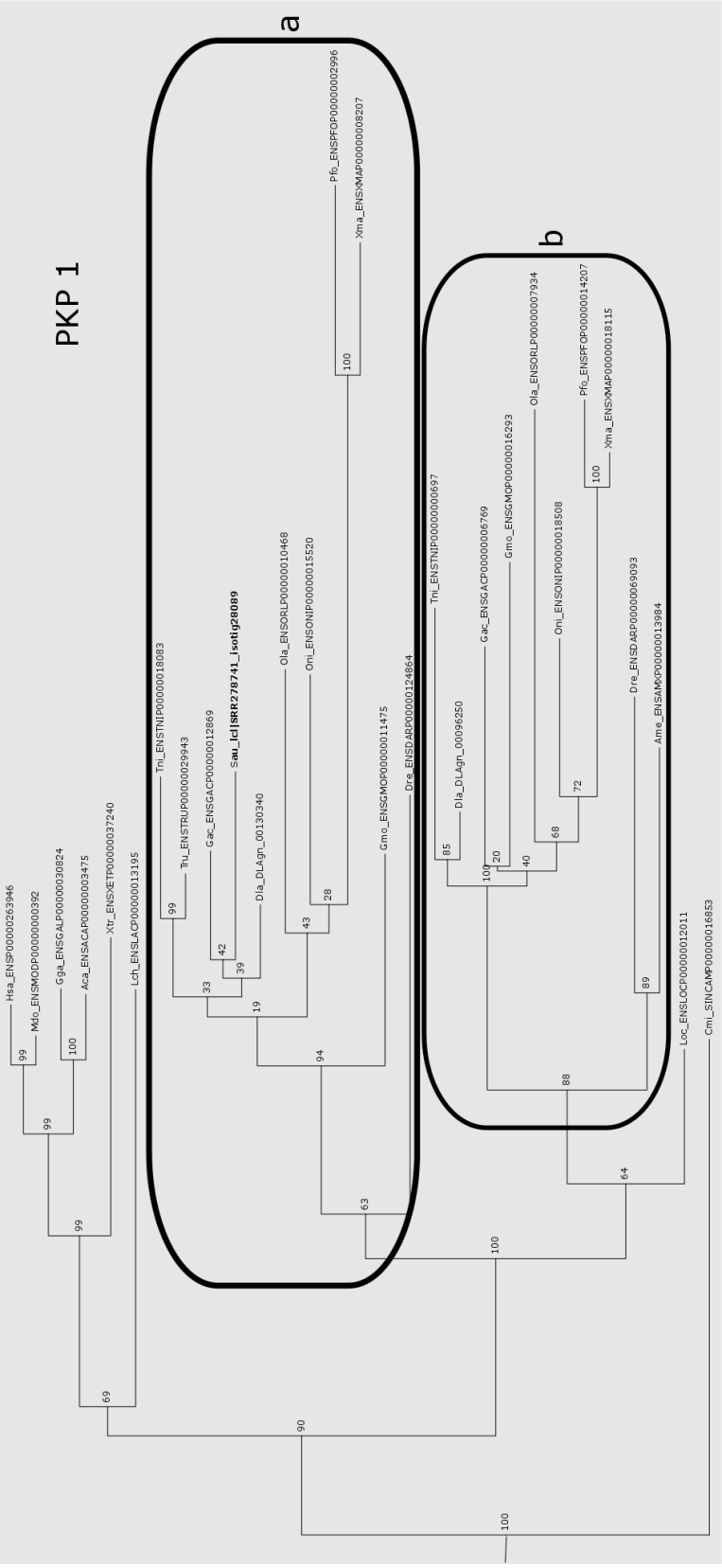
Annex Figure III.1: Complete version of the Figure 3.5. Phylogenetic tree of *pkp* members of fish and other vertebrates, obtained with the model JTT+I+G+F using the method maximum likelihood with a bootstrap analysis of 100 replicates. The tree was rooted with the human δ-catenin 1. Due the large dimensions of this phylogenetic tree, zoom in of each cluster (PKP4, PKP3, PKP1 and PKP2) is provided respectively in Annex Figures III.2, III.3, III.4 and III.5. Accession numbers available in Annex Table I.5 and Table 3.1.



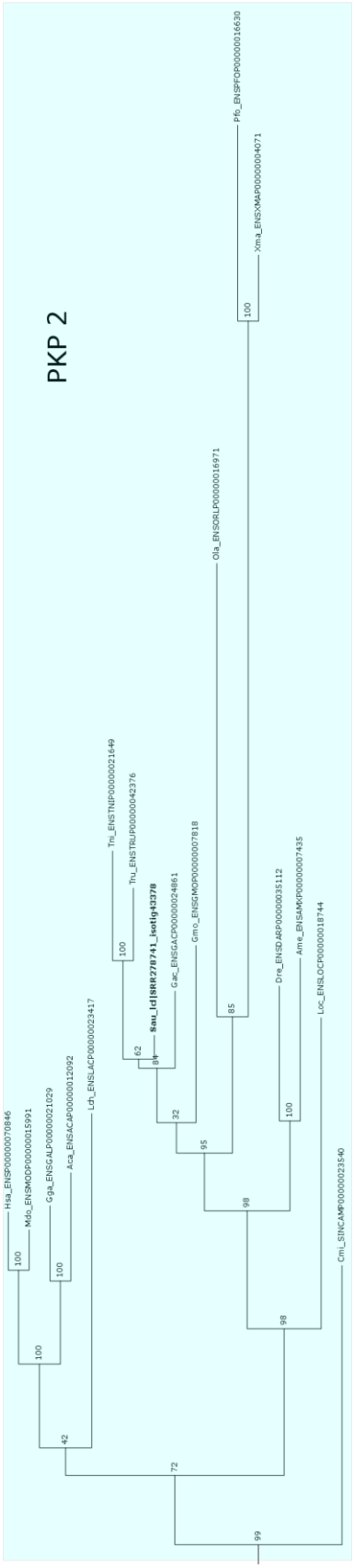
Annex Figure III.2**:** Zoom in of the *pkp4* cluster from Annex Figure III.1. *a* and *b* represent the different *pkp4* duplicates identified by the clustering organization.



Annex Figure III.3: Zoom in of the *pkp3* cluster from Annex Figure III.1. *a* and *b* represent the different PKP 3 duplicates identified by the clustering organization.



Annex Figure III.4: Zoom in of the *pkp1* cluster from Annex Figure III.1. a and b represent the different PKP 1 duplicates identified by the clustering organization.



Annex Figure III.5: Zoom in of the *pkp2* cluster from Annex Figure III.1.