Norwegian University
of Life Sciences

Master's Thesis 2017 60 ECTS
The Department of Animal and Aquacultural Sciences (IHA)

Characterization and expression analysis of the key genes for early development of swim bladder in Atlantic cod

Yihang Wang
Aquaculture

ACKNOWLEDGEMENT

The work here presented was performed at Nofima ÅS, Norway, during 2016-2017.

I would like to express my sincere gratitude to my supervisor, Øivind Andersen, for leading me into such an interesting and significant project. The experience of working with him was motive and fun. Thank you for all the constructive suggestion along the entire procedure of my work and constant encouragement during my study here.

I am also thankful to Katrine Hånes Kirste, for her endless patience and kindness to help on my work in lab. Besides, I sincerely thank Ifrat Jahan Tamanna, Gerrit Timmerhaus, and Hanne Johnsen for all the answers of my questions.

At last, I would like to thank Norwegian University of Life Sciences for offering me the opportunity to study and Nofima for supporting me to accomplish my experiments and thesis.

Yihang Wang
Ås, August 2017

ABSTRACT

Some genes have been proved critical for swim bladder inflation during early stages. Hence, our researches were focused on the investigation into the genetic features of the key genes to early development of Atlantic cod swim bladder. The *elovl1*, *pbx1*, *psap*, and *sftpb* genes were selected, and the expression modes during embryonic and larval stages were studied by qPCR quantification procedure. Genetic structures, multiple alignment, phylogeny of these genes were also investigated. Researches of Atlantic cod *sftpb* was further examined by studying relative expression levels in different organs and multiple genomic conserved synteny.

Cod *elovl1* and *psap* genes showed similar expression patterns during early development, with significant decline before early somitogenesis and stable expression thereafter. Cod *pbx1a* declined dramatically at late gastrula stage, while it was stable at relatively high level during larval period, while *pbx1b* seems to be stable throughout the early development. Cod *sftpb* dropped significantly before onset gastrula, thereafter, *sftpb* expression was stable before hatching, and abundant during larval stage. Cod *sftpb* was much enriched in male reproductive organ than in ovary, *sftpb* was also enriched in heart, spleen, head kidney, and gas gland. However, *sftpb* expression was weak in ovary, pancreas, brain, and liver, and gill.

The phylogenetic tree showed Atlantic cod Pbx1 proteins seem to have closer evolutionary relation with coelacanth Pbx1 than ballan wrasse Pbx1, tilapia Pbx1b, or platyfish Pbx1b. Cod Psap seems to intimate with most of Actinopterygii Psap except for spotted gar Psap, and elephantfish Psap. Fish Sftpb and Psap proteins were closely clustered in the phylogenetic trees. Whale shark Sftpb and Psap may exhibit the differentiation process between these two genes in the long- term evolution history. Atlantic cod Sftpb seems to be intimate with the Sftpb existed in stickleback group XIII, tetraodon chromosome 12, Amazon molly scaffold KI519905.1, platyfish scaffold JH556662.1, and medaka chromosome 9, and their genomic environments were highly conserved as well. Highly conserved orthologous pairwise between Atlantic cod scaffold 2788 and stickleback group XIII was observed. Extensive pairwise between the *sftpb* paralogons in stickleback group XIII and XIV may indicate a large-scale duplication event in the evolutionary history.

Key words: Atlantic cod, swim bladder, phylogeny, synteny

CONTENT

LIST OF ABBREVIATIONS

aa: Amino acids

BLAST: Basic local alignment search tool

BR: Brain

C9: Cleavage at 9 ℃

Cers: Ceramide

CoAs: Coenzyme As

Chro/Chr: Chromosome

dpf: Days post fertilization

DPPC: Dipalmitoylphosphatidylcholine

elovl: Elongation of very long chain fatty acids

ER: Endoplasmic reticulum

ES9: Early Somitogenesis at 9 ℃

EtOH: Ethyl alcohol

GI: Gill

Ha9: Hatching at 9 ℃

HE: Heart

hES: human embryonic stem cells

HK: Head kidney

HOX: Homeobox

hpf: Hours post fertilization

LB9: Late Blastula at 9 ℃

LG9: Late Gastrula at 9 ℃

LI: Liver

LS9: Late Somitogenesis at 9 ℃

MAPK: Mitogen-activated protein kinase

MB9: Mid Blastula at 9 ℃

min: Minutes

OG9: Onset Gastrula at 9 ℃

OV: Ovary

PA: Pancreas

pbxs: Pre- B cell leukemia transcription factors

Psap: Prosaposin

PUFA: Polyunsaturated fatty acid

qPCR: Quantitative real-time polymerase chain reaction

Sca: Scaffold

sec: Seconds

sftp: surfactant- associated protein

SPs: Saposins

SP: Spleen

St4: Stage 4

St8: Stage 8

St11: Stage 11

St12: Stage 12

TALE: Three amino acid loop extension

TE: Testis

UE: Unfertilised Egg

UTR: Untranslated region

VLCFA: Very long chain fatty acids

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Genetic roles of surfactant- related genes

Saposins make up many saposin- like proteins that are highly related to surfactant, such as prosaposin and surfactant protein (SP) B. The precursor of SP-B has been proved homologous with prosaposin (Patthy 1991), which is consistent with the major role of SP- B proteins in the biophysical activity of the surfactant complex.

1.1.1 Functions of prosaposin and surfactant proteins

Human prosaposin (PSAP), encoded by multifunctional gene *PSAP*, is the common precursor of four cleavage lysosomal activator proteins known as saposins (SPs), including SP- A, B, C, and D (Rorman *et al.* 1992). PSAP is glycosylated in the endoplasmic reticulum and the Golgi apparatus. With the help of cathepsin D, proteolytic cleavage of PSAP takes place in the lysosomes, following the transportation to the lysosomes (Zhu *et al.* 1994; Igdoura *et al.* 1996; Leonova *et al.* 1996; Hiraiwa *et al.* 1997). During the proteolytic process, the excision of SP- A happens first after the N- terminus of PSAP is removed (Leonova *et al.* 1996). SP- B, C, and D were then cleaved in tandem (O'Brien *et al.* 1988). All SPs are small nonenzymatic proteins involved in the breakdown of sphingolipids as the cofactors for lysosome (Kolter *et al.* 2005). In human and mouse, PSAP comprises the basic requirement for epidermal barrier formation (Doering *et al.* 1999). The deficiency of PSAP in the formation causes a series of abnormality, including excessive glucosylceramide, insufficient ceramide in epithelium, and deformed stratum corneum (Chang-Yi *et al.* 1997; Doering *et al.* 1999). In zebrafish, and the function of epidermal permeability barrier has been linked largely to the swim bladder and kidney development (Bhandari *et al.* 2016). It is highly possible that zebrafish prosaposin also contributes to the formation of epidermal permeability barrier, and thus affects the development of swim bladder.

In human, PSAP is critical for the homeostasis of nervous system by participating in the storage of lysosomal and sphingolipid. Mutation of *PSAP* has been proved associated with several severe diseases involved in lysosomal or sphingolipid storage, including Gaucher disease (mutations in the Sap- C domain), metachromatic leukodystrophy (mutations in the SP- B domain), and Krabbe disease (mutations in the SP- A domain) (Horowitz *et al.* 1994; Regis *et al.* 1999; Spiegel *et al.* 2005). PSAP is also an cofactor for lysosomal enzymes involved in the degradation of

sphingolipids (Siri *et al.* 2014). Sphingolipids have a high melting temperature and thus increase membrane rigidity. Hence, Sphingolipids are often frequent targets to treat many lungs disorders (Uhlig *et al.* 2008). Sphingolipid synthesis has also been found an key factor that influence the early development of swim bladder in zebrafish (Bhandari *et al.* 2016). In zebrafish, prosaposin was identified as Psap, encoded by *psap*. The deficiency of zebrafish *psap* may disorder the swim bladder by failing to store sphingolipids.

In mouse, PSAP also plays an important role in maintaining male reproductive system, including the differentiation of reproductive organs, genesis of sperm, and fertilization (Guo *et al.* 2007). For example, PSAP has ability to bind Rhox5 (Guo *et al.* 2007), which is effective in decreasing male germ cell apoptosis and in increasing sperm vitality (Shanker *et al.* 2008). The deformities of the prostate gland and other reproductive organs in mice caused by the inactivation of PSAP also gives the evidence that prove the importance of Psap in maintaining reproductive system (Morales *et al.* 2000).

Genes involved in MAPK signaling pathways has been identified highly enriched in zebrafish swim bladder and mammalian lung (Zheng *et al.* 2011). Psap has ability to trigger MAPK and PI3K/ Akt signaling pathways and gives play to their functions (Morales *et al.* 2000). For example, in prostate cancer cells, the enrichment of Psap leads the cancer to carcinogenesis and progression at early stage by activating the MAPK and PI3K/ Akt signaling pathways (Koochekpour *et al.* 2005).

Pulmonary surfactant plays critical roles at the air- tissue interface in lungs during inspiration and expiration, including lower the surface tension in air breathing tissues, reduce the energy of inspiration and the risk of atelectasis or alveolar edema (Floros *et al.* 1998). Surfactant homeostasis is indispensable to lung health and normal function. Normal surfactants usually have both hydrophilic heads clinging to the water and hydrophobic tails facing towards the air, maintaining the air sacs filling with air, preventing the tissues from sticking together. Qualitative or quantitative changes in surfactant composition are associated with respiratory disorders or diseases (Floros *et al.* 2001; Whitsett *et al.* 2010; Silveyra *et al.* 2012).

Surfactant consists of disaturated and unsaturated phospholipids, neutral lipids, and surfactant proteins. Surfactant proteins make up around 5% of the surfactant. Within the four types of the hydrophilic SFTP- A, B, C, and D, SFTPA and SFTPD contribute to immune functions within alveoli in mammalians (Crouch 1998;

McCormack 1998), while SFTPB and SFTPC regulate the surface activity of lung surfactant lipids, promote the spread of surfactant over the surface, and make breathing process easily (Hawgood *et al.* 1998; Johansson 1998).

SFTPB and SFTPC are small proteins with highly hydrophobicity produced from high molecular weight precursors (Biscotti *et al.* 2016). Both proteins are required for proper biophysical function of the lung in packaging phospholipid, organizing surfactant, and lowering the surface tension of lungs after exhalation.

Similar to SFTPA, SFTPB was also found specific to DPPC that helps to stabilize the phospholipid monolayer (Kishore *et al.* 2006). Among the surfactant proteins, SFTPB is of vital importance for life (Clark *et al.* 1995). The deficiency or mutations of SFTPB causes pulmonary surfactant metabolism dysfunction type 1, associated with fatal respiratory disorder in the neonatal period (Kebaabetswe *et al.* 2015; Wang *et al.* 2016).

1.1.2 Distribution of prosaposin and surfactant B protein

Despite of the mature saposins in mice are dominant in liver, lung, kidney, and spleen (Morales *et al.* 2000; Kolter *et al.* 2005), PSAP exhibits a distinct distribution pattern. PSAP is mainly distributed in skeletal muscle, heart, and brain, and it is also enriched in various biological fluids as the component, such as blood, milk, semen, cerebrospinal fluid, bile, and pancreatic juice (Hineno *et al.* 1991). PSAP exists also in sensory organs, including the organ of Corti, ears, and eyes (Van Den Berghe *et al.* 2004; Terashita *et al.* 2007). The concentration of PSAP presents differently in the tissues that contain PSAP (Sun *et al.* 1994). Prosaposin exhibits high concentration in the nervous and reproductive systems, esophagus, upper trachea, and lungs, while the concertration of prosaposin in the organs of heart, kidney, liver, adrenals and lymphoid exhibits lower concentration (Sprecher-Levy *et al.* 1993).

Prosaposin was also identified in the zebrafish as Psap, encoded by *psap*, and it was enriched in the swim bladder (Zheng *et al.* 2011). The author also suggested that *psap* were highly expressed in the zebrafish swim bladder than in human lung.

In mammalian, SFTPB and SFTPC have been found regulate the surface activity of lung surfactant lipids (Hawgood *et al.* 1998; Johansson 1998). Except for mammalian lung, surfactant B has also been identifid in fish swim bladder and gas gland as Sftpb, encoded by *sftpb* (Perry *et al.* 2001; Zheng *et al.* 2011). Gas gland is a compact structure and gas gland cells are connected to the swim bladder lumen via

small canals, lamellar bodies were also found in flattened cells forming the swim bladder epithelium. Gas gland cells of physostome have been proved to produce surfactant in vivo and culture (Perry *et al.* 2001). Just like prosaposin, surfactant B was also found in male production systems, including the kidney and testis, involved in the maintainance of the reproduction systems (Holstein *et al.* 2003; Liu *et al.* 2015).

1.1.3 Structures of prosaposin and surfactant proteins

Human *PSAP*, which is highly conserved, localizes in the chromosome 10q22.1 (Bar-Am *et al.* 1996). Human PSAP comprises 524, 526 or 527 amino acids, resting with the splicing region of mRNA (Hiraiwa *et al.* 1997). Compared to human PSAP, mouse and chicken PSAP share an identity of 77% and 59%, respectively (Zhao *et al.* 1997; Azuma *et al.* 1998). 15 exons encodes the four central and two peripheral domains of PSAP (Patthy 1991). Among the 15 exons, the first two and the last exons encode the N- and C- terminus, respectively, the exons 3- 5 encode SP- A, exons 6- 9 encode SP- B, exons 10 and 11 encode SP- C, exons 12, 13, and 14 encode SP- D (Zhao *et al.* 1997). The central domains, termed as B- type domains (SP- A, B, C, and D), are linked in tandem by three connecting regions (Patthy 1991; Kishimoto *et al.* 1992; Tsuda *et al.* 1992). According to Pfam online searching tool, each Saposin- like type B domain consists of region 1 (Accession ID: PF05184) and region 2 (Accession ID: PF03489). The connecting regions of the central domains can be cleaved to liberate four saposins into the lysosomes (O'Brien *et al.* 1991). The peripheral domains, termed as A- type domains (N- terminus and C- terminus), link the central domains at both ends respectively (Patthy 1991). In human, the B- type and A- type domains comprise 80- 84 and 34 amino acid residues, respectively (Patthy 1991; Kishimoto *et al.* 1992).

Based on the alternatively spliced exon 8, three PSAP forms are generated with three, two, and zero extra amino acid residues respectively (Madar-Sharipo *et al.* 1999). In addition, two other forms of PSAP, following post-translational glycosylated, exist as the fully glycosylated protein (70 kDa), and the partially glycosylated protein (65 kDa) (Tsuda *et al.* 1992). PSAP is initially a 53 kDa protein which is further modified to a 65 kDa protein and can be further transferred into a 70 kDa protein eventually. The 65 kDa protein binds to Golgi membranes is combined with lysosomes, where the saposins hydrolyze the GM1 ganglioside,

glucosylceramide, galactosylceramide, sulfatide, sphingomyelin and other sphingolipids (Hiraiwa *et al.* 1992; Zhao *et al.* 2000). However, the 70 kDa PSAP, exists in many biological fluids, is an uncleaved protein which is not affinitive with lysosomes (Kishimoto *et al.* 1992). The 70 kDa PSAP is involved in the activation of MAPK and PI3K/Akt signalling pathways (Campana *et al.* 1998), in the biogenesis of membrane (Soeda *et al.* 1993), and also in the maintenance of the male reproductive system (Morales *et al.* 2000).

Surfactant B is highly hydrophobic protein belonging to the saposin super family with several variation in tertiary structures (Ahn *et al.* 2003; Ahn *et al.* 2006). Human SFTPB is encoded by *SFTPB* gene and synthesized in the pneumocytes of 381 hydrophobic amino acids and 42 kDa with 79 residue (eight conserved positively-charged residues and one conserved negatively- charged residue) (Khatami *et al.* 2016). Seven cysteines are involved in the SFTPB structure, six of which form the intra- molecular disulphide bridges C8–C77, C11–C71 and C35–C46 and stabilize the tertiary structure (Johansson *et al.* 1991; Andersson *et al.* 1995; Serrano *et al.* 2005), while the seventh cysteine is involved in the covalent dimerisation of the protein (Serrano *et al.* 2005). The seventh cysteine makes one of the major difference from other saposins. Reciprocal ion pairs formed by E51 and R52n is also functional as the disulphide bridges that involve in the dimerisation of SFTPB, which is supported by the strict conservation of those residues (Zaltash *et al.* 2000). The second structure of SFTPB contains four amphipathic α- helics (Andersson *et al.* 1995), while the third dimensional structure of Sftpb contains a globular closed saposin fold of five packed helices (Zaltash *et al.* 2000), which is supported by the structural analysis of the largest fragment of Sftpb termed Mini- B (Fig. 1) (Sarker *et al.* 2007). According to Zaltash suggestion, Sftpb has an opening compact fold may be important in lipid interactions (Khatami *et al.* 2016). The feasible structures within or near bilayer may be in response to surface tension variations (Zaltash *et al.* 2000; Fullagar *et al.* 2003; Wang *et al.* 2003; Walther *et al.* 2007).

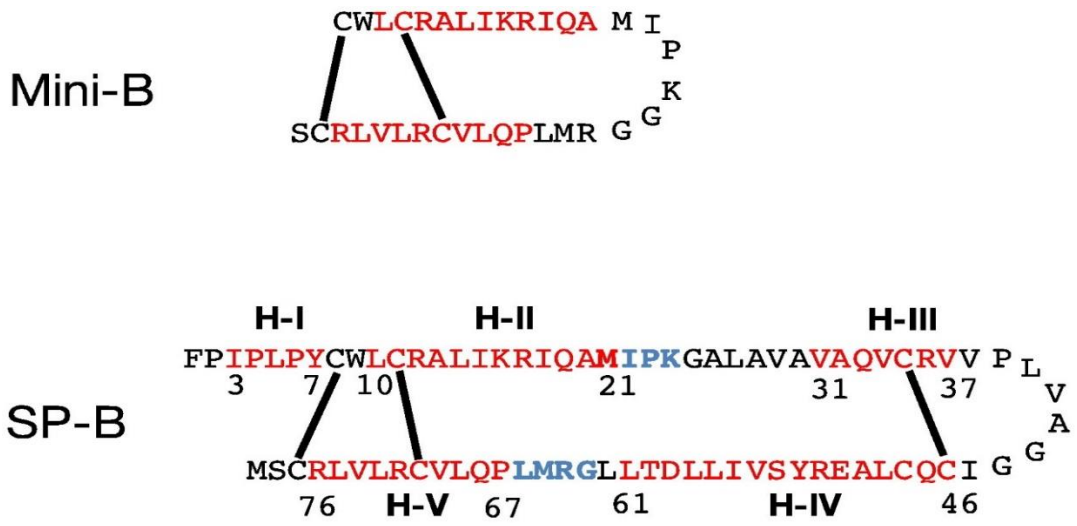


Fig. 1. Amino acid sequence of SP- B and Mini- B (Revived from Khatami et al. 2016)

Another structural difference of SFTPB from other saposin super family proteins is the N- terminal 7- residue. The residue helps SFTPB strongly associate to surfactant phospholipid membranes (Walther *et al.* 2010). Various investigations into N-terminal 7- residue have shown its activities in lipid bilayers interactions, including membrane binding, lysis, fusion, lipid adsorption, stabilization of monomolecular surface films (Hawgood *et al.* 1998).

Despite the specific characters of SFTPB have been known, a further understanding of functional and structural mechanism of SFTPB is still lacking. Besides, the studies of SFTPB structure are still focused on human, it is still unknown whether the structural differences from other species exist. The characteristics of surfactant B structure in other species, such as fishes, are expected for research.

1.1.4 Homologous evidences of surfactant proteins in tetrapods and teleosts.

Surfactants have also been identified and characterized in fishes either with swim bladders that are used for buoyancy, such as Goldfish (*Carassius auratus*) (Daniels *et al.* 1994), European eel *(Anguilla anguilla*) (Prem *et al.* 2000) and common carp (Cyprinus carpio) (Rubio *et al.* 1996), or for air breathing, such as gar (*Atractosteus spatula*) (Smits *et al.* 1994). Interestingly, surfactants have also been localized in some fishes without swim bladder, such as physoclist perch (*Perca fluviatilis*) (Prem *et al.* 2000), and elefant shark identified in Ensembl.fugu (Gene ID: SINCAMG00000002317). In air- breathing fishes, such as rope fish, bichirs, and lungfishes, surfactants are also been identified (Smits *et al.* 1994; Orgeig *et al.* 1995;

Sullivan *et al.* 1998).

Surfactant A, C, D were mostly identified in tetrapods, whereas surfactant B is conserved across many vertebrates (Biscotti *et al.* 2016). Lungs and swim bladders that can be used either for air breathing or for buoyancy is the evidence of convergent evolution. In mammals, surfactant B is specific enriched in lung (Biscotti *et al.* 2016). Latest studies have identified the surfactant system in the swim bladders of either air-breathing or non- air- breathing teleost fish, including pirarucu (*Arapaima gigas*), tarpon (*Megalops cyprinoides*), goldfish (*Carassius auratus*) and New Zealand snapper (*Pagrus auratus*) (Daniels *et al.* 2004). Besides, surfactant B is both identified in tarpon and pirarucu lavage, and in swim bladder tissue of the tarpon. Gas gland cells of physostome have also been shown to produce surfactant in vivo and in culture (Prem *et al.* 2000). The presence of the surfactant system in Actinopterygii and the Sarcopterygii provides strong evidence that the surfactant system is homologous in teleost fish and in tetrapods.

1.2 Genetic roles of Elovl1

In vertebrates, the long chain fatty acids (LCFA) or very long chain fatty acids (VLCFA) play key roles in phototransduction, skin permeability, and fertility (Agbaga *et al.* 2010; Monroig *et al.* 2011). Up to C16 fatty acids that are either derived directly from the diet or synthesized in the cytosol by fatty acid synthase are further elongated into LCFA C18 and C20, or VLCFA with ≥ C22 (Jakobsson *et al.* 2006). VLCFA is mainly formed in the endoplasmic reticulum (ER) by enzymes (Morais *et al.* 2009). The fatty acid elongases, elongation of very long chain fatty acids (elovl), is the condensing enzyme that specify the substrate in the cytoplasmic side of ER membranes through catalyzing the driven formation of a 3- ketoacyl- CoA from malonyl- CoA on the first step of the enzymatic reactions in the fatty acid elongation. Elovl also determines the rate of the condensation (Schackmann *et al.* 2015). Seven elovl proteins (Elovl1- 7) are found with differing fatty acid substrate specificities (Morais *et al.* 2009). Generally, elongation of very long chain fatty acids protein 1 (Elovl1) involve in the synthesis of saturated and monounsaturated VLCFA (Jakobsson *et al.* 2006).

In mammals, Elovl1 was found predominant in the elongation of C22:0 to C26:0 (Ofman *et al.* 2010). Like Psap, Elovl1 also comprises the basic requirement for epidermal barrier formation (Doering *et al.* 1999). Elovl1 shows high activity toward

saturated and monounsaturated C20 and C22 coenzyme As (CoAs), enzymes essential for C24 sphingolipid synthesis. Elovl1 activity is regulated with the ceramide synthase CerS2, which is also essential for C24 sphingolipid synthesis. Knockdown of *elovl1* will cause a reduction in the activity of the Src kinase family LYN, which functions in leukocytes membrane (Ohno *et al.* 2010). The result found on human ELOVL1 is consistent with the finding on mouse ELOVL1 and the functional orthologue of yeast elongase Elo3p (Oh *et al.* 1997; Tvrdik *et al.* 2000).

Studies on mice showed that *Elovl1* gene is ubiquitously expressed in the stomach, lung, kidney, skin, and intestine (Tvrdik *et al.* 2000; Asadi *et al.* 2002). It also suggested that *Elovl1* is determinant of epidermal sphingolipid backbone ceramide (Cers) chain length and is essential for epidermal barrier formation (Tvrdik *et al.* 2000), but the precise functions in vertebrate development or in different organs have not been well studied because of the early death of *Elovl1* knockout mice (Sassa *et al.* 2013).

Fish species have distinct abilities in producing polyunsaturated fatty acid (PUFA) or LC-PUFA probably due to their evolutionarily natural diets (Agaba *et al.* 2005). Most marine fish, including Atlantic cod, derive LC-PUFA from their natural diets as they are incapable or inefficient of producing LC-PUFA themselves (Agaba *et al.* 2005; Tocher *et al.* 2006). It might be caused by the limitation of elongation from C18 to C20 (Ghioni *et al.* 1999). Although Elovl1 has been widely identified in bony fishes, cartilaginous fishes, and coelacanths, further molecular characterization and functional analysis of fish Elovl1 were only studied on several species, including Atlantic cod (*Gadus morhua*) (Xue *et al.* 2014), sea bream (*Sparus aurata*) (Benedito-Palos *et al.* 2014), and zebrafish (*Danio rerio*) (Bhandari *et al.* 2016). Sequencing analysis indicates that elovl 1 is encoded by two genes in some bony fishes, *elovl1a* and *elovl1b*.

In zebrafish, *elovl1* genes are highly expressed in swim bladder from two days post fertilization (dpf), and *elovl1b* is expressed in the pronephric kidney as well. Gene *elovl1* is linked to yolk consumption, lipid metabolism, and swim bladder development (Bhandari *et al.* 2016). Deficiency of *elovl1* results in a deflated swim bladder, and deficiency of e*lovl1b* affects kidney development as well. Interestingly, *elovl1* knockout does not affect neuronal myelination in zebrafish, which is distinct with the result in mice (Yeh 1988; Bhandari *et al.* 2016).

1.2.1 Structures of Elovl1 in fish

Fish Elovl1 encoded by *elovl1* gene was found highly conserved in vertebrates (Bhandari *et al.* 2016). In Atlantic cod, *elovl1a* and *elovl1b* encode 306 and 319 AA proteins, respectively. 55% identity is shared between this two genes. In addition, *elovl1b* has two cDNA variants that showed 100% identity over the 1639 bp aligned at the 5′ end. Both *elovl1a* and *elovl1b* consist of 8 exons and 7 introns (Xue *et al.* 2014).

1.2.2 Transcript expression of *elovl1* in Atlantic cod

Qualitative RT- PCR analysis in Atlantic cod showed that *elovl1b* transcript is ubiquitously expressed in all tissues tested, including skeletal muscle, skin, eye, brain, head kidney, posterior kidney, spleen, pyloric caecum, midgut, hindgut, stomach, liver, blood, heart, and gill (Xue *et al.* 2014). This result is consistent with the broad range of transcript expression in other teleost fishes (Benedito-Palos *et al.* 2014; Bhandari *et al.* 2016) or human (Ohno *et al.* 2010). The broad expression of cod *elovl1b* suggests that Elovl1b may be required as 'housekeeping elongase' to stabilize the specific fatty acids as suggested in mammals (Jakobsson *et al.* 2006; Guillou *et al.* 2010). Unlike *elovl1b*, *elovl1a* transcript distribute in a much narrow range of tissue (Xue *et al.* 2014; Bhandari *et al.* 2016). In addition, the paralogues of *elovl1* were only found in several evolutionarily diverged fish species, such as Atlantic cod, pufferfish, and zebrafish. The distinctive differences in cod *elovl1a* and *elovl1b* transcript expression suggest that the paralogues might be undergoing the regulatory divergence.

1.3 Genetic roles of Pbx1

Pre- B cell leukemia transcription factors (pbxs) belong to PBC group of the three amino acid loop extension (TALE) superclass of homeodomain proteins (Bürglin 1997). *Pbx* gene is highly conserved in Mammalian. *Pbx* isoforms have highly similar sequences, which are identical within or flanking their DNA- binding homeodomain (Monica *et al.* 1991; Wagner *et al.* 2001). Besides, mammalian *Pbx* is a ortholog of *ceh- 20* in *Caenorhabditis elegans* (Burglin *et al.* 1992), *Exd* in *Drosophila melanogaster* (Rauskolb *et al.* 1993), and *lazarus* in *Danio rerio* (Pöpperl *et al.* 2000; Vlachakis *et al.* 2000).

As an essential homeobox (HOX) cofactor, Pbx interacts efficiently both with the

TALE domain proteins in the first helix of homeodomain, and with Hox partners through a short C- terminal tail comprising 16 amino acid residues (Moens *et al.* 2006). Besides, short C- terminal tail is also essential for maximal monomeric binding of Pbx1 to DNA. Pbx also interacts with other homeodomain proteins, such as the Meis/Prep subfamily of TALE domain proteins (Chang *et al.* 1997; Berthelsen *et al.* 1998; Kilstrup‐Nielsen *et al.* 2003), or PDX1 and Engrailed (Swift *et al.* 1998). An increasing studies also suggest that Pbx acts a broader role as partner of non-homeodomain transcription factors (Laurent *et al.* 2003).

PBX1 gene was firstly identified in human (Kamps *et al.* 1990; Nourse *et al.* 1990). Based on the conserved sequence of *PBX1, PBX2, 3, 4* genes were then identified (Monica *et al.* 1991; Wagner *et al.* 2001). Shortly after, *pbx1a* and *pbx1b*, derived from C- terminal differential splicing of *pbx1*, have been identified in bony fishes, including zebrafish (Teoh *et al.* 2010) and common carp (Kolder *et al.* 2016). In addition, *pbx1a* has a longer sequence than *pbx1b* (Monica *et al.* 1991; Wagner *et al.* 2001).

The importance of *PBX1* was subsequently reported. For example, the disruption of *PBX1* caused by t(1;19) chromosomal translocation leads to pre- B acute lymphoblastic leukemia in human (Kamps *et al.* 1990; Nourse *et al.* 1990). To date, *PBX1* was reported to regulate *Nanog* directly in human embryonic stem cells (hES) (Chan *et al.* 2009). *Pbx1* gene was also identified in mice, and in *Pbx1* knockout mice, various abnormalities and deadly embryonic lethality were observed. It has been proved that *Pbx1* is critical in the formation and development of several tissues or organs, such as skeleton (Selleri *et al.* 2001), pancreas (Kim *et al.* 2002), kidney, adrenal (Schnabel *et al.* 2003), and caudal pharyngeal pouch- derived organ (Manley *et al.* 2004). *Pbx1* also maintains the hematopoiesis in fetal liver, serves to modulate early stages of B- cell development (Selleri *et al.* 2001), and differentiates the urogenital organs (Schnabel *et al.* 2003). These findings, together with the roles of *PBX1* in hematopoiesis and development of organs (Chan *et al.* 2009), indicates the function of *Pbx1* that maintain the pluripotentcy of hES cells.

PBX1 has shown great activity in the mouse lung mesenchyme for the expression of *Fgf10* (Park et al. 2008). The lack of PBX1 in the lung mesenchyme of mouse embryos led to compact terminal saccules and perinatal lethality with failure of postnatal alveolar expansion (Li *et al.* 2014).

Despite of the various characterizations of mammalian *Pbx1*, the function of *pbx1* in bony fishes has not been well characterized. The tetrapod lung and the teleost swim bladder are homologous organs derived from the foregut during embryonic period (Shannon *et al.* 2004). The important roles that Pbx1 plays in the development of swim bladder might be similar with the development of mouse lung organogenesis (Schnabel *et al.* 2001). In zebrafish, based on the available cDNA sequences of *pbx1a* and *pbx1*b, the functional characterization was speculated (Teoh *et al.* 2010). At embryonic stages, *pbx1* could be detected in the central nervous system and pharyngeal arches since 24 hpf, and in the swim bladder since 28 hpf. The knockdown of *pbx1* deactivates the inflation of swim bladder, with fetal consequence at 8 dpf. Besides, the *pbx1* knockdown abolishes the expression of *anxa5* completely, which is essential for swim bladder development, at 60 hpf, suggesting that *pbx1* may be also essential for later development of swim bladder (Teoh *et al.* 2010).

1.3.1 Structure of zebrafish *pbx1*

In zebrafsih, *pbx1a* and *pbx1b* with two different C- termini are two isoforms of *pbx1*. Both isoforms are consist of 7 exons and 5 exons, respectively, while human *PBX1* comprises 9 exons (Kamps *et al.* 1990; Teoh *et al.* 2010). It was found that *pbx1b* is shorter than *pbx1a* by 37 amino acids (aa) at exon 5 splicing (Fig. 2). Two conserved domains, PBC and HOX domain, are both included in the N- terminus of two isoforms. PBC domain is located at 38 to 231 aa, and the HOX is at 223 to 292 aa.

Fig. 2. Protein domains of zebrafish Pbx1a and Pbx1b (Revived from Teoh *et al.* 2010).

1.4 Early development of Atlantic cod

In some fish species like Atlantic cod (*Gadus morhua*), dramatic body changes in structure and function happen at the early life stages. Meanwhile at these stages,

Atlantic cod is extremely sensitive to abiotic factors, such as incubation densities, temperature, salinity, and oxygen (Puvanendran *et al.* 2015). High incidence of deformities and high mortality rates are easily seen during these life stages in farming (Taranger *et al.* 2010). Even so, it is confirmed that the early life stages of Atlantic cod is operable primarily to maintain the fish stocks at their observed levels (Hunter 1976). The research of early development of Atlantic cod is therefore necessary to guarantee the quality and quantity of stocks.

1.4.1 Stages of the embryonic development

Egg quality is the key factor for the populations of both wild and farmed fish, although the factors that affect egg quality of wild and farmed fish are quite different (Avery *et al.* 2009). Normally the batches of eggs that have high probability of low hatching rate are discarded before costly devotion to their culture (Avery *et al.* 2005).

The first cell cleavage of the fertilized cod egg can be observed at about 5 to 6 hours post fertilization (hpf) at 7 ℃. The midblastula transition is reached, followed by nine or ten consecutive cell divisions. During the blastula period, the cells increase and gather around the blastoderm rim, getting ready to move toward the surface of the yolk. At about 84 h, the egg reaches 50% epiboly. At about 94 h, the first somite can be observed. The following somites are formed anterior to posterior, until the 50th somite pairs are observed at about 220 hpf. At the following segmentation period, the hatching begins at approximately 250 hpf (Hall *et al.* 2004). The description of cod embryogenesis was further simplified into six stages, including fertilization, cleavage, blastulation, gastrulation, somitogenesis, and prehatching and hatching (Gorodilov *et al.* 2008).

1.4.2 Stages of the larval development

After hatching, the larval of Atlantic cod experiences the most rapid growth and changes (Wieser 1991). The morphologic and functional development of feeding and skeletal structures during larval stages has been well studied (Herbing *et al.* 1996a; Herbing *et al.* 1996b; Herbing 2001). Herbing *et al*. (1996a) has described nine external morphological structures as landmarks for a total of three larval periods, including 12 stages from hatching to complete exhaustion of the yolk sac.

At first period (0- 6 days old, stage 1- 4), newly hatched larvae form and develop mouth and gill quickly. The larval nutrition fully depends on yolk sac. All the

cranial structures increases in length and diameter. At the end of this first period, structural networks have developed and linked the lower jaw apparatus, the hyoid apparatus and the suspensory apparatus together.

At second period (1- 4 weeks old, stage 5- 8), internal skeletal structures form, thicken and strengthen. Rapid growth has also increased the size and thickness of all cartilages, muscles, and the intestine. The larvae must feed exogenously since internal nutrition source from yolk sac is not sufficient for growth and survival. At the end of the period, only small remnant of yolk sac is present.

At third period (5- 10 weeks old, stage 9- 12). At this stage, larval cods start to capture prey. At the end of period, general body shape of larvae have developed very close to young juveniles. Most cranial structures present in juveniles have formed and are mineralizing, but still not completed.

1.4.3 Temperature of the embryonic and larval development

North Atlantic cod has been found to spawn within a range of -1.5 to 12 ℃ (Galloway *et al.* 1998). Early study suggested that the cumulative mortality during embryonic developments of Atlantic cod was minimum within 5- 10 ℃ (Iversen *et al.* 1984). More current studies have confirmed 9.5 ℃ as the optimal incubation temperature for Atlantic cod eggs (Puvanendran *et al.* 2015) and is suitable for larvae development (Juliussen 2016). The incubation temperature can be elevated from 4.5 ℃ to 9.5 ℃ using gradual increment periods 8- 96 h (optimal periods is 32 h), without disturbing or altering the normal embryonic and organic development, and without reducing hatching rate and the quality or quantity of larvae.

1.4.4 The importance of swim bladder in Atlantic cod

Atlantic cod is a physoclistous species that has a closed, compliant swim bladder which accounts for 4%- 5% of the volume of body (Harden Jones *et al.* 1985). Although they normally live in deep ocean, the swim bladder can be filled or emptied to achieve neutral buoyancy at any depth (Midling *et al.* 2012). The exchange of gas in cod swim bladder is achieved by the secretion or excretion of gas from two specialized vascular structures in the swim bladder, gas gland and oval, respectively (Heffernan *et al.* 2004). Without well- developed swim bladder, adult Atlantic cod may suffer damages from barotrauma during vertically transfer, autonomously or manually (Ferter *et al.* 2015). In larval stage, the failure of swim bladder inflation can

be fatal (Howell 1984).

Larvae of Atlantic cod form and fill the swim bladder with air at first feeding around 5 days after hatched (Yin *et al.* 1986). Pneumatic duct closes and therefore the swim bladder is capable of inflation within 6- 15 days from hatched, and no inflation occurs either before or beyond this period. Those cod larvae, which do not inflate their swim bladder, exhibit retarded growth rates during the larval stages and normally die at the start of metamorphosis (Howell 1984). Similar consequence is observed in zebrafish that the deactivation of swim bladder inflation during early stages is also fetal (Teoh et al. 2010). Further study on Atlantic cod has also found the improper speed of increasing water temperature during the incubation may retard the inflation at larval stage (Puvanendran *et al.* 2015).

2. MATERIAL AND METHODS

2.1 Materials

2.11 Chemicals

Table 1. Chemicals that were used in the experiments

Chemicals	Producer
PureLink™ RNA Mini Kit	Thermo Fisher Scientific, USA
PureLink™ DNase Kits	Thermo Fisher Scientific, USA
TaqMan® Reverse Transcription Reagents	Thermo Fisher Scientific, USA
LightCycler®480 SYBR Green I Mater	Roche Applied Science, Germany
70% EtoH (ethanol), 30%DEPC (Dietyl-Poly Carbonate 0.1%)	

2.1.2 Equipment

Table 2. Equipments that were used in the experiments

Equipments	Producer
Precellys 24 Lysis and homogenization	Birtin Technology, France
Centrifuge 5424 and 5415R eppendorf	VWR
LightCycler®480	Roche Diagnostics Gmbh, Germany
AvantiTM J-30 I Centrifuge	Beckman Coulter, USA
Veriti 96 Well Thermal Cycler	Applied Biosystem
NanoDrop 1000 Spectrophotometer	BMG LAVTECH, Ortenberg, Germany

2.2 Methods

2.2.1 Preparation of experimental samples

To investigate the expression level of the key genes (*elovl1a*, *elovl1b*, *pbx1a*, *pbx1b*, *psap*, and *sftpb*) that might be essential for the early development of swim bladder or gas gland in Atlantic cod, eggs were recieved from the National Breeding Program for Atlantic cod, Tromsø, and were sampled as part of the CODE project. Sampling started from unfertiliezation (30-03-2011) to End metamorphosis (Stage 12) (29-06-2011). Temperature for incubation and for laval development was 9 ℃ according to the studies of Puvanendran *et al.* (2015) and Juliussen (2016). Based on the grading method for embryonic stages worked out by Gorodilov *et al.* 2008, the

embryonic development was divided into nine stages in our experiment (Table 3), including Unfertilised Egg, Cleavage, Mid Blastula, Late Blastula, Onset Gastrula, Late Gastrula, Early Somitogenesis, Late Somitogenesis, and Hatching. After hatching, the larval development was divided into three periods including 12 stages based on the studies of Herbing *et al.* (1996a). The larval samples were collected at Stage 4, Stage 8, Stage 11, and Stage 12 (Table 3). A minimum amount of 10 eggs/ larvae per stage were collected. The samples were moved into petri dish and transfered to Eppendorff tube. 250 µl TRK lysis buffer was added to each pipe in chemical hood. 20 µl ProteinaseK was then added and vortexed well, making sure all the eggs were broken. The samples were incubated at 37 ℃ for 90 min, and transferred to -80 ℃ freezer for RNA extraction.

In order to analyze the *sftpb* expression level in different tissues of Atlantic cod, heart, spleen, head kidney, gill, liver, pancrease, brain, testis, ovary, and gas gland samples were collected from grown up Atlantic cod. Three individuals of each gendar were used for tissue- collection. All the samples were stored in -20 ℃ freezer immediately after the collection.

Table 3. Stages of embryonic and larval development in Atlantic cod

Schedule	Stages	ID
30/03/2011	Unfertilised Egg	UE
31/03/2011	Cleavage	C9
31/03/2011	Mid Blastula	MB9
01/04/2011	Late Blastula	LB9
01/04/2011	Onset Gastrula	OG9
02/04/2011	Late Gastrula	LG9
05/04/2011	Early Somitogenesis	ES9
06/04/2011	Late Somitogenesis	LS9
12/04/2011	Hatching	Ha9
09/05/2011	Stage 4 (before metamorphosis)	St4
23/05/2011	Stage 8 (Start metamorphosis)	St8
23/06/2011	Stage 11 (Mid metamorphosis)	St11
29/06/2011	Stage 12 (End metamorphosis)	St12

2.2.2 RNA isolation and measurement

The total RNA from Atlantic cod heart, spleen, head kidney, gill, liver, pancreas, brain, testis, ovary, and gas gland tissue was isolated using PureLink™ RNA Mini

Kit. Add approximately 20 mg tissue to a 2 ml tube with beads and 600 µl lysis buffer, and then homogenize the sample in Precyllis 24 at 5500 rpm for 2 × 30 sec. After centrifuge at 12, 000 g for 2 min at room temperature, transfer 500 µl of the supernatant to the micro- tube containing 70% EtOH and mix well. The mixture was then transferred to the filter tube and centrifuged at 12, 000 g for 30 sec for RNA attaching. Centrifuge again at 12, 000 g after 300 µl was buffer I was added. For efficient removal of DNA from RNA, 80 µl DNase solution (PureLink™ DNase kit) was added, incubating at room temperature for 15 min. Thereafter, the RNA was washed by 300 µl wash buffer I once and 500 µl wash buffer II twice at 12, 000 g for 30 sec. For the removal of the redundant reagent, the tube was centrifuged empty in 12, 000 g for 1 min, and dried in laminar flow cabinet for 2 min. 80 µl RNase free water was added and incubated at room temperature for 1 min. Afterwards, RNA was eluted by centrifugation at 12, 000 g for 2 min. The eluted RNA samples were kept at -70 ℃ for use.

The RNA concentration and quality was measured by NanoDrop 1000 Spectrophotometer. 260/280 ratio was used to assess the purity of RNA, and 260/230 was used to assess the nucleic acid purity. In our experiments, high quality RNA was for further experiments. The 260/280 ratio and 260/230 ratio was between 2.1~ 2.3 and 1.7~ 2.3, respectively, and the concerntration was higher than 60 ng/ µl.

2.2.3 cDNA synthesis and quantitative real time PCR (qPCR)

Reverse transcription was performed by TaqMan® Reverse Transcription Reagents. A 10 µl mixture per sample was made and added into 96 well plate, each mixture contains 150 ng RNA. The mixture was made as Table 4. Before cDNA synthesis, the RNA samples were centrifuged at AvantiTM J-30 I Centrifuge at 1500 g for 2 min. The cDNA synthesis was processed in Veriti 96 Well Thermal Cycler machine under setting of 10 min at 25 ℃, 60 min at 48 ℃, and 5 min at 95 ℃. The synthesized cDNA was diluted (1: 10) with nuclease- free H_2O to a final volume of 100 µl, and then stored at -20 ℃ for quantitative real-time PCR (qPCR) study.

Table 4. Mixture setting for cDNA synthesis

Reagens	1x (10 µl)
10 ×RT-buffer	1
$MgCl_2$ (25 mM)	2.2
dNTP (2.5 mM)	2

Oligo-dT	0.25
Hexamer	0.25
Rnase inhibi	0.35
RT	0.2
RNA+RNase free water (150 ng/ μl)	3.75

To validate the expression level of 6 target genes (*elovl1a, elovl1b, pbx1a, pbx1b, psap, sftpb*) in samples, qPCR was performed using LightCycler® 480 system. The reaction mix for qPCR consisted of 4 μl of diluted cDNA, 5 μl SYBR Green I Master chemistry, 0.5 μl forward primer, and 0.5 μl reverse primer. The amplification program consisted of 40 cycles of amplification (95 ℃ for 15 sec and 60 ℃ for 30 sec), 1 cycle of melting curve (95 ℃ for 5 sec, 60 ℃ for 1 min, and 97 ℃ continue). PCR amplification efficiency of target genes were measured by running a template dilution series. The cDNA samples from testis were diluted by 5 series of 1:1, 1:5, 1:25, 1:125, 1:625 and then used as templates for 7 primers of 6 target genes and 1 reference gene: ubiquitin. Ubiquitin was chosen as the internal control gene because of its stable expression in Atlantic cod (Sæle *et al.* 2009). After PCR amplification, the efficiency of different primers were measured by the LightCycler®480 and calculated by the equation $E= 10^{[-1/slope]}$ (Meuer *et al.* 2012) (Appendix 2). The primers of target genes were designed by Primer Premier 5 (Table 5). Ubiquitin was chosen as the internal control gene because of its stable expression in Atlantic cod (Sæle *et al.* 2009)

Table 5. Sequences of primers used for qPCR

Gene	Forward primer (5'- 3')	Reverse primer (3'- 5')
elovl1a	TTCCTCTTCGCCCACTTC	AGCACTTTGCCGTTTGAC
elvo11b	GGTCATTCACCTCATCTGGGTA	CACTGTGCTTGCCGTTGG
pbx1a	CCCGCTACTCCAAACTCT	CCGCCTGTCTGGTTGATA
pbx1b	GAGATCGAGCGCATGGTG	TCCGTCGCCTGCTTGTTG
psap	AGACAGAGTGTGACCAGCTC	CCCCACTTTCATGCACACAA
sftpb	GTTGGTTCTCAGCAGGCTTC	GACAGACAGACCCGAGGAAA
Ubiquitin	GGCCGCAAAGATGCAGAT	CTGGCTCGACCTCAAGAGT

2.2.4 Relative expression of target genes

The log2 Pfaffl method was accounted for actual PCR efficiency (Pfaffl 2001). As the Ct value of the chosen reference gene (Ubiquitin) was the same in the control

and the sample, Pfaffl value was calculated using the equation of Pfaffl value= $(E_{target})^{\Delta Ct\ target\ (control-\ sample)}$ (Pfaffl 2001). E_{target} represents the PCR amplification efficiency of target genes, and Ct value represents the number of cycles required for the fluorescent signal to cross the threshold. Ct values higher than 36, which indicated a weak reaction and insufficient amount of target genes, were excluded.

2.2.5 Studies on genomic structure, phylogeny, and synteny

Six target genes (*elovl1a*, *elovl1b*, *pbx1a*, *pbx1b*, *psap*, and *sftpb*) related to the development of swim bladder were selected for the study in Atlantic cod. Transcript sequences of these genes in Atlantic cod were original retrieved from Ensembl database (http://www.ensembl.org), and then were predicted by Basic Local Alignment Search Tool (BLAST) in Ensembl, NCBI (https://www.ncbi.nlm.nih.gov/), and CEES database (http://cees-genomes.hpc.uio.no). Relatively complete transcript sequences were selected from one of the database, and most related genes were selected for nucleotide homology. Correspondent proteins of the genes retrieved from CEES were predicted by NCBI orffinder online tool. The corresponding protein sequences were used for multiple alignments (Appendix 2). The ClustalW alignment method worked in Mega7 was used for the analysis of genetic structure, the Muscle alignment method worked in Mega7 was used for the analysis of phylogeny, and the Mauve alignment method worked in Megalign Pro was used for synteny analysis between the partial genome of Atlantic cod and Stickleback. Phylogenetic tree was built by Mega7 (Construct/Test Neighbor Joining Tree). The conserved synteny of *sftpb* gene was analyzed on Genomicus v88.01 platform (www.ibens.ens.fr/). Protein structure was predicted with the SMART online tool (http://smart.embl-heidelberg.de/) or Pfam online tool (http://pfam.xfam.org/). The functional sites in proteins were predicted with the Eukaryotic Linear Motif resource online tool (http://elm.eu.org/), and the signal peptide site was predicted with SignalP 4.1 on line tool (http://www.cbs.dtu.dk/). All the reference proteins used for analysis were listed in Appendix 1.

2.2.6 Data analysis

Microsoft Office Excel 2013 was used to integrate data, and IBM SPSS statistics 19 was used for One-way analysis of variance (ANOVA) ($P < 0.05$ means significant difference) and presented in means ± standard errors. Adobe Illustrator CS5, together

with Microsoft Office Excel 2013, was used for figures editing. GeneDOC was used for modification and export of the alignment.

3. RESULTS

3.1 Relative expression of target genes

3.1.1 Expression of *sftpb* in different tissues

The expression of cod *sftpb* gene was quantified by real- time qPCR in nine different tissues in adult females and males sampled in March 2016 (Fig. 3). The relative expression of cod *sftpb* was detectable in all nine tissues, and no significant differences were detected between males and females, except for testis and ovary. *Sftpb* was much expressed in testis than in ovary.

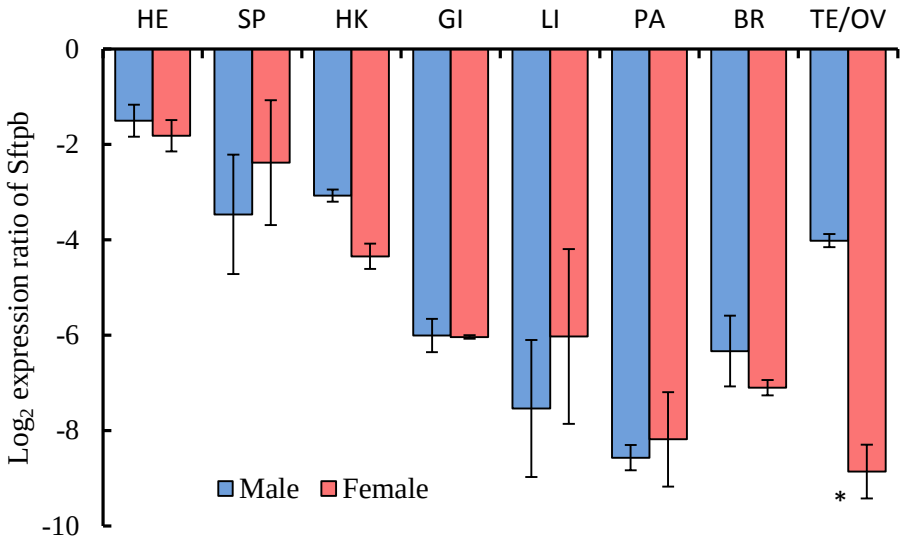


Fig. 3. The overview of *sftpb* gene relative expression level in different tissues between female and male Atlantic cod. HE, SP, HK, GI, LI, PA, BR, TE, and OV represents heart, spleen, head kidney, gill, liver, pancreas, testis, and ovary, respectively. Asterisks indicate a significant difference between females and males: n=3, Welch's two-sample t test, P <0.01

Without regard to the gender, *sftpb* was mostly enriched in heart, followed by spleen, head kidney, gas gland, and testis, while *sftpb* in ovary, pancreas, brain, and liver, and gill were weak in expression (Fig. 4). Significant difference was detected between the less- expressed tissues and most- expressed tissues.

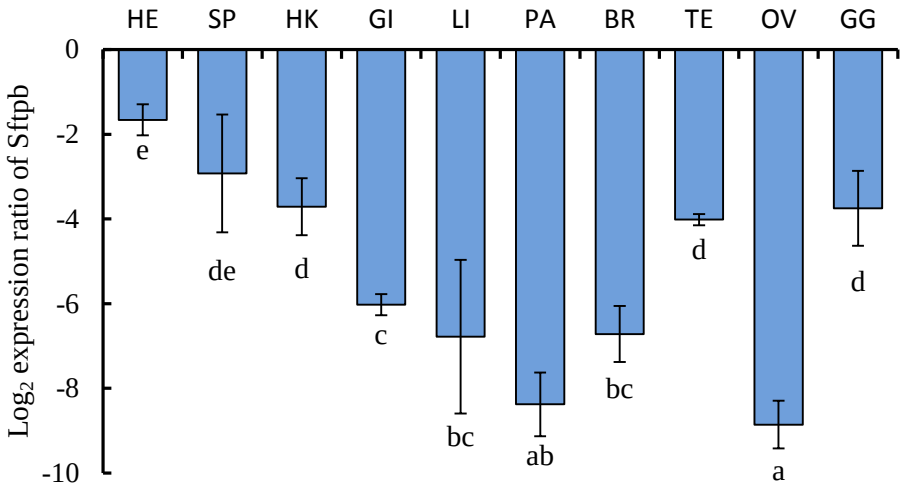


Fig. 4. The overview of *sftpb* gene relative expression level in different tissues of Atlantic cod. HE, SP, HK, GI, LI, PA, BR, TE, and OV represents heart, spleen, head kidney, gill, liver, pancreas, testis, and ovary, respectively. Different letters indicate a significant difference between tissues: n=3, Duncan, P <0.05

3.1.2 Expression of key genes in swim bladder at early stages of Atlantic cod

The expression of key genes in the swim bladder was quantified at the embryonic and larval stages of Atlantic cod by real- time qPCR (Fig. 5). Relative expression of cod *elovl1a* peaked at around 0 (log 2 ratio) the first three stages, Unfertilized egg (UE), Cleavage (C9), and Mid Blastula (MB9) (Fig. 5). At next two stages of Late Blastula (LB9) and Onset Gastrul (OG9), the expression level of *elovl1a* dropped dramatically. *Elovl1a* kept down- regulated significantly (P < 0.05) at Late Gastrula, and the tendency lasted until Early Somitogenesis (ES9). Embryonic *elovl1a* level tend to stay stable since ES9, although a modest growth there was followed by Late Somitogenesis (LS9), Hatching (Ha9), Stage 4 (before metamorphosis- St4), Stage 8 (Start metamorphosis-St8). Significant difference (P < 0.05) was detected between ES9 and post- Hatching stages (Hatching- Stage 8). At larval stage, *elovl1a* started to fall at Stage 8, Stage 11 (Mid metamorphosis- St11), and Stage 12 (End metamorphosis- St12), significant difference (P < 0.05) existed between Stage 8 and Stage 12. Relative expression of embryonic *elovl1a* decreased continuously and dramatically, while *elovl1a* was maintained at relatively low and stable levels at larval stage. At larval stage of Atlantic cod, *elovl1a* was most enriched at stage 8.

Elovl1b seemed to have a similar tendency with *elovl1a* at early stages of Atlantic cod. *Elovl1b* stayed at around 0 log2 ratio at the first three stages: UE, C9, and MB9, but dropped dramatically until ES9 (Fig. 5). Since then, embryonic *elovl1b* level increased smoothly till the end of larval stage. Larval *elovl1b* Peaked at St12 with the log2 ratio of around -4.6, although no significant difference was there between St12 and ES9.

Pbx1a showed the highest level at UE stage with the log2 ratio of -2.83 (Fig. 5). The expression level of *pbx1a* dropped significantly ($P < 0.05$) at C9, MB9, LB9 stages. The dramatically falling continued at OG9, and eventually stopped at LG9, with the lowest log2 ratio of -16.6. The expression level then increased significantly until hatching. After Ha9, the larval *pbx1a* was maintained at stable level with the log2 ratio around -8.

Pbx1b peaked at C9 and MB9 stages with the log 2 ratio around -1 (Fig. 5). At the next stage, LB9, the expression of *pbx1b* seemed to be unstable. At stages of OG and LG, the expression dropped slightly, and made no difference with C9 or MB9. After a slightly decrease, the relative expression of *pbx1b* tended to be stable at stages of ES9, LS9, Ha9, and larval stages, with the log 2 ratio around -5.

The relative log2 ratio of *psap* was above 0 before LB9 stage (Fig. 5). The expression of *psap* seemed to be unstable at LB9 and had significant difference ($P < 0.05$) with stages of UE, C9, and MB9. The expression of *psap* declined continuously and significantly ($P < 0.05$) since MB9 to ES9. However, no significant difference was found during stages of LG9 to St12, and the relative log2 ratio was between -2 and -4.

Like the other genes in Fig. 5, *sftpb* decreased sharply during the early embryonic stage. Significant differences was detected between every two stages from UE to OG9 (the log2 ratio ranged from -2.8 to -8.8). *Sftpb* gene expressed stably during the stages from OG9 to LS9, with the log2 ratio ranged from -8.8 to -9.1. Interestingly, *sftpb* level rose significantly after hatching ($P < 0.05$), and remained stable during larval stage, with the log2 ratio ranged from -6.1 to -6.9.

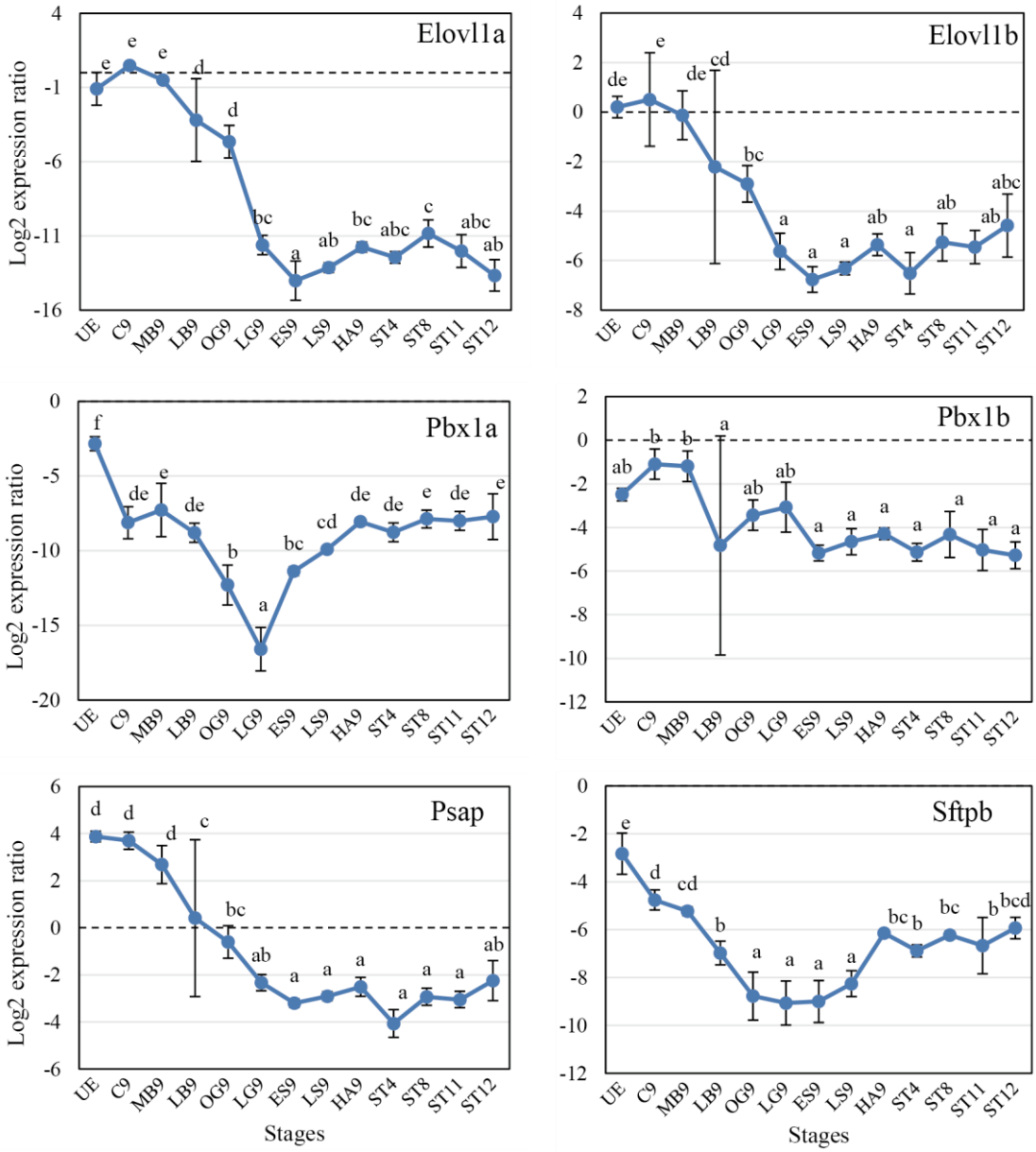


Fig. 5. Relative expression level of Atlantic cod *elovl1a*, e*lovl1b*, *pbx1a*, *pbx1b*, *psap*, and *sftpb* genes in different development stages. UE, C9, MB9, LB9, OG9, LG9, ES9, HA9, ST4, ST8, ST11, and ST12 represents Unfertilised Egg, Cleavage, Mid Blastula, Late Blastula, Onset Gastrula, Late Gastrula, Early Somitogenesis, Late Somitogenesis, Hatching, Stage 4 (before metamorphosis), Stage 8 (Start metamorphosis), Stage 11 (Mid metamorphosis), ans Stage 12 (End metamorphosis), respectively. Different letters indicate a significant difference between tissues: n=3, Duncan, P <0.05

3.2 Genetic characterization, orthology, and phylogenetic analysis

3.2.1 Atlantic cod *psap*

Atlantic cod *psap* (Gene ID: ENSGMOG00000017246), zebrafish *psap* (Gene ID: ENSDARG00000013968), and human *Psap* (Gene ID: ENSG00000197746) were

identified in Ensembl database. Full length transcript sequence of Atlantic cod *psap* was predicted by BLAST tool in Atlantic cod genome database CEES. Atlantic cod *psap* (Transcript ID: GAMO_00033032_RA) consists of 14 exons and codes for a protein of 663 aa (Fig. 6). Similar to Atlantic cod *psap*, zebrafish *psap* (Transcript ID: ENSDART00000045069.6) also consists of 16 exons, although intron 4- 5 (1 bp) and intron 5- 6 (2 bp) were too short to display in Fig 9. Human *Psap* (Transcript ID: ENST00000394934.4) consists of 14 exons. The last exon of each of the three gene contains both untranslated UTR portion and very short translated portion at 3' downstream. No UTR portion was found in the first exon at 5' upstream of Atlantic cod *psap*.

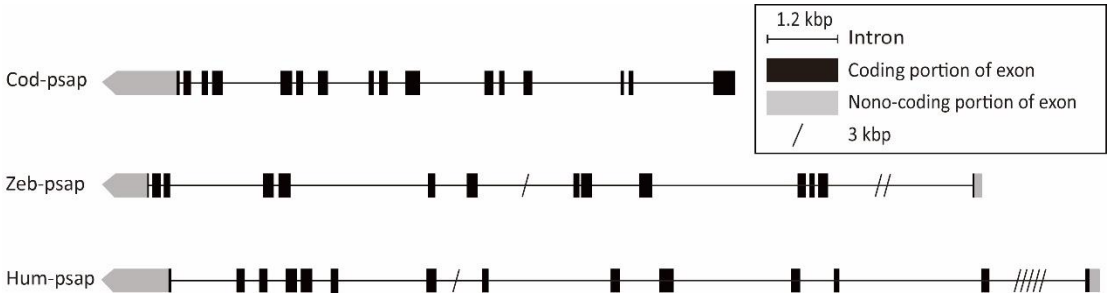


Fig. 6. Schematic representation of genetic organization for full-length *psap* gene in different organisms (Cod- Atlantic cod, Zeb- zebrafish, Hum- human). Boxes represent exons, while lines represent introns. The gray and black colors are used to distinguish non-coding and coding portion of exons, respectively.

Atlantic cod Elovl1 proteins was predicted by NCBI orffinder online tool. Multiple alignment of Atlantic cod Psap protein with orthologous sequence from zebrafish and human was performed to investigate the conserved domains (Fig. 7). By using SignalP 4.1 on line tool , the signal peptide (residues in red box) was predicted in each Psap protein. In cod Psap, a total of 135 translated amino acid residues present before the signal peptide, which makes difference from zebrafish or human Psap protein. A total of six domains from superfamily of Saposin-like protein were predicted in SMART online tool. The first and last domain, are peripheral domains termed as Saposin A- type domain (Accession ID: SM000162) that link the central domains at N- terminus and C- terminus, respectively. The central domains, termed as Saposin B- type domains (Accession ID: SM000741), are linked in tandem by three connecting regions. The connecting regions of the central domains can be cleaved to

liberate four saposins (Sap- A, B, C, and D, respectively) into the lysosomes.

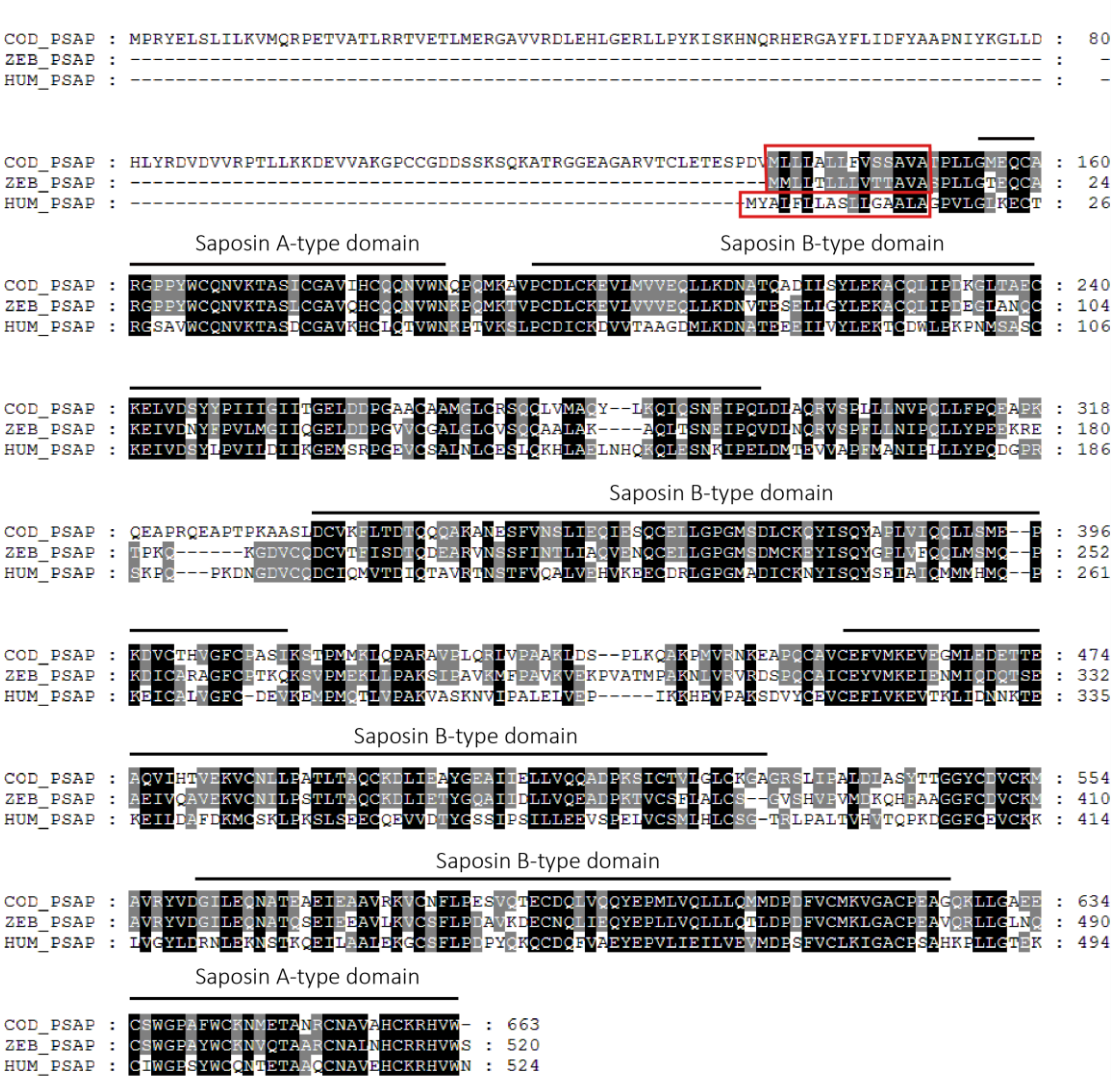


Fig. 7. ClustalW multiple alignment of predicted Atlantic cod Psap proteins with orthologous sequences from zebrafish, and human. Black shading with white font is used to denote identical residues. Gray shading with white font is used for residues with 80% conservative substitution. Light gray with black font specifies that 60% conservative substitutions. Residues under black line representing saposin A- type and B- type domain from superfamily of saposin- like protein predicted in SMART were highly conserved. Residues in red box represents the signal peptide of the protein.

As shown in the phylogenetic tree (Fig. 8), Atlantic cod Psap was clustered with its putative orthologues of other fish species. Atlantic cod Psap grouped with Psap from most of Actinopterygii (Osteichthyes) fish species, including large yellow croaker, Asian sea bass, stickleback, fugu, tilapia, turquoise killfish, amazon molly, platyfish, northern pike, Atlantic herring, zebrafish, and cave fish, except for the

spotted gar. Interestingly, Spotted gar Psap had a relatively less intimate evolutionary relationship with other Actinopterygii fish species than Elephantfish Psap, a member from Chondrichthyes. In other words, cod Psap is relatively close to elephantfish Psap during the long- term evolution. However, cod Psap was relatively distant from coelacanth or shark Psap than elephantfish or spotted gar Psap in evolutionary history.

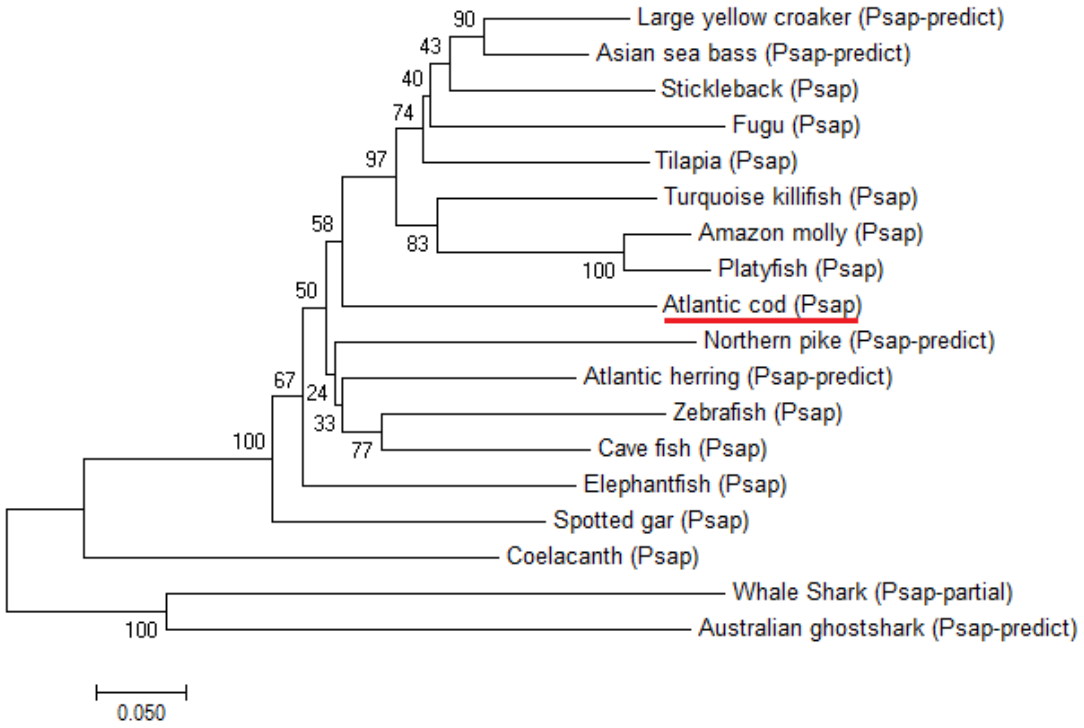


Fig. 8. Phylogenetic analysis of the Atlantic cod Psap. The predicted proteins of cod Psap were aligned against homologous proteins from other fish species using MEGA7. Based on the multiple sequence alignment, an unrooted phylogenetic tree was constructed by the maximum likelihood method. The tree was bootstrapped 5000 times, and the bootstrap values are shown at the branch points.

3.2.2 Atlantic cod *sftpb*

Atlantic cod *sftpb* gene (Gene ID: ENSGMOT00000012190) was identified in Ensembl online database. Full length transcript was predicted by BLAST tool in CEES. Zebrafish *sftpba* (Gene ID: ENSDARG00000094708) and *sftpbb* (Gene ID: ENSDARG00000067566), and human *sftpb* (Gene ID: ENSG00000168878) were identified. Atlantic cod *sftpb* (Transcript ID: GAMO_00041380_RA) consists of 9 exons that codes for a 317 aa protein (Fig. 9). Both untranslated UTR portion and translated portion were contained in the first and last exon at 5’ upstream and 3’

downstream respectively. Zebrafish *sftpba* (Transcript ID: ENSDART00000141824.1) contains 8 exons, including 7 coding exons and an UTR at the 3' downstream, while *sftpbb* (Transcript ID: ENSDART00000097254.4) consists of 10 coding exons including two UTR portions in the first and last exon, respectively. Human *sftpb* (Transcript ID: ENST00000393822.7) consists of 10 coding exons with an UTR at both ends.

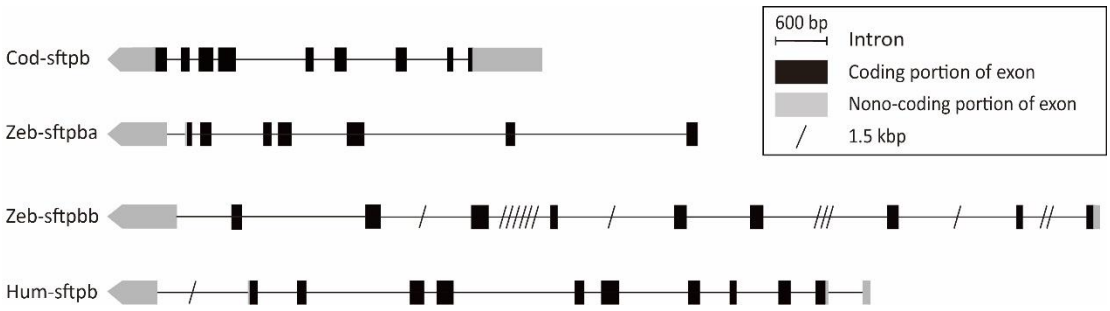


Fig. 9. Schematic representation of genetic organization for full-length *sftpb* gene in different organisms (Cod- Atlantic cod, Zeb- zebrafish, Hum- human). Boxes represent exons, while lines represent introns. The gray and black colors are used to distinguish non-coding and coding portion of exons, respectively.

Atlantic cod Sftpb proteins was predicted by NCBI orffinder online tool. Multiple alignment of Atlantic cod Sftpb protein with orthologous protein sequence from zebrafish and human was performed, and showed in Fig. 10. SMART online tool was used to predict the conserved domains. One Saposin A- type domain (Accession ID: SM000162) was predicted at 5'downstream, and two Saposin B- type domains (Accession ID: SM000741) were predicted before Saposin A- type domain as shown in Fig. 10. SignalP 4.1 on line tool was used to predict the signal peptide (residues in red box), which existed in each Sftpb protein except for zebrafish sftpbb.

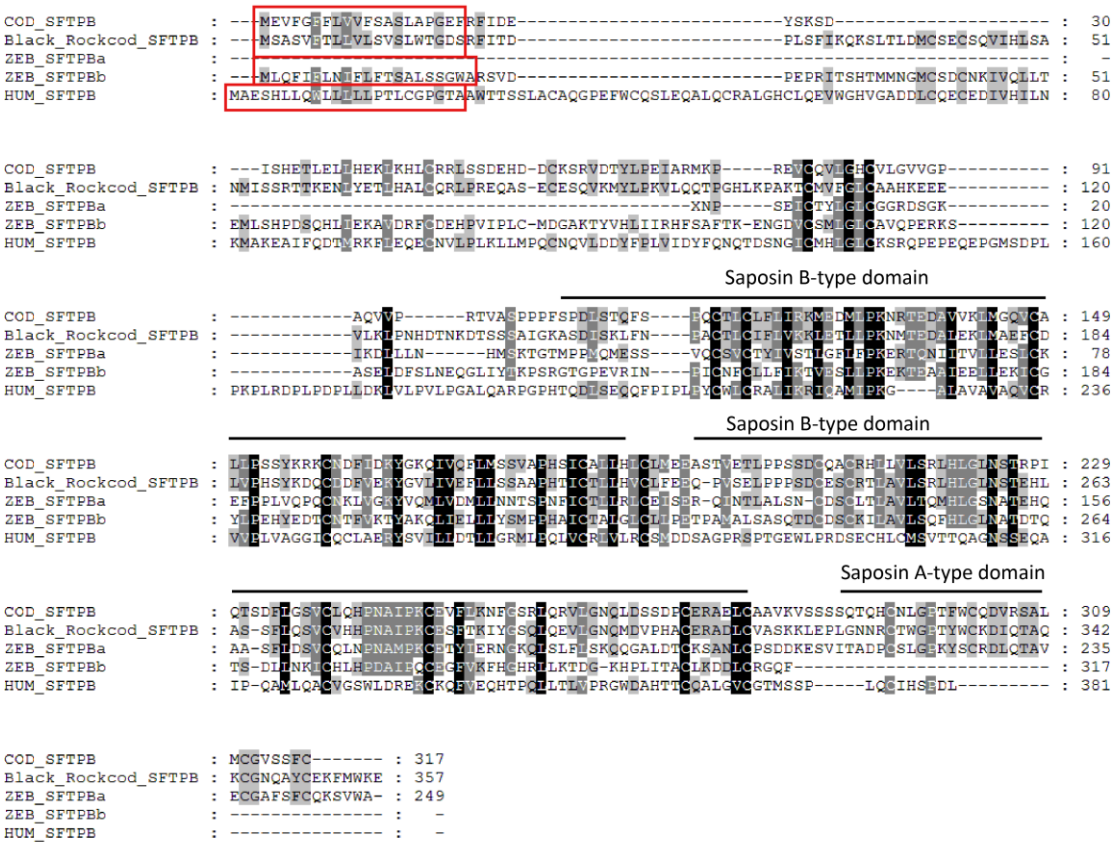


Fig. 10. ClustalW multiple alignment of predicted Atlantic cod Sftpb proteins with orthologous sequences from black rockcod, zebrafish, and human. Black shading with white font is used to denote identical residues. Gray shading with white font is used for residues with 80% conservative substitution. Light gray with black font specifies that 60% conservative substitutions. Residues under black line representing saposin A- type and B-type domain from superfamily of saposin- like protein predicted in Pfam were highly conserved. Residues in red box represents the signal peptide of the protein.

In the molecular phylogenetic tree, fish Sftpb and Psap protein were closely clustered, and all fish proteins were classified into two major groups with each Atlantic cod Sftpb or Psap clustered with its putative orthologues of other fish species (Fig. 11). Sftpb protein consists of two major groups. Atlantic cod Sftpb protein grouped with Mudskipper, Swamp eel, and Black rockcod Sftpb, and other fish predicted surfactant or prosaposin- like protein, including prosaposin- like protein in Ballan wrasse, predicted Sftpb in killfish, Pundamilla nyererei, Zebra mbuna, Medaka, Platyfish, Tetraodon, Stickleback, and Amazon molly. Zebrafish sftpba, Cave fish sftpba, and predicted Japanese puffer Sftpb grouped together with predicted Sftpb on another chromosome in Medaka, Platyfish, Tetraodon, Stickleback, Amazon

molly. Atlantic cod Sftpb has a distant relation with coelacanth and Whale shark Sftpb. Interestingly, Whale shark Sftpb was not only close to whale shark Psap, but also grouped with other fish Psap proteins.

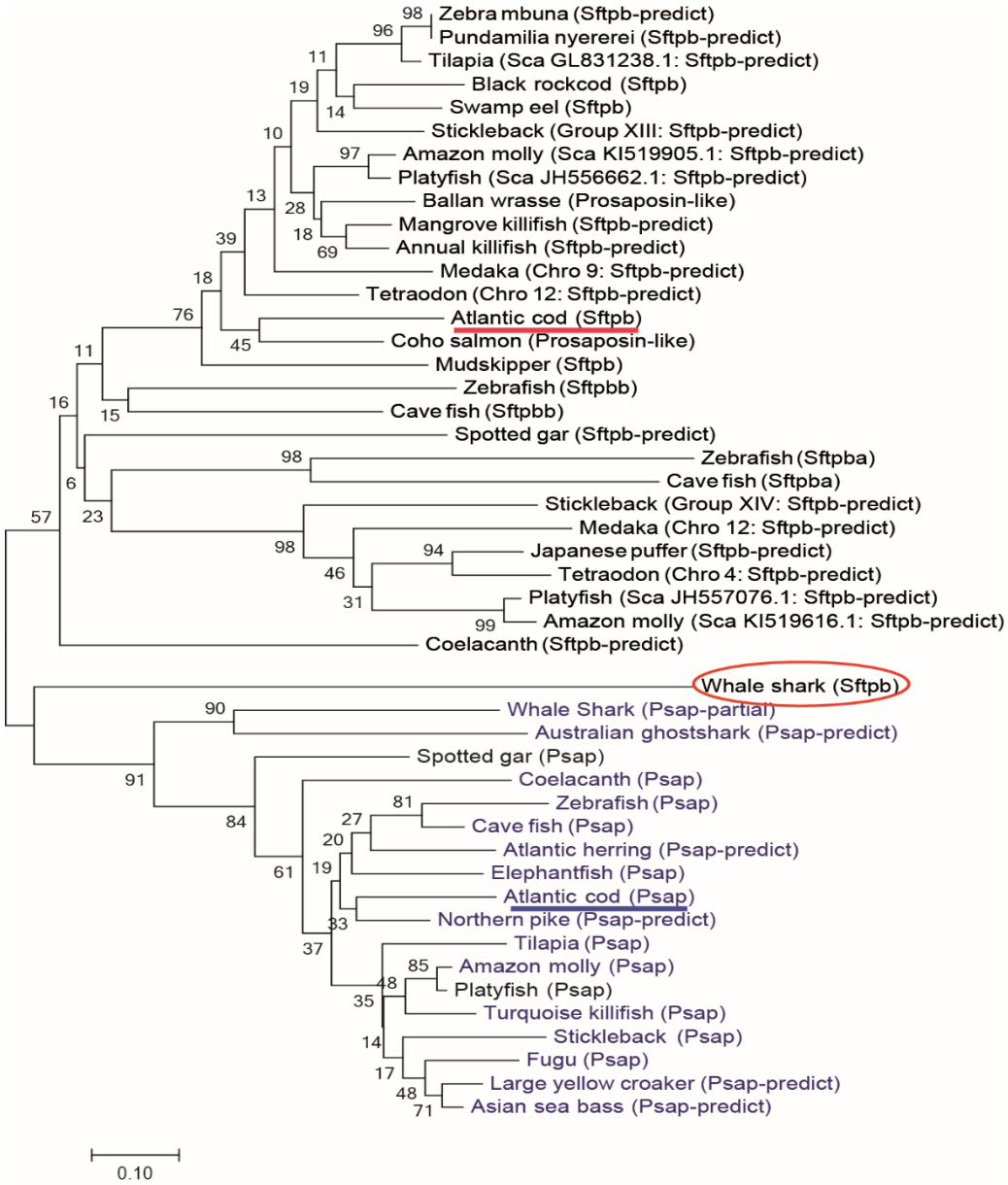


Fig. 11. Phylogenetic analysis of the Atlantic cod Psap and Sftpb. Whale shark Sftpb was circled because of its special position in the tree. The predicted proteins of cod Psap and Sftpb were aligned against homologous proteins from other fish species using MEGA7. Based on the multiple sequence alignment, an unrooted phylogenetic tree was constructed by the maximum likelihood method. The tree was bootstrapped 5000 times, and the bootstrap values are shown at the branch points.

3.2.3 Atlantic cod *elovl1a* and *elovl1b*

Two Atlantic cod *elovl1* genes were identified in Ensembl online database (http://www.ensembl.org/), named *elovl1a* (Gene ID: ENSGMOG00000010488) and *elovl1b* (Gene ID: ENSGMOG00000018234). Full length transcript sequences were predicted by BLAST tool in Atlantic cod genome database CEES (http://cees-genomes.hpc.uio.no/). Zebrafish *elovl1a* (Transcript ID: ENSDART00000166073.1) and *elovl1b* (Transcript ID: ENSDART00000169052.1), human *Elovl1* (Transcript ID: ENST00000621943.4), and mouse *elovl1* (Transcript ID: ENSMUST00000006557.12) were also identified in Ensembl database. Atlantic cod *elovl1a* (Transcript ID: GAMO_00033084_RA) consists of 8 exons and codes for a protein of 317 amino acid residues (aa), while Atlantic cod *elovl1b* (Transcript ID: GAMO_00029213_RA) consists of 9 exons and codes for a protein of 319 aa (Fig. 12, 4). Zebrafish, human, and mouse *elovl1* genes consist of 8 exons respectively, show the similar structures with cod *elovl1a*, although zebrafish *elovl1* genes contain much longer introns (Fig. 12). Each *elovl1* genes of Atlantic cod, zebrafish, human, and mouse consists of one untranslated region (UTR) at both 5' upstream and 3' downstream side.

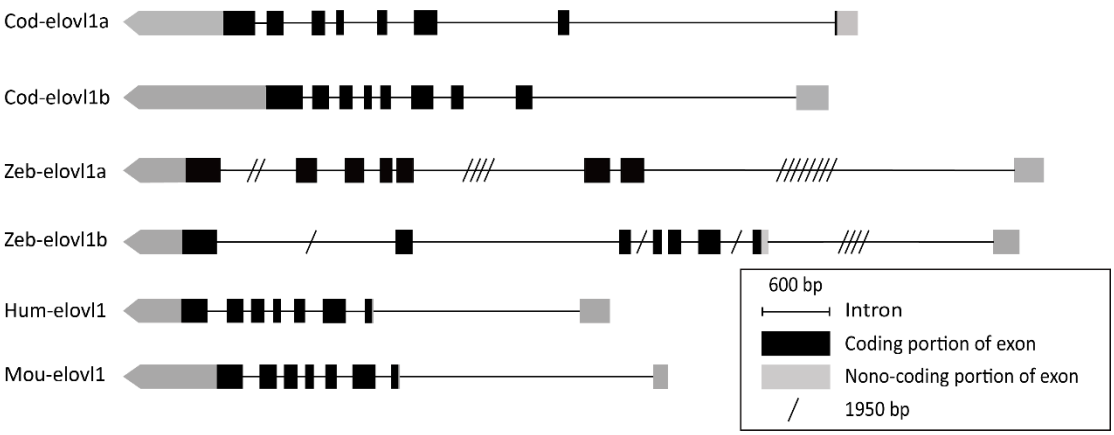


Fig. 12. Schematic representation of genetic organization for full-length *elovl1* gene in different organisms (Cod- Atlantic cod, Zeb- zebrafish, Hum- human, Mou- mouse). Boxes represent exons, while lines represent introns. The gray and black colors are used to distinguish non-coding and coding portion of exons, respectively.

Multiple alignment of Atlantic cod Elovl1 proteins with orthologous sequences from zebrafish, human, and mouse (Fig. 13) revealed highly conserved GNS1/ ELO family protein (Accession NO. PF01151). Members of GNS1 family play a part in

long chain fatty acid elongation systems that produce C26 precursors for the synthesis of ceramide and sphingolipid (Oh *et al.* 1997). Seven transmembrane domains (with black lines on the top of the sequences) from GNS1 family were identified. The signal peptide (residues in red box) was also identified in each Elovl1 protein, by SignalP 4.1 on line tool (http://www.cbs.dtu.dk/).

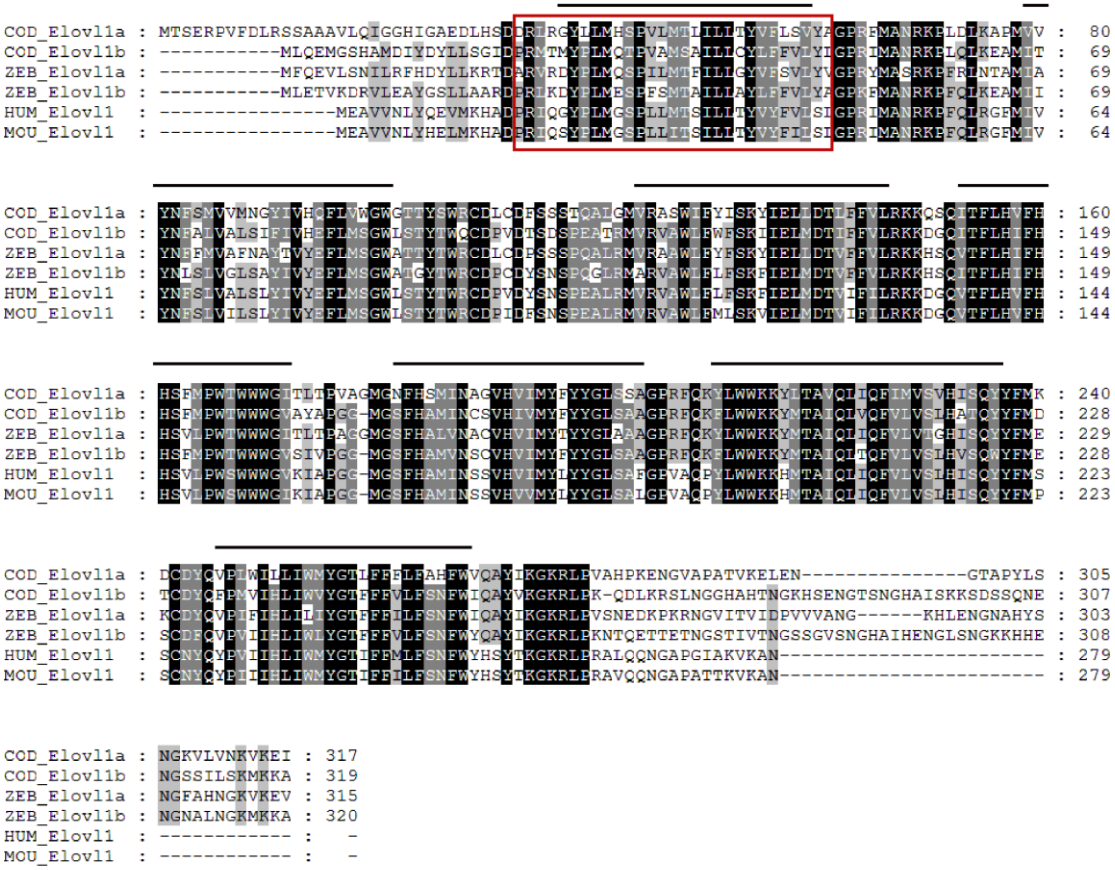


Fig. 13. ClustalW multiple alignment of predicted Atlantic cod Elovl1 proteins with orthologous sequences from zebrafish, human, and mouse. Black shading with white font is used to denote identical residues. Gray shading with white font is used for residues with 80% conservative substitution. Light gray with black font specifies that 60% conservative substitutions. Seven transmembrane domain from GNS1 family with black line on the top, predicted in Pfam, are highly conserved in all organisms from this figure. Residues in red box represents the signal peptide of the protein.

In the molecular phylogenetic tree, all fish Elovl1 proteins were classified into two major groups with each Atlantic cod Elovl1a or Elovl1b clustered with its putative orthologues of other fish species (Fig. 14). Three major groups with cod

Elovl1b clustered with its putative orthologues of other fishes. Ballan wrasse Elovl1, European bass Elovl1, Atlantic cod Elovl1b, Mummichog Elovl1, Beira killfish Elovl1b, zebrafish elovlb, and Atlantic herring predicted Elovl were grouped together. Coelacanth predicted Elovl1, Whale shark Elovl1, and Australian ghostshark predicted Elovl1 were highly closed to each other and shared a single branch. Spotted gar Elovl1 was the only member in the third group of Elovl1b, and was relatively distant from the other two groups. Atlantic cod Elovl1a clustered with zebrafish Elovl1a, which was grouped together with Atlantic cod Elovl1a in single branch.

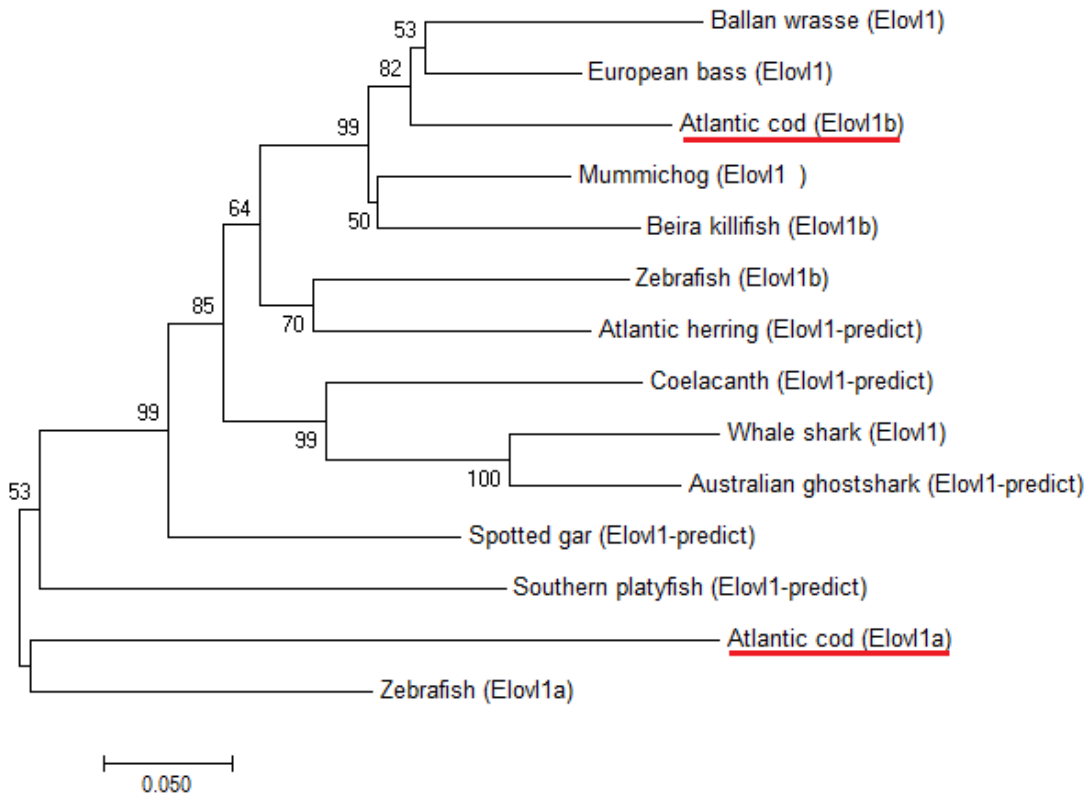


Fig. 14. Phylogenetic analysis of the Atlantic cod Elovl1. The predicted proteins of cod Elovl1 were aligned against homologous proteins from other fish species using MEGA7. Based on the multiple sequence alignment, an unrooted phylogenetic tree was constructed by the maximum likelihood method. The tree was bootstrapped 10000 times, and the bootstrap values are shown at the branch points.

3.2.4 Atlantic cod *pbx1a* and *pbx1b*

Atlantic cod *pbx1a* (Gene ID: ENSGMOG00000010178) and *pbx1b* (Gene ID: ENSGMOG00000007627), zebrafish *pbx1a* (Gene ID: ENSDARG00000100494) and *pbx1b* (Gene ID: ENSDARG00000101131), and human *pbx1* (Gene ID:

ENSG00000185630) were identified in Ensembl database. Transcript sequence of Atlantic cod *pbx1a* (Transcript ID: ENSGMOT00000011191.1) consists of 11 exons and codes for a protein of 411 aa (Fig. 15). In Atlantic cod *pbx1a* of Fig. 15, exon 6 (counted from 5' upstream on the right side) consists of only four base pairs (bp), which was CTCC. Intron 3- 4 and intron 6- 7 consists of 48 bp and 1 bp, respectively, which were both too short to display. Transcript sequence of Atlantic cod *pbx1b* (Transcript ID: ENSGMOT00000008396.1) consists of 10 exons with 1,254 bp transcript length and 417 aa translation length. Zebrafish *pbx1a* (Transcript ID: ENSDART00000165698.1), *pbx1b* (Transcript ID: ENSDART00000165223.1) and human *pbx1* (Transcript ID: ENST00000420696.6) transcripts consist of nine, ten, and nine exons, respectively. Each of these three gene contains one UTR at 3' downstream (Fig. 15). Besides, the first exon of zebrafish *pbx1a* or *pbx1b* contains both translated portion and untranslated portion.

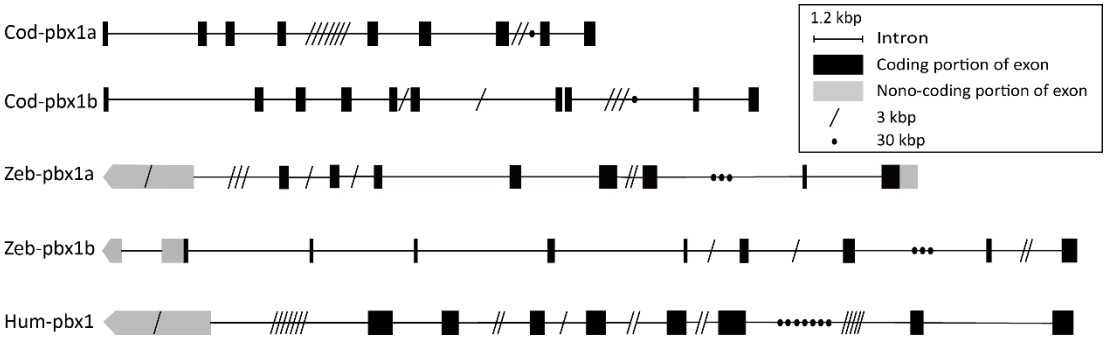


Fig. 15. Schematic representation of genetic organization for full-length *pbx1* gene in different organisms (Cod- Atlantic cod, Zeb- zebrafish, Hum- human). Boxes represent exons, while lines represent introns. The gray and black colors are used to distinguish non-coding and coding portion of exons, respectively.

Multiple alignment of Atlantic cod Pbx1 proteins with orthologous sequences from zebrafish and human revealed highly conserved domains. PBC family (Accession NO. PF03792) and homeobox (HOX) domain (Accession NO. PF00046) were predicted in Pfam online tool (Fig. 16). PBC domain of Atlantic cod Pbx1a and Pbx1b protein is located at 29 to 213 aa, and 30 to 218 aa, respectively. The HOX domian of Atlantic cod Pbx1a and Pbx1b protein is at 216 to 281 aa, and 221 to 284 aa, respectively. PBC domain is a member of the TALE superclass of homeodomain proteins which are involved in growth and differentiation during vertebrate embryogenesis and tumorigenesis (Burglin *et al.* 1992). HOX genes encode

homeodomain protein products that are transcription factors sharing a characteristic protein fold structure that binds DNA (Bürglin *et al.* 2016). Helix- turn- helix motif domain (with red line on the top of the sequences) that functions as DNA- binding motif was also predicted. No signal peptide was identified in Pbx1 protein.

Fig. 16. ClustalW multiple alignment of predicted Atlantic cod Pbx1 proteins with orthologous sequences from zebrafish, and human. Black shading with white font is used to denote identical residues. Gray shading with white font is used for residues with 80% conservative substitution. Light gray with black font specifies that 60% conservative substitutions. Residues under black line representing PBC domain and Homeobox domain predicted in Smart/Pfam are highly conserved in all organisms from this figure. Residues under red line represents the Helix- turn-helix motif.

In the phylogenetic tree, Atlantic cod Pbx1a and Pbx1b were clustered with their putative orthologues of other fish species (Fig. 17). Atlantic cod Pbx1a clustered with Platyfish Pbx1a, Tilapoia Pbx1a, Tetrodon Pbx1a, Fugu Pbx1a, zebrafish Pbx1a, and Coelacanth Pbx1 grouped together. Interestingly, Atlantic cod Pbx1b clustered with

zebrafish Pbx1b was closer to Atlantic cod Pbx1a group than clustered with Ballan wrasse Pbx1, Tilapia Pbx1b, or Platyfish Pbx1b. Platyfish Pbx1b has further distance than any other fish species in this figure.

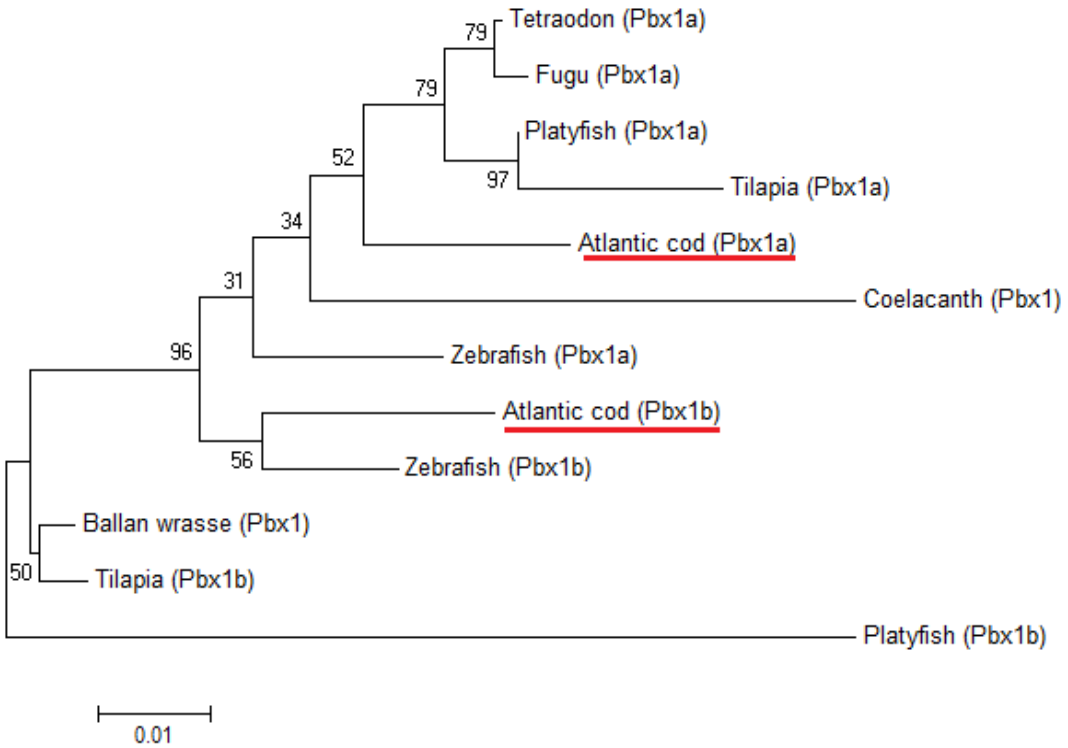


Fig. 17. Phylogenetic analysis of the Atlantic cod Pbx1. The predicted proteins of cod Pbx1 were aligned against homologous proteins from other fish species using MEGA7. Based on the multiple sequence alignment, an unrooted phylogenetic tree was constructed by the maximum likelihood method. The tree was bootstrapped 10000 times, and the bootstrap values are shown at the branch points.

3.3 Syntenic analysis of Sftpb

Genomicus (Version 89.01) was used for the analysis of genomic environments of vertebrate Sftpb genes. As we seen in phylogenetic tree in Fig. 11, Atlantic cod only contains a single Sftpb protein, while some fish species contain two Sftpb isoforms on different chromosomes, such as stickleback, medaka, tetraodon, platyfish, and Amazon molly. Alternative isoform can provide information of phylogenetic characters independent of the environment of genome neighborhoods. The Sftpb existed in Stickleback Group XIII, Tetraodon chromosome 12, Amazon molly scaffold KI519905.1, Platyfish scaffold JH556662.1, and Medaka chromosome 9, were grouped genes with Atlantic cod Sftpb. Atlantic cod Sftpb has a total of 28

neighbor genes on scaffold 2877, including *ewsr1b*, *GAS2L1*, *ap1b1*, *rasl10a*, *si:ch211-106a19.1*, *si:ch211-102c2.7*, *tcn2*, *dguok*, *ndr1*, *eif4epb1*, *ank1a*, *ebf2*, *mpx*, *usp39*, *si:ch211-283h6.4*, *mat2ab*, *inpp5jb*, *selm*, *SMTN*, *trafd1*, *mthfd2*, *arl6ip4*, *PITPNM2*, *si:ch211-275j6.5*, *setd8b*, *rilpl2*, *RILPL1*, and *hip1rb*. The direct genomic environment of the *sftpb* on these chromosomes were also similar and shared 23- 26 neighbor genes with Atlantic cod scaffold 2877 (Fig. 18). However in the phylogenetic tree, zebrafish chromosome 10, Cave fish scaffold KB871593.1, Stickleback Group XIV, Tetraodon chromosome 4, Amazon molly scaffold KI519616.1, Platyfish scaffold JH557076.1, and Medaka chromosome 12 tend to place Sftpb as outgroup to Atlantic cod. The direct genomic environment of the *sftpb* gene on these chromosomes were similar but contribute only 4- 5 neighbor genes to conserved synteny with Atlantic cod *sftpb*, including *mpx*, *mat2aa*, *inpp5ja*, *smtna*, *osbp2*, *ewsr1a*. Both zebrafish *sftpba and sftpbb* (on chromosome 5 and 10 respectively) seems to have relatively distant genomic environment to Atlantic cod s*ftpb*, and only 7 and 6 common genes of *sftpb* on zebrafish chromosome 5 and 10, respectively, were contributed to conserved syteny with cod *Sftpb* on scaffold 2877. It was expected that spotted gar and coelacanth contributed direct conserved genomic environment with Atlantic cod *sftpb*. In the genomic environment of Atlantic cod *sftpb*, genes of *ank1b*, *ebf2*, *mps*, *usp39*, and *mat2ab* were commonly shared with spotted gar, genes of *USP39* and *C2orf68* were commonly shared with coelacanth.

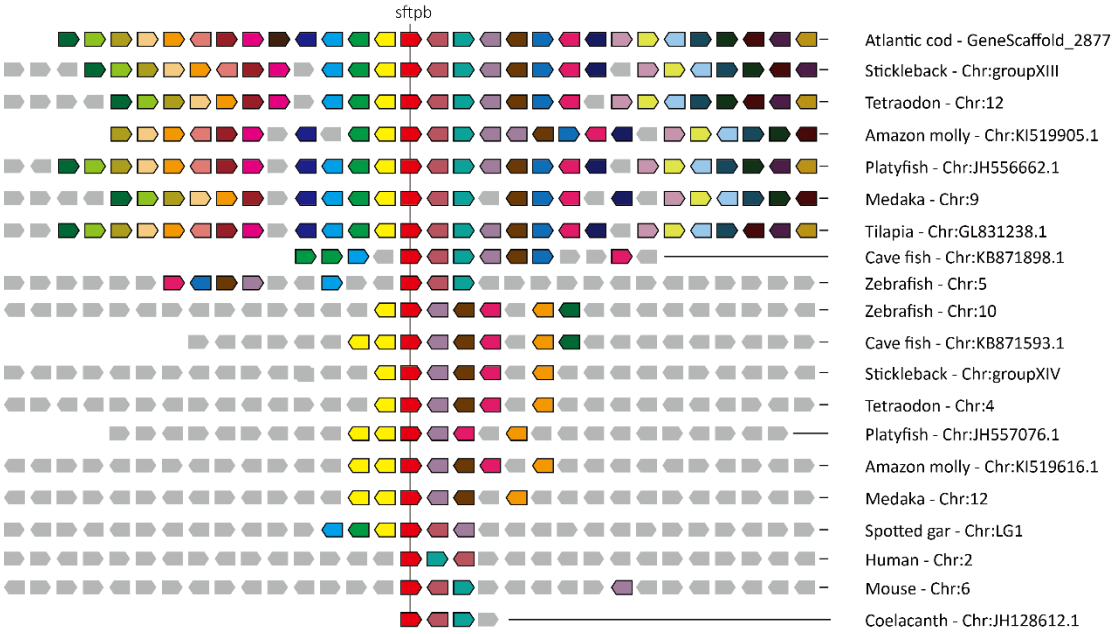


Fig. 18. Genomic environments of vertebrate *sftpb* genes. Orthologous genes contributing to conserved synteny are coded in similarly color. Gray boxes represent genes that are not

related.

As intragenomic patterns of conserved synteny can be informative to infer teleost ancestor and gene family history with regard to the whole genome duplications (Lagman *et al.* 2013), we therefore constructed multiple genome alignments, with the help of MAUVE (http://darlinglab.org/mauve/mauve.html), and the genomic database in Ensembl, to generate the orthologous pairwise between Atlantic cod *sftpb* and Stickleback *sftpb*, and the *sftpb* paralogons in different chromosomes of stickleback (Fig. 19). Although only partial genome of Atlantic cod in chromosome was available (Scaffold 2877 in Fig. 19), the genes that were available in Atlantic cod genome seemed to also cluster in Stickleback group XIII. Within Stickleback group XIII and XIV, the *sftpb* gene regions show extensive pairwise paralogy with each other, indicating that the *sftpb* paralogons were occurred in a large-scale duplication event.

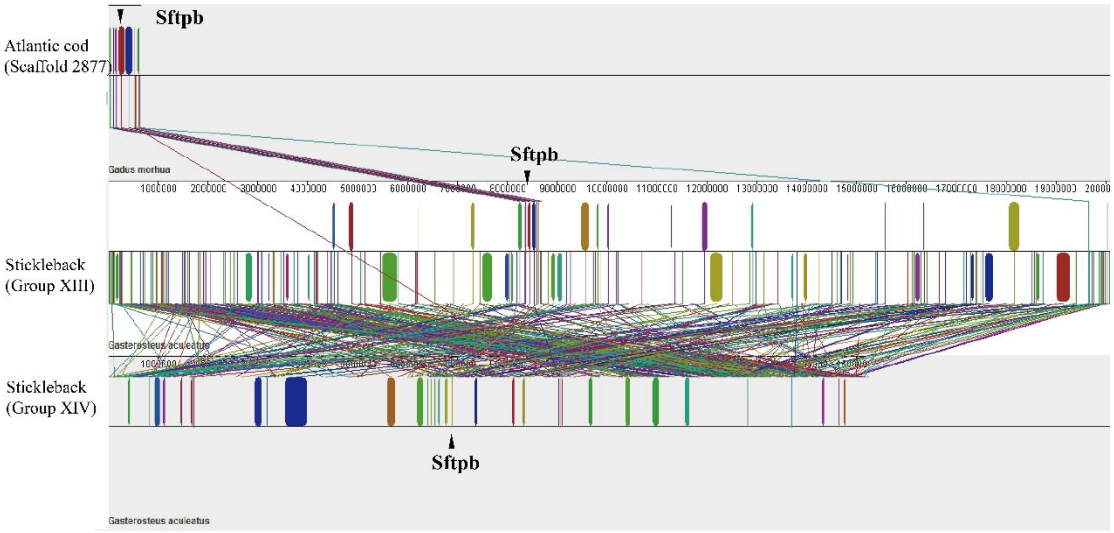


Fig. 19. Conserved synteny between Atlantic cod and Stickleback Sftpb gene. Orthologous pairwise clusters between Atlantic cod and Stickleback and conserved synteny of Sftpb paralogons in Stickleback Group XIII and XIV chromosomes are shown in the figure.

4. DISCUSSION

In the present study, qRT- PCR was utilized to investigate the expression level *sftpb, psap, elovl1*, and *pbx1* genes in Atlantic cod. Multiple genetic tools, such as BLAST, ClustalW, Muscle, Mauve, and Genomicus were utilized to explore the properties of these genes. The greatest challenge during the investigation into the genetic properties of these genes is the fact that most of these genes, such as *sftpb* and *psap,* have not been well characterized in Atlantic cod or in fish species. In this chapter, some important or interesting findings will be summarized, and more emphasis will be given to *sftpb* and *psap* because of their special relation and interesting properties.

4.1 Genetic analysis on Atlantic cod *sftpb* and *psap*

Although saposin is homologous with prosaposin and surfactant B, the expression modes was largely different during early stage of Atlantic cod. Significant decline of *sftpb* was observed before onset gastrula, while decline of *psap* was observed before early somitogenesis. Gene *psap* remain abundant since late gastrula till the end of the larval stage, while *sftpb* remain at low level between onset gastrula and hatching, and remain at higher level during larval period.

The expression modes of *sftpb* were similar between male and female Atlantic cod, while *sftpb* was much enriched in male reproductive organ than in female. This finding indicated that surfactant B may play an role similar with Psap, in maintaining male reproductive system (Guo et al. 2007). Another evidence which may prove the hypothesis was inferred in the phylogenetic tree in Fig. 11. Fish Sftpb and Psap protein were closely clustered, indicating a close relationship between Atlantic cod *sftpb* and *psap*, both of which are from superfamily of Saposin-like protein. Except for testis, Sftpb was mostly enriched in heart, spleen, head kidney, and gas gland. However in human and/ or mouse, SFTPB was mostly enriched in lung (Song *et al.* 2013). SFTPB found in mammalian regarding to regulate the surfactant surface activity in lung and promote the surfactant to spread over the surface (Hawgood *et al.* 1998; Johansson 1998). The finding of Sftpb in swim bladder or gas gland in fish species support the great possibility that the teleost fish Sftpb is homologous with tetrapods Sftpb. Atlantic cod *sftpb* consists of 9 coding exons, which is distinct from either zebrafish *sftpb* or human *SFTPB.* Both Saposin A- type and B- type domains, contained also in Psap, were predicted in Atlantic cod, zebrafish, and human

surfactant B proteins (Fig. 10). However, either zebrafish or Atlantic cod contains only two Saposin B- type and one Saposin A- type domain on C- terminus, while one Saposin A- type and three Saposin B- type domains was predicted on C- terminus in human SFTPB.

As surfactant B and prosaposin belong to saposin-like family, fish Sftpb and Psap protein were closely clustered as expected in our research (Fig. 11). However, just like Atlantic cod Psap and Elovl1, Sftpb has a distant evolutionary relation with Coelacanth and Whale shark Sftpb. Interestingly, not only has Whale shark Sftpb and Psap a close distance, they also seem to stand in the middle of fish Sftpb and Psap proteins. The findings in Whale shark Sftpb and Psap may indicate the interim differentiation process between Sftpb and Psap in the long- term evolution history. Besides, highly conserved saposin domains among Atlantic cod, zebrafish, and human prosaposin indicated conserved evolution on *psap* gene (Fig. 7). However, Elephantfish Psap, a member from Chondrichthyes, seems to has a closer evolutionary distant with Atlantic cod Psap than spotted gar Psap (Fig. 8).

In some fish species, *sftpb* exists in two different chromosomes, while in Atlantic cod only a single *sftpb* was predicted. According to the molecular phylogenetic tree and genomic (Fig. 11, 18), Atlantic cod *sftpb* seems to be more conserved with the Sftpb existed in Stickleback Group XIII, Tetraodon chromosome 12, Amazon molly scaffold KI519905.1, Platyfish scaffold JH556662.1, and Medaka chromosome 9, than the Sftpb isoforms existed in another chromosomes of these species. It indicates that the duplication event of *sftpb* among chromosomes is common in fish species. Multiple genomic alignments of *sftpb* showed highly conserved orthologous pairwise between Atlantic cod scaffold 2788 and Stickleback group XIII, and extensive pairwise between the *sftpb* paralogons in Stickleback group XIII and XIV (Fig. 19). The genomic information of the *sftpb* paralogons provided in stickleback may indicate a large-scale duplication event in the history of evolution.

The identification of elefant shark *sftpb* in Ensembl.fugu (Gene ID: SINCAMG00000002317), together with the whale shark *sftpb*, indicate *sftpb* also present in fish species without swim bladder. It indicates that surfactant B may not just function as the regulator of the respiration by maintaining the air sacs filling with air, like what it does in human or mouse lungs (Kebaabetswe *et al.* 2015; Wang *et al.* 2016), it may affect other organs in fish by activating the surfactant lipids on the surface as well.

4.2 Genetic analysis on Atlantic cod *elovl1*

In our investigation, Atlantic cod *elovl1* was encoded by two isoforms, *elovl1a* and *elovl1b*. Similar to the expression mode of Atlantic cod *psap* in our study, relative expression of Atlantic cod *elovl1* declined to a nadir at early somitogenesis stage, while it was stable and remained at relative low level at the larval stage. During early development of Atlantic cod, it is highly possible that *elovl1* genes are mostly enriched before early somitogenesis. The evidence support this hypothesis was reported in zebrafish (Bhandari *et al.* 2016), it was conceivable that zebrafish *elovl1* expressed before 24 hpf affected the later organogenesis or development of swim bladder, similar process may also exist in Atlantic cod embryo *elovl1*, *psap*, and *sftpb genes*.

Atlantic cod *elovl1* shows similar genetic structures with zebrafish, human, and mouse *elovl1* genes (Fig. 12). Multiple alignment of Elovl1 proteins also revealed highly conserved transmembrane domains from GNS1 family proteins in these species (Fig. 13). GNS1 family proteins are involved in produce C26 precursors for the synthesis of ceramide and sphingolipid (Oh *et al.* 1997). It indicates the important roles Atlantic cod Elovl1 played on the epithelial development of some organs, such as swim bladder. Phylogenetic analysis indicated that Atlantic cod Elovl1 shows a closer evolutionary distance with Actinopterygii, while Coelacanth, Whale shark, and Australian ghostshark Elovl1 seemed to distant from Atlantic cod Elovl1. (Fig. 14).

4.3 Genetic analysis on Atlantic cod *pbx1*

At embryonic stages of zebrafish, *pbx1* could be detected in the swim bladder as early as 28 hpf (Teoh *et al.* 2010). The absence of *pbx1* deactivated the inflation of swim bladder, with fetal consequence at 8 days post fertilization (dpf).

In our research, Atlantic cod *pbx1* are encoded by *pbx1a* and *pbx1b*, of which the relative expression exhibited distinct patterns during early period. During the embryonic period, relative expression of *pbx1a* declined dramatically to a nadir at late gastrula stage and increased significantly until hatching, while *pbx1a* remained high during larval period. However, *pbx1b* seems to remain stable through the early developmental stages. Different expressed patterns may indicate distinct roles of these two isoforms for organogenesis.

In zebrafish, Atlantic cod *pbx1*, and human *PBX1* genes, highly conserved PBC family protein, homeobox domain, and Helix- turn- helix motif domain were

predicted (Fig. 16). PBC domain belongs to homeodomain proteins which has been involved in regulating the expression of targeted genes and direct the formation of many body structures during vertebrate embryogenesis (Burglin *et al.* 1992). *Pbx1* has shown great activity in the lung, and the lack of *Pbx1* in lung mesenchyme during mouse embryos led to lethal consequence, with failure of postnatal alveolar expansion (Li et al. 2014). Tetrapod lung and teleost swim bladder are homologous organs derived from the foregut during embryonic period (Shannon et al. 2004). Hence, the roles *pbx1* plays in the development of fish swim bladder might be similar with the roles *Pbx1* plays in mouse lung might be similar.

5. CONCLUTION

In present study, the expression levels of *sftpb*, *psap*, and *elovl1* give detailed clues of the expression modes during the early development and in different organs of Atlantic cod, and the genetic analysis of structures, multiple alignment, phylogenetic tree, and conserved synteny of these genes contributes to a deeper understanding of their natures. These findings provide clues to their performances in fish swim bladder. The further work should be focused on the verification and function of those key genes in swim bladders, compared with their functions in lungs, and to find out the differences, similarities and homologous evidences.

REFERENCE

Agaba, M. K., Tocher D. R., Zheng X., Dickson C. A., Dick J. R. and Teale A. J. (2005). Cloning and functional characterisation of polyunsaturated fatty acid elongases of marine and freshwater teleost fish. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *142*(3), 342-352.

Agbaga, M.-P., Mandal M. N. A. and Anderson R. E. (2010). Retinal very long-chain PUFAs: new insights from studies on ELOVL4 protein. *Journal of lipid research*, *51*(7), 1624-1642.

Ahn, V. E., Faull K. F., Whitelegge J. P., Fluharty A. L. and Privé G. G. (2003). Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proceedings of the National Academy of Sciences*, *100*(1), 38-43.

Ahn, V. E., Leyko P., Alattia J. R., Chen L. and Privé G. G. (2006). Crystal structures of saposins A and C. *Protein science*, *15*(8), 1849-1857.

Andersson, M., Curstedt T., Jörnvall H. and Johansson J. (1995). An amphipathic helical motif common to tumourolytic polypeptide NK‐lysin and pulmonary surfactant polypeptide SP‐B. *FEBS letters*, *362*(3), 328-332.

Asadi, A., Jörgensen J. and Jacobsson A. (2002). Elovl1 and p55Cdc genes are localized in a tail-to-tail array and are co-expressed in proliferating cells. *Journal of Biological Chemistry*, *277*(21), 18494-18500.

Avery, T. and Brown J. (2005). Investigating the relationship among abnormal patterns of cell cleavage, egg mortality and early larval condition in Limanda ferruginea. *Journal of Fish Biology*, *67*(3), 890-896.

Avery, T. S., Killen S. S. and Hollinger T. R. (2009). The relationship of embryonic development, mortality, hatching success, and larval quality to normal or abnormal early embryonic cleavage in Atlantic cod, Gadus morhua. *Aquaculture*, *289*(3), 265-273.

Azuma, N., Hee-Chan S., Øystein L., Qiang F., Gould M. R., Hiraiwa M., . . . O'Brien J. S. (1998). Cloning, expression and map assignment of chicken prosaposin. *Biochemical Journal*, *330*(1), 321-327.

Bar-Am, I., Avivi L. and Horowitz M. (1996). Assignment of the human prosaposin gene (PSAP) to 10q22. 1 by fluorescence in situ hybridization. *Cytogenetic and Genome Research*, *72*(4), 316-318 (1996.

Benedito-Palos, L., Ballester-Lozano G. and Pérez-Sánchez J. (2014). Wide-gene expression analysis of lipid-relevant genes in nutritionally challenged gilthead sea bream (Sparus aurata). *Gene*, *547*(1), 34-42.

Berthelsen, J., Zappavigna V., Ferretti E., Mavilio F. and Blasi F. (1998). The novel homeoprotein Prep1 modulates Pbx–Hox protein cooperativity. *The EMBO journal*, *17*(5), 1434-1445.

Bhandari, S., Lee J. N., Kim Y.-I., Nam I.-K., Kim S.-J., Kim S.-J., . . . Yoo H. J. (2016). The fatty acid chain elongase, Elovl1, is required for kidney and swim bladder development during zebrafish embryogenesis. *Organogenesis*, *12*(2), 78-93.

Biscotti, M. A., Gerdol M., Canapa A., Forconi M., Olmo E., Pallavicini A., . . . Schartl M. (2016). The lungfish transcriptome: a glimpse into molecular evolution events at the transition from water to land. *Scientific reports*, *6*.

Burglin, T. and Ruvkun G. (1992). New motif in PBX genes. *Nature genetics*, *1*, 319-320.

Bürglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC,

KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic acids research*, *25*(21), 4173-4180.
Bürglin, T. R. and Affolter M. (2016). Homeodomain proteins: an update. *Chromosoma*, *125*(3), 497-521.
Campana, W. M., Hiraiwa M. and O'Brien J. S. (1998). Prosaptide activates the MAPK pathway by a G-protein-dependent mechanism essential for enhanced sulfatide synthesis by Schwann cells. *The FASEB Journal*, *12*(3), 307-314.
Chan, K. K. K., Zhang J., Chia N. Y., Chan Y. S., Sim H. S., Tan K. S., . . . Choo A. B. H. (2009). KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. *Stem Cells*, *27*(9), 2114-2125.
Chang-Yi, C., Kusuda S., Seguchi T., Takahashi M., Aisu K. and Tezuka T. (1997). Decreased level of prosaposin in atopic skin. *Journal of investigative dermatology*, *109*(3), 319-323.
Chang, C.-P., Jacobs Y., Nakamura T., Jenkins N. A., Copeland N. G. and Cleary M. L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Molecular and cellular biology*, *17*(10), 5679-5687.
Clark, J. C., Wert S. E., Bachurski C. J., Stahlman M. T., Stripp B. R., Weaver T. E. and Whitsett J. A. (1995). Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. *Proceedings of the National Academy of Sciences*, *92*(17), 7794-7798.
Crouch, E. C. (1998). Structure, biologic properties, and expression of surfactant protein D (SP-D). *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1408*(2), 278-289.
Daniels, C. B., Orgeig S., Sullivan L. C., Ling N., Bennett M. B., Schürch S., . . . Brauner C. J. (2004). The origin and evolution of the surfactant system in fish: insights into the evolution of lungs and swim bladders. *Physiological and Biochemical Zoology*, *77*(5), 732-749.
Daniels, C. B. and Skinner C. H. (1994). The composition and function of surface-active lipids in the goldfish swim bladder. *Physiological Zoology*, *67*(5), 1230-1256.
Doering, T., Holleran W. M., Potratz A., Vielhaber G., Elias P. M., Suzuki K. and Sandhoff K. (1999). Sphingolipid activator proteins are required for epidermal permeability barrier formation. *Journal of Biological Chemistry*, *274*(16), 11038-11045.
Ferter, K., Weltersbach M. S., Humborstad O.-B., Fjelldal P. G., Sambraus F., Strehlow H. V. and Vølstad J. H. (2015). Dive to survive: effects of capture depth on barotrauma and post-release survival of Atlantic cod (Gadus morhua) in recreational fisheries. *ICES Journal of Marine Science: Journal du Conseil*, *72*(8), 2467-2481.
Floros, J. and Kala P. (1998). Surfactant proteins: molecular genetics of neonatal pulmonary diseases. *Annual review of physiology*, *60*(1), 365-384.
Floros, J. and Wang G. (2001). A point of view: quantitative and qualitative imbalance in disease pathogenesis; pulmonary surfactant protein A genetic variants as a model. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *129*(1), 295-303.
Fullagar, W. K., Aberdeen K. A., Bucknall D. G., Kroon P. A. and Gentle I. R. (2003). Conformational changes in SP-B as a function of surface pressure. *Biophysical journal*, *85*(4), 2624-2632.
Galloway, T., Kjørsvik E. and Kryvi H. (1998). Effect of temperature on viability and

axial muscle development in embryos and yolk sac larvae of the Northeast Arctic cod (Gadus morhua). *Marine Biology*, *132*(4), 559-567.
Ghioni, C., Tocher D. R., Bell M. V., Dick J. R. and Sargent J. R. (1999). Low C 18 to C 20 fatty acid elongase activity and limited conversion of stearidonic acid, 18: 4 (n–3), to eicosapentaenoic acid, 20: 5 (n–3), in a cell line from the turbot, Scophthalmus maximus. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1437*(2), 170-181.
Gorodilov, Y. N., Terjesen B. F., Krasnov A. and Takle H. (2008). Description of embryogenesis of Atlantic cod Gadus morhua. *Open Mar Biol J*, *2*, 43-53.
Guillou, H., Zadravec D., Martin P. G. and Jacobsson A. (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Progress in lipid research*, *49*(2), 186-199.
Guo, F., Huang X., Li S., Sun L., Li Y., Li H., . . . Zhou T. (2007). Identification of prosaposin as a novel interaction partner for Rhox5. *Journal of Genetics and Genomics*, *34*(5), 392-399.
Hall, T. E., Smith P. and Johnston I. A. (2004). Stages of embryonic development in the Atlantic cod Gadus morhua. *Journal of morphology*, *259*(3), 255-270.
Harden Jones, F. and Scholes P. (1985). Gas secretion and resorption in the swimbladder of the codGadus morhua. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, *155*(3), 319-331.
Hawgood, S., Derrick M. and Poulain F. (1998). Structure and properties of surfactant protein B. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1408*(2), 150-160.
Heffernan, O., Righton D. and Michalsen K. (2004). Use of data storage tags to quantify vertical movements of cod: effects on acoustic measures. *ICES Journal of Marine Science*, *61*(7), 1062-1070.
Herbing, I. (2001). Development of feeding structures in larval fish with different life histories: winter flounder and Atlantic cod. *Journal of Fish Biology*, *59*(4), 767-782.
Herbing, I. H. V., Miyake T., Hall B. K. and Boutilier R. (1996a). Ontogeny of feeding and respiration in larval Atlantic cod Gadus morhua (Teleostei, Gadiformes): II. Function. *Journal of morphology*, *227*, 37-50.
Herbing, I. H. V., Miyake T., Hall B. K. and Boutilier R. G. (1996b). Ontogeny of feeding and respiration in larval Atlantic cod Gadus morhua (Teleostei, Gadiformes): I. Morphology. *Journal of morphology*, *227*(1), 15-35.
Hineno, T., Sano A., Kondoh K., Ueno S.-i., Kakimoto Y. and Yoshida K.-i. (1991). Secretion of sphingolipid hydrolase activator precursor, prosaposin. *Biochemical and biophysical research communications*, *176*(2), 668-674.
Hiraiwa, M., Martin B. M., Kishimoto Y., Conner G. E., Tsuji S. and O'Brien J. S. (1997). Lysosomal proteolysis of prosaposin, the precursor of saposins (sphingolipid activator proteins): its mechanism and inhibition by ganglioside. *Archives of biochemistry and biophysics*, *341*(1), 17-24.
Hiraiwa, M., Soeda S., Kishimoto Y. and O'Brien J. S. (1992). Binding and transport of gangliosides by prosaposin. *Proceedings of the National Academy of Sciences*, *89*(23), 11254-11258.
Holstein, A.-F., Schulze W. and Davidoff M. (2003). Understanding spermatogenesis is a prerequisite for treatment. *Reproductive Biology and Endocrinology*, *1*(1), 107.
Horowitz, M. and Zimran A. (1994). Mutations causing Gaucher disease. *Human mutation*, *3*(1), 1-11.

Howell, B. (1984). The intensive rearing of juvenile cod, Gadus morhua L. *Flødevigen Rapportserie*, *1*, 657-675.
Hunter, J. R. (1976). Report of a colloquium on larval fish mortality studies and their relation to fishery research, January 1975, Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.
Igdoura, S. A., Rasky A. and Morales C. R. (1996). Trafficking of sulfated glycoprotein-1 (prosaposin) to lysosomes or to the extracellular space in rat Sertoli cells. *Cell and tissue research*, *283*(3), 385-394.
Iversen, S. and Danielssen D. (1984). Development and mortality of cod(Gadus morhua L.) eggs and larvae in different temperatures. *Floedevigen Rapportserie. 1984*.
Jakobsson, A., Westerberg R. and Jacobsson A. (2006). Fatty acid elongases in mammals: their regulation and roles in metabolism. *Progress in lipid research*, *45*(3), 237-249.
Johansson, J. (1998). Structure and properties of surfactant protein C. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1408*(2), 161-172.
Johansson, J., Curstedt T. and Joernvall H. (1991). Surfactant protein B: disulfide bridges, structural properties and kringle similarities. *Biochemistry*, *30*(28), 6917-6921.
Juliussen, E. H. (2016). Thermal reaction norms for larval Atlantic cod (Gadus morhua)-Exploring population differences on a micro-geographic scale.
Kamps, M. P., Murre C., Sun X.-h. and Baltimore D. (1990). A new homeobox gene contributes the DNA binding domain of the t (1; 19) translocation protein in pre-B ALL. *Cell*, *60*(4), 547-555.
Kebaabetswe, L. P., Haick A. K., Gritsenko M. A., Fillmore T. L., Chu R. K., Purvine S. O., . . . Waters K. M. (2015). Proteomic analysis reveals down-regulation of surfactant protein B in murine type II pneumocytes infected with influenza A virus. *Virology*, *483*, 96-107.
Khatami, M. H., Saika-Voivod I. and Booth V. (2016). All-atom molecular dynamics simulations of lung surfactant protein B: Structural features of SP-B promote lipid reorganization. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1858*(12), 3082-3092.
Kilstrup‐Nielsen, C., Alessio M. and Zappavigna V. (2003). PBX1 nuclear export is regulated independently of PBX‐MEINOX interaction by PKA phosphorylation of the PBC‐B domain. *The EMBO journal*, *22*(1), 89-99.
Kim, S. K., Selleri L., Lee J. S., Zhang A. Y., Gu X., Jacobs Y. and Cleary M. L. (2002). Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. *Nature genetics*, *30*(4), 430-435.
Kishimoto, Y., Hiraiwa M. and O'brien J. (1992). Saposins: structure, function, distribution, and molecular genetics. *Journal of lipid research*, *33*(9), 1255-1267.
Kishore, U., Greenhough T. J., Waters P., Shrive A. K., Ghai R., Kamran M. F., . . . Chakraborty T. (2006). Surfactant proteins SP-A and SP-D: structure, function and receptors. *Molecular immunology*, *43*(9), 1293-1315.
Kolder, I., Van der Plas-Duivesteijn S., Tan G., Wiegertjes G., Forlenza M., Guler A., . . . Irnazarow I. (2016). A full-body transcriptome and proteome resource for the European common carp. *BMC Genomics*, *17*(1), 701.
Kolter, T. and Sandhoff K. (2005). Principles of lysosomal membrane digestion: stimulation of sphingolipid degradation by sphingolipid activator proteins and anionic lysosomal lipids. *Annu. Rev. Cell Dev. Biol.*, *21*, 81-103.

Kolter, T., Winau F., Schaible U. E., Leippe M. and Sandhoff K. (2005). Lipid-binding proteins in membrane digestion, antigen presentation, and antimicrobial defense. *Journal of Biological Chemistry*, *280*(50), 41125-41128.

Koochekpour, S., Zhuang Y. J., Beroukhim R., Hsieh C. L., Hofer M. D., Zhau H. E., . . . Luftig R. B. (2005). Amplification and overexpression of prosaposin in prostate cancer. *Genes, Chromosomes and Cancer*, *44*(4), 351-364.

Lagman, D., Daza D. O., Widmark J., Abalo X. M., Sundström G. and Larhammar D. (2013). The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC evolutionary biology*, *13*(1), 238.

Laurent, A., Bihan R., Omilli F., Deschamps S. and Pellerin I. (2003). PBX proteins: much more than Hox cofactors. *International Journal of Developmental Biology*, *52*(1), 9-20.

Leonova, T., Qi X., Bencosme A., Ponce E., Sun Y. and Grabowski G. A. (1996). Proteolytic processing patterns of prosaposin in insect and mammalian cells. *Journal of Biological Chemistry*, *271*(29), 17312-17320.

Li, W., Lin C. Y., Shang C., Han P., Xiong Y., Lin C. J., . . . Chang C. P. (2014). Pbx1 activates Fgf10 in the mesenchyme of developing lungs. *genesis*, *52*(5), 399-407.

Liu, J., Abdel-Razek O., Liu Z., Hu F., Zhou Q., Cooney R. N. and Wang G. (2015). Role of surfactant proteins A and D in sepsis-induced acute kidney injury. *Shock (Augusta, Ga.)*, *43*(1), 31.

Madar-Sharipo, L., Pasmanik-Chor M., Vaccaro A. M., Dinur T., Dagan A., Shimon G. and Horowitz M. (1999). Importance of splicing for prosaposin sorting. *Biochemical Journal*, *337*(3), 433-443.

Manley, N. R., Selleri L., Brendolan A., Gordon J. and Cleary M. L. (2004). Abnormalities of caudal pharyngeal pouch development in Pbx1 knockout mice mimic loss of Hox3 paralogs. *Developmental biology*, *276*(2), 301-312.

McCormack, F. X. (1998). Structure, processing and properties of surfactant protein A. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1408*(2), 109-131.

Meuer, S., Wittwer C. and Nakagawara K.-I. (2012). Rapid cycle real-time PCR: methods and applications, Springer Science & Business Media.

Midling, K. Ø., Koren C., Humborstad O.-B. and Sæther B.-S. (2012). Swimbladder healing in Atlantic cod (Gadus morhua), after decompression and rupture in capture-based aquaculture. *Marine Biology Research*, *8*(4), 373-379.

Moens, C. B. and Selleri L. (2006). Hox cofactors in vertebrate development. *Developmental biology*, *291*(2), 193-206.

Monica, K., Galili N., Nourse J., Saltman D. and Cleary M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Molecular and cellular biology*, *11*(12), 6149-6157.

Monroig, Ó., Webb K., Ibarra-Castro L., Holt G. J. and Tocher D. R. (2011). Biosynthesis of long-chain polyunsaturated fatty acids in marine fish: Characterization of an Elovl4-like elongase from cobia Rachycentron canadum and activation of the pathway during early life stages. *Aquaculture*, *312*(1), 145-153.

Morais, S., Monroig O., Zheng X., Leaver M. J. and Tocher D. R. (2009). Highly unsaturated fatty acid synthesis in Atlantic salmon: characterization of

ELOVL5-and ELOVL2-like elongases. *Marine Biotechnology*, *11*(5), 627-639.

Morales, C., Zhao Q., Lefrancois S. and Ham D. (2000). Role of prosaposin in the male reproductive system: effect of prosaposin inactivation on the testis, epididymis, prostate, and seminal vesicles. *Archives of andrology*, *44*(3), 173-186.

Morales, C. R., Zhao Q., El-Alfy M. and Suzuki K. (2000). Targeted disruption of the mouse prosaposin gene affects the development of the prostate gland and other male reproductive organs. *J Androl*, *21*(6), 765-775.

Nourse, J., Mellentin J. D., Galili N., Wilkinson J., Stanbridge E., Smith S. D. and Cleary M. L. (1990). Chromosomal translocation t (1; 19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, *60*(4), 535-545.

O'Brien, J. S. and Kishimoto Y. (1991). Saposin proteins: structure, function, and role in human lysosomal storage disorders. *The FASEB Journal*, *5*(3), 301-308.

O'Brien, J. S., Kretz K. A., Dewji N., Wenger D. A., Esch F. and Fluharty A. L. (1988). Coding of two sphingolipid activator proteins (SAP-1 and SAP-2) by same genetic locus. *Science*, *241*(4869), 1098.

Ofman, R., Dijkstra I. M., van Roermund C. W., Burger N., Turkenburg M., van Cruchten A., . . . Kemp S. (2010). The role of ELOVL1 in very long‐chain fatty acid homeostasis and X‐linked adrenoleukodystrophy. *EMBO molecular medicine*, *2*(3), 90-97.

Oh, C.-S., Toke D. A., Mandala S. and Martin C. E. (1997). ELO2 and ELO3, Homologues of theSaccharomyces cerevisiae ELO1 Gene, Function in Fatty Acid Elongation and Are Required for Sphingolipid Formation. *Journal of Biological Chemistry*, *272*(28), 17376-17384.

Ohno, Y., Suto S., Yamanaka M., Mizutani Y., Mitsutake S., Igarashi Y., . . . Kihara A. (2010). ELOVL1 production of C24 acyl-CoAs is linked to C24 sphingolipid synthesis. *Proceedings of the National Academy of Sciences*, *107*(43), 18439-18444.

Orgeig, S. and Daniels C. B. (1995). The evolutionary significance of pulmonary surfactant in lungfish (Dipnoi). *American journal of respiratory cell and molecular biology*, *13*(2), 161-166.

Park, J. T., Shih I.-M. and Wang T.-L. (2008). Identification of Pbx1, a potential oncogene, as a Notch3 target gene in ovarian cancer. *Cancer research*, *68*(21), 8852-8860.

Patthy, L. (1991). Homology of the precursor of pulmonary surfactant-associated protein SP-B with prosaposin and sulfated glycoprotein 1. *Journal of Biological Chemistry*, *266*(10), 6035-6037.

Perry, S. F., Wilson R. J., Straus C., Harris M. B. and Remmers J. E. (2001). Which came first, the lung or the breath? *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *129*(1), 37-47.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic acids research*, *29*(9), e45-e45.

Pöpperl, H., Rikhof H., Cheng H., Haffter P., Kimmel C. B. and Moens C. B. (2000). Lazarus is a novel pbx gene that globally mediates hox gene function in zebrafish. *Molecular cell*, *6*(2), 255-267.

Prem, C., Salvenmoser W., Würtz J. and Pelster B. (2000). Swim bladder gas gland cells produce surfactant: in vivo and in culture. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *279*(6),

R2336-R2343.
Puvanendran, V., Falk‐Petersen I. B., Lysne H., Tveiten H., Toften H. and Peruzzi S. (2015). Effects of different step‐wise temperature increment regimes during egg incubation of Atlantic cod (Gadus morhua L.) on egg viability and newly hatched larval quality. *Aquaculture Research*, *46*(1), 226-235.
Rauskolb, C., Peifer M. and Wieschaus E. (1993). extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. *Cell*, *74*(6), 1101-1112.
Regis, S., Filocamo M., Corsolini F., Caroli F., Keulemans J. L., Van Diggelen O. P. and Gatti R. (1999). An Asn> Lys substitution in saposin B involving a conserved amino acidic residue and leading to the loss of the single N-glycosylation site in a patient with metachromatic leukodystrophy and normal arylsulphatase A activity. *European Journal of Human Genetics*, *7*(2), 125-130.
Rorman, E. G., Scheinker V. and Grabowski G. A. (1992). Structure and evolution of the human prosaposin chromosomal gene. *Genomics*, *13*(2), 312-318.
Rubio, S., Chailley-Heu B., Ducroc R. and Bourbon J. R. (1996). Antibody against pulmonary surfactant protein A recognizes proteins in intestine and swim bladder of the freshwater fish, carp. *Biochemical and biophysical research communications*, *225*(3), 901-906.
Sæle, Ø., Nordgreen A., Hamre K. and Olsvik P. A. (2009). Evaluation of candidate reference genes in Q-PCR studies of Atlantic cod (Gadus morhua) ontogeny, with emphasis on the gastrointestinal tract. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *152*(1), 94-101.
Sarker, M., Waring A. J., Walther F. J., Keough K. M. and Booth V. (2007). Structure of Mini-B, a Functional Fragment of Surfactant Protein B, in Detergent Micelles†. *Biochemistry*, *46*(39), 11047-11056.
Sassa, T., Ohno Y., Suzuki S., Nomura T., Nishioka C., Kashiwagi T., . . . Shimizu H. (2013). Impaired epidermal permeability barrier in mice lacking elovl1, the gene responsible for very-long-chain fatty acid production. *Molecular and cellular biology*, *33*(14), 2787-2796.
Schackmann, M. J., Ofman R., Dijkstra I. M., Wanders R. J. and Kemp S. (2015). Enzymatic characterization of ELOVL1, a key enzyme in very long-chain fatty acid synthesis. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1851*(2), 231-237.
Schnabel, C. A., Godin R. E. and Cleary M. L. (2003). Pbx1 regulates nephrogenesis and ureteric branching in the developing kidney. *Developmental biology*, *254*(2), 262-276.
Schnabel, C. A., Selleri L., Jacobs Y., Warnke R. and Cleary M. L. (2001). Expression of Pbx1b during mammalian organogenesis. *Mechanisms of development*, *100*(1), 131-135.
Selleri, L., Depew M. J., Jacobs Y., Chanda S. K., Tsang K. Y., Cheah K. S., . . . Cleary M. L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development*, *128*(18), 3543-3557.
Serrano, A. G., Cruz A., Rodríguez-Capote K., Possmayer F. and Pérez-Gil J. (2005). Intrinsic structural and functional determinants within the amino acid sequence of mature pulmonary surfactant protein SP-B. *Biochemistry*, *44*(1), 417-430.
Shanker, S., Hu Z. and Wilkinson M. (2008). Epigenetic regulation and downstream

targets of the Rhox5 homeobox gene. *International journal of andrology*, *31*(5), 462-470.
Shannon, J. M. and Hyatt B. A. (2004). Epithelial-mesenchymal interactions in the developing lung. *Annu. Rev. Physiol.*, *66*, 625-645.
Silveyra, P. and Floros J. (2012). Genetic variant associations of human SP-A and SP-D with acute and chronic lung injury. *Frontiers in bioscience: a journal and virtual library*, *17*, 407.
Siri, L., Rossi A., Lanza F., Mazzotti R., Costa A., Stroppiano M., . . . Filocamo M. (2014). A novel homozygous splicing mutation in PSAP gene causes metachromatic leukodystrophy in two Moroccan brothers. *neurogenetics*, *15*(2), 101-106.
Smits, A. W., Orgeig S. and Daniels C. B. (1994). Surfactant composition and function in lungs of air-breathing fishes. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *266*(4), R1309-R1313.
Soeda, S., Hiraiwa M., O'Brien J. S. and Kishimoto Y. (1993). Binding of cerebrosides and sulfatides to saposins AD. *Journal of Biological Chemistry*, *268*(25), 18519-18523.
Song, Y., Ahn J., Suh Y., Davis M. E. and Lee K. (2013). Identification of novel tissue-specific genes by analysis of microarray databases: a human and mouse model. *PLoS One*, *8*(5), e64483.
Spiegel, R., Bach G., Sury V., Mengistu G., Meidan B., Shalev S., . . . Zeigler M. (2005). A mutation in the saposin A coding region of the prosaposin gene in an infant presenting as Krabbe disease: first report of saposin A deficiency in humans. *Molecular genetics and metabolism*, *84*(2), 160-166.
Sprecher-Levy, H., Orr-Urtreger A., Lonai P. and Horowitz M. (1993). Murine prosaposin: expression in the reproductive system of a gene implicated in human genetic diseases. *Cellular and molecular biology (Noisy-le-Grand, France)*, *39*(3), 287-299.
Sullivan, L. C., Daniels C. B., Phillips I. D., Orgeig S. and Whitsett J. A. (1998). Conservation of surfactant protein A: evidence for a single origin for vertebrate pulmonary surfactant. *Journal of Molecular Evolution*, *46*(2), 131-138.
Sun, Y., Witte D. P. and Grabowski G. A. (1994). Developmental and tissue-specific expression of prosaposin mRNA in murine tissues. *The American journal of pathology*, *145*(6), 1390.
Swift, G. H., Liu Y., Rose S. D., Bischof L. J., Steelman S., Buchberg A. M., . . . MacDonald R. J. (1998). An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2). *Molecular and cellular biology*, *18*(9), 5109-5120.
Taranger, G. L., Carrillo M., Schulz R. W., Fontaine P., Zanuy S., Felip A., . . . Norberg B. (2010). Control of puberty in farmed fish. *General and comparative endocrinology*, *165*(3), 483-515.
Teoh, P. H., Shu‐Chien A. C. and Chan W. K. (2010). Pbx1 is essential for growth of zebrafish swim bladder. *Developmental Dynamics*, *239*(3), 865-874.
Terashita, T., Saito S., Miyawaki K., Hyodo M., Kobayashi N., Shimokawa T., . . . Gyo K. (2007). Localization of prosaposin in rat cochlea. *Neuroscience research*, *57*(3), 372-378.
Tocher, D. R., Zheng X., Schlechtriem C., Hastings N., Dick J. R. and Teale A. J. (2006). Highly unsaturated fatty acid synthesis in marine fish: cloning,

functional characterization, and nutritional regulation of fatty acyl Δ6 desaturase of Atlantic cod (Gadus morhua L.). *Lipids*, *41*(11), 1003-1016.
Tsuda, M., Sakiyama T., Endo H. and Kitagawa T. (1992). The primary structure of mouse saposin. *Biochemical and biophysical research communications*, *184*(3), 1266-1272.
Tvrdik, P., Westerberg R., Silve S., Asadi A., Jakobsson A., Cannon B., . . . Jacobsson A. (2000). Role of a new mammalian gene family in the biosynthesis of very long chain fatty acids and sphingolipids. *The Journal of cell biology*, *149*(3), 707-718.
Uhlig, S. and Gulbins E. (2008). Sphingolipids in the lungs. *American journal of respiratory and critical care medicine*, *178*(11), 1100-1114.
Van Den Berghe, L. c., Sainton K., Gogat K. n., Marchant D., Dufour E., Bonnel S., . . . Abitbol M. (2004). Prosaposin gene expression in normal and dystrophic RCS rat retina. *Investigative ophthalmology & visual science*, *45*(5), 1297-1305.
Vlachakis, N., Ellstrom D. R. and Sagerström C. G. (2000). A novel pbx family member expressed during early zebrafish embryogenesis forms trimeric complexes with Meis3 and Hoxb1b. *Developmental Dynamics*, *217*(1), 109-119.
Wagner, K., Mincheva A., Korn B., Lichter P. and Pöpperl H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mechanisms of development*, *103*(1), 127-131.
Walther, F. J., Waring A. J., Hernandez-Juviel J. M., Gordon L. M., Wang Z., Jung C.-L., . . . Sharma S. (2010). Critical structural and functional roles for the N-terminal insertion sequence in surfactant protein B analogs. *PLoS One*, *5*(1), e8672.
Walther, F. J., Waring A. J., Sherman M. A., Zasadzinski J. A. and Gordon L. M. (2007). Hydrophobic surfactant proteins and their analogues. *Neonatology*, *91*(4), 303-310.
Wang, J., Mei H., Liu C., Zhang Y., Liu C., Song D. and Zhang Y. (2016). Relationship between R236C site in exon 7 of SP-B gene and respiratory distress syndrome in Han newborns in western Inner Mongolia. *Zhongguo dang dai er ke za zhi= Chinese journal of contemporary pediatrics*, *18*(9), 802.
Wang, Y., Rao K. M. K. and Demchuk E. (2003). Topographical organization of the N-terminal segment of lung pulmonary surfactant protein B (SP-B1-25) in phospholipid bilayers. *Biochemistry*, *42*(14), 4015-4027.
Whitsett, J. A., Wert S. E. and Weaver T. E. (2010). Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annual review of medicine*, *61*, 105-119.
Wieser, W. (1991). Limitations of energy acquisition and energy use in small poikilotherms: evolutionary implications. *Functional Ecology*, *5*(2), 234-240.
Xue, X., Feng C. Y., Hixson S. M., Johnstone K., Anderson D. M., Parrish C. C. and Rise M. L. (2014). Characterization of the fatty acyl elongase (elovl) gene family, and hepatic elovl and delta-6 fatty acyl desaturase transcript expression and fatty acid responses to diets containing camelina oil in Atlantic cod (Gadus morhua). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *175*, 9-22.
Yeh, Y.-Y. (1988). Long chain fatty acid deficits in brain myelin sphingolipids of undernourished rat pups. *Lipids*, *23*(12), 1114-1118.

Yin, M. and Blaxter J. (1986). Morphological changes during growth and starvation of larval cod (Gadus morhua L.) and flounder (Platichthysflesus L.). *Journal of Experimental Marine Biology and Ecology*, *104*(1-3), 215-228.

Zaltash, S., Palmblad M., Curstedt T., Johansson J. and Persson B. (2000). Pulmonary surfactant protein B: a structural model and a functional analogue. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1466*(1), 179-186.

Zhao, Q., Bell A. W., El-Alfy M. and Morales C. R. (1997). Mouse testicular sulfated glycoprotein-1: sequence analysis of the common backbone structure of prosaposins. *Journal of andrology*, *19*(2), 165-174.

Zhao, Q., Hay N. and Morales C. R. (1997). Structural analysis of the mouse prosaposin (SGP‐1) gene reveals the presence of an exon that is alternatively spliced in transcribed mRNAs. *Molecular reproduction and development*, *48*(1), 1-8.

Zhao, Q. and Morales C. R. (2000). Identification of a novel sequence involved in lysosomal sorting of the sphingolipid activator protein prosaposin. *Journal of Biological Chemistry*, *275*(32), 24829-24839.

Zheng, W., Wang Z., Collins J. E., Andrews R. M., Stemple D. and Gong Z. (2011). Comparative transcriptome analyses indicate molecular homology of zebrafish swimbladder and mammalian lung. *PLoS One*, *6*(8), e24019.

Zhu, Y. and Conner G. E. (1994). Intermolecular association of lysosomal protein precursors during biosynthesis. *Journal of Biological Chemistry*, *269*(5), 3846-3851.

APPENDIX

Appendix 1: Accession NO. or transcript ID of different proteins

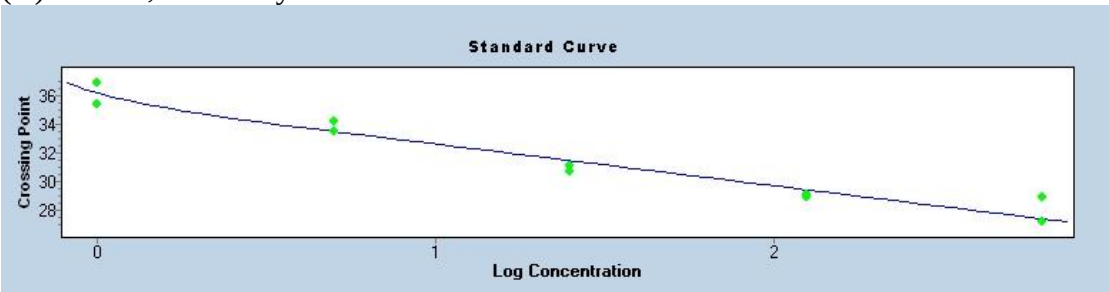
Protein type	Species	Accession NO. /Transcript ID
Elovl1	Atlantic cod (Elovl1a)	GAMO_00033084_RA
	Atlantic cod (Elovl1b)	GAMO_00029213_RA
	zebrafish (Elovl1a)	NP_001005989.1
	zebrafish (Elovl1b)	AAI55206.1
	Mummichog (Elovl1)	XP_021173338.1
	Ballan wrasse (Elovl1)	XP_020505695.1
	European bass (Elovl1)	AIT82973.1
	Spotted gar (Elovl1-predict)	XP_015210705.1
	Southern platyfish (Elovl1-predict)	XP_014328001.1
	Atlantic herring (Elovl1-predict)	XP_012677319.1
	Beira killifish (Elovl1b)	SBQ98112.1
	Whale shark (Elovl1)	P_020383509.1
	Australian ghostshark (Elovl1-predict)	XP_007885271.1
	Coelacanth (Elovl1-predict)	XP_014352157.1
Pbx	Atlantic cod (Pbx1a)	ENSGMOT00000011191.1
	Atlantic cod (Pbx1b)	ENSGMOT00000008396.1
	zebrafish (Pbx1a)	NP_571689.1
	zebrafish (Pbx1b)	NP_001077322.2
	Blackstripe livebearer (Pbx1)	JAO75987.1
	Ballan wrasse (Pbx1)	XP_020487568.1
	Beira Killifish (Pbx1)	SBQ86523.1
	Zebra mbuna (Pbx1-predict)	XP_004552971.1
	Tetraodon (Pbx1a)	ENSTNIT00000009492.1
	Platyfish (Pbx1a)	ENSXMAT00000016342.1
	Platyfish (Pbx1b)	ENSXMAT00000003079.1
	Fugu (Pbx1a)	ENSTRUT00000017794.1
	Tilapia (Pbx1a)	ENSONIT00000012317.1
	Tilapia (Pbx1b)	ENSONIT00000000322.1
	Coelacanth (Pbx1)	ENSLACT00000014334.1
Psap	Atlantic cod (Psap)	GAMO_00033032_RA
	zebrafish (Psap)	ENSDART00000056295.7
	Amazon molly (Psap)	ENSPFOT00000024381.1
	Cave fish (Psap)	ENSAMXT00000015466.1
	Coelacanth (Psap)	ENSLACT00000014532.1
	Fugu (Psap)	ENSTRUT00000028542.1
	Platyfish (Psap)	ENSXMAT00000005633.1
	Spotted gar (Psap)	ENSLOCT00000010545.1
	Stickleback (Psap)	ENSGACT00000006810.1
	Tilapia (Psap)	ENSONIT00000012479.1
	Turquoise killifish (Psap)	SBP62631.1
	Northern pike (Psap-predict)	XP_019902042.1
	Elephantfish (Psap)	AHI42569.1
	Large yellow croaker (Psap-predict)	XP_010728542.2
	Atlantic herring (Psap-predict)	XP_012669904.1
	Asian sea bass (Psap-predict)	XP_018515599.1
	Whale Shark (Psap-partial)	XP_020390811.1
	Australian ghostshark (Psap-predict)	XP_007902815.1
Sftpb	Atlantic cod (Sftpb)	GAMO_00041380_RA

Black rockcod (Sftpb)	XP_010765022.1
Coho salmon (prosaposin-like)	XP_020328019.1
Ballan wrasse (prosaposin-like)	XP_020506357.1
Mudskipper (Sftpb)	XP_020794066.1
Swamp eel (Sftpb)	XP_020458197.1
Whale shark (Sftpb)	XP_020378026.1
Mangrove killifish (Sftpb-predict)	XP_017269442.1
Spotted gar (Sftpb-predict)	XP_015196624.1
Coelacanth (Sftpb-predict)	XP_006009554.1
Zebra mbuna (Sftpb-predict)	XP_014269510.1
Annual killifish (Sftpb-predict)	XP_013872774.1
Pundamilia nyererei (Sftpb-predict)	XP_005722192.1
Japanese puffer (Sftpb-predict)	XP_003965562.1
zebrafish (sftpba)	ENSDART00000141824.1
zebrafish (sftpbb)	ENSDART00000097254.4
Stickleback (Group XIV: Sftpb-predict)	ENSGACT00000022710.1
Stickleback (Group XIII: Sftpb-predict)	ENSGACT00000011156.1
Tetraodon (Chro 4: Sftpb-predict)	ENSTNIT00000012161.1
Tetraodon (Chro 12: Sftpb-predict)	ENSTNIT00000019156.1
Platyfish (Sca JH557076.1: Sftpb-predict)	ENSXMAT00000004166.1
Platyfish (Sca JH556662.1: Sftpb-predict)	ENSXMAT00000017872.1
Amazon molly (Sca KI519616.1: Sftpb-predict)	ENSPFOT00000019886.2
Amazon molly (Sca KI519905.1: Sftpb-predict)	ENSPFOT00000010269.2
Medaka (Chro 12: Sftpb-predict)	ENSORLT00000011597.1
Medaka (Chro 9: Sftpb-predict)	ENSORLT00000007966.1
Tilapia (Sca GL831238.1: Sftpb-predict)	ENSONIT00000016785.1
Cave fish (sftpbb)	ENSAMXT00000021781.1
Cave fish (sftpba)	ENSAMXT00000019120.1

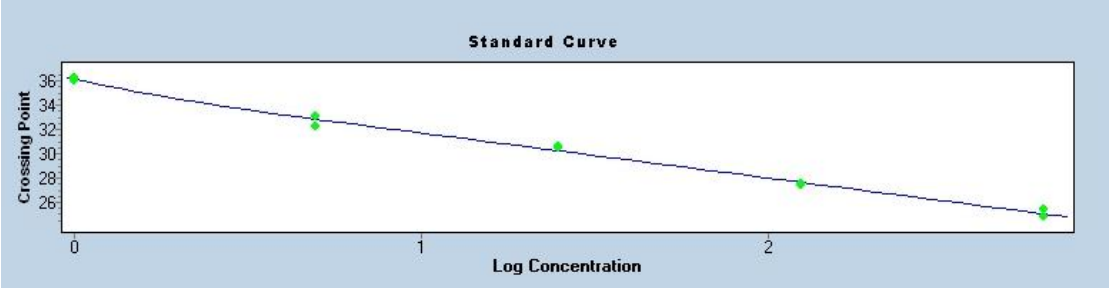
Note: Accession NO. revived from NCBI database (https://www.ncbi.nlm.nih.gov/) is in BLACK, transcript ID revived from Atlantic cod genome database (http://cees-genomes.hpc.uio.no/) is in RED, transcript ID revived from Ensmbl database (http://www.ensembl.org) is in BLUE; “Sca” is the abbreviation of Scaffold, “Chro” is the abbreviation of Chromosome

Appendix 2: Standard curve of Atlantic cod *elovl1a*, *elovl1b*, *pbx1a*, *pbx1b*, *psap*, and *sftpb*

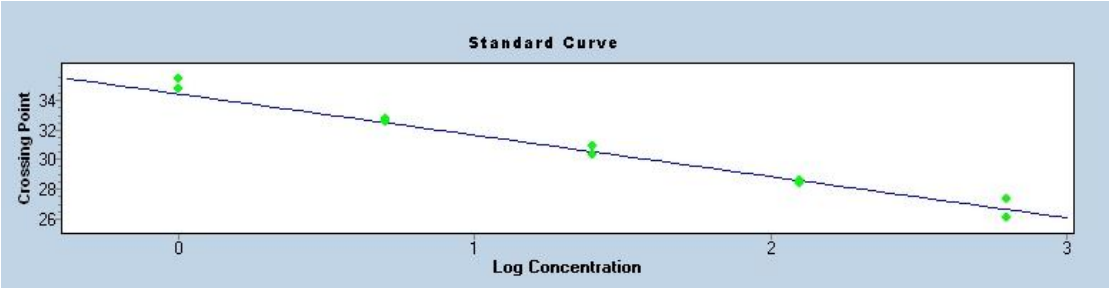
(A) *elovl1a*, efficiency= 2.206



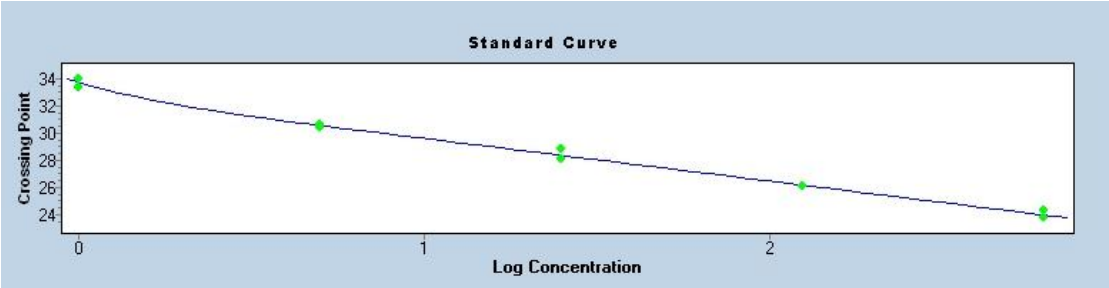
(B) *elovl1b*, efficiency= 1.874



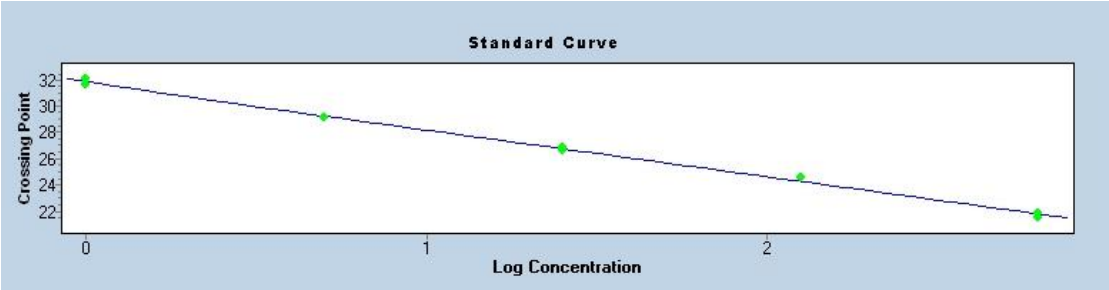
(C) *pbx1a*, efficiency= 2.306



(D) *pbx1b*, efficiency= 2.082



(E) *psap*, efficiency= 1.922



(F) *sftpb*, efficiency= 1.826