Gene expression, phylogenetic and syntenic analyses of pantophysin (*Pan* I) and synaptophysin-like2 (*Sypl2*) genes in Atlantic cod (*Gadus morhua* L.)

Suzana Stjelja

Norwegian University of Life Sciences
Department of Animal and Aquacultural Sciences
Master Thesis 30 credits 2012

Suzana Stjelja

Gene expression, phylogenetic and syntenic analyses of pantophysin (*Pan* I) and synaptophysin-like2 (*Sypl*2) genes in Atlantic cod (*Gadus morhua* L.)

Norwegian University of Life Sciences
Ås, Norway, June 2012

This thesis I dedicate to my father,
Savo Stjelja

Acknowledgements

The presented thesis was carried out at Nofima Marin in Ås, Norway as a part of the European Master in Animal Breeding and Genetics (EM-ABG) and joint collaboration between Norwegian University of Life Sciences in Ås and Swedish University of Agricultural Sciences in Uppsala.

The work with this thesis would not have been possible without becoming admitted at EM-ABG programme and receiving the Erasmus Mundus scholarship. I am deeply grateful for an amazing opportunity to be a part of this great international community. My special thanks go to Prof. Johan A.M. van Arendonk at the Wageningen University for his support and encouragement.

I would like to express my sincere gratitude to my supervisors for the excellent guidance and all the support I received. Dr. Øivind Andersen at Nofima Marin deserves special thanks for introducing me into this project, for having always enthusiastic approach to science and for giving very quick responses. Dr. Stefan Örn at SLU deserves great thanks for valuable discussions and for being always very supportive.

Further I would like to thank Dr. Jacob Torgesen, Hege Munck and Kåtrine Hanes Kirste who were excellent guides through novel techniques in the lab.

I would like to thank my family and friends for their support, encouragement and love.

Uppsala, June 2012

Suzana Stjelja

Abstract

The pantophysin (*Pan* I) locus reveals differences between Norwegian coastal cod (NCC) and Northeast Arctic cod (NEAC) populations that are highly significant, temporally stable and larger than for any other genetic marker. However, the biochemical basis for the selection and the functional role of pantophysin in fishes are unknown. The observed polymorphism might be related to different expression and/or function of the different *Pan* I genotypes.

We applied several different approaches to investigate evolutionary history and spatio-temporal expression patterns of *Pan* I and synaptophysin-like2 (*Sypl*2) gene in Atlantic cod. *In silico* analyses identified pantophysin, synaptophysin and synaptophysin-like genes in various vertebrates, including Atlantic cod. Multiple alignment of amino acid sequences in Atlantic cod and other teleosts revealed a conserved structure characterized by the MARVEL domain, indicating functionally important parts of the examined proteins. Phylogenetic analysis showed that a number of examined vertebrates classified into clade of teleosts and tetrapods. In the teleost clade, separated from tetrapods two clusters were formed. Various taxa from the class *Actinopterygii* with *Pan* I clustered together with *Sypl*1b and *Sypl*2b, while *Sypl*1 and *Sypl*2 clustered together with *Sypl*1a and *Sypl*2a. Teleosts with *Sypl*1 sequences clustered together and formed a separate clade. Similar, tetrapods with *Sypl*1 clustered together, except *Xenopus Sypl*1 that formed a separate branch. The tetrapods *Sypl*2 also cluster together, forming a separate clade. The teleosts *Syp* clustered together with tetrapod orthologs, indicating a common origin. Regions closely linked to Atlantic cod *Pan* I, *Sypl*2 and *Syp*1 and ortologhous genes in medaka, zebrafish, *Xenopus* and human displayed well conserved synteny. The constructed phylogenetic tree and syntenic analysis revealed incorrect sequence naming in several species, including cod *Pan* I.

Since pantophysin expression in fish has not yet been investigated, this study provides the first characterization of the *Pan* I and *Sypl*2 spatio-temporal expression patterns in Atlantic

cod. Using quantitative PCR analysis expression of *Pan* I and *Sypl*2 was quantified in early life stages in NEAC. Both genes were expressed from late gastrula with 8 somites until the larval size of 10-15 mm body length, indicating functional significance during early cod development. The *Sypl*2 transcript was identified by whole mount *in situ* hybridization (WISH) in the embryonic head, heart and liver. In the cod larvae *Sypl*2 mRNA was detected in the eye, liver, heart, neuromast, ceratohyal and ceratobranchial arches, mandible and pectoral fin. Ubiquitous expression of *Pan* I and *Sypl*2 was revealed in various tissues in adult NEAC and NCC. The highest *Pan* I mRNA levels were detected in the head kidney, spleen, epithelial mucus and ovaries while the muscle, testicles and epithelial mucus exhibited the highest *Sypl*2 expression. Interestingly, an inverse expression pattern between *Pan* I and *Sypl*2 was observed in the muscle, head kidney and spleen in examined NEAC, while in NCC both genes showed similar expression patterns with the highest gonadal and epithelial expression.

Table of contents

Abbreviations

A	Adenin
AP	Alkaline phosphatase
BLAST	Basic local alignment search tool
bp	Base pare
cDNA	Complementary DNA
Ct	Threshold cycle
d°	Day degrees, number of days multiplied by the water temperature
ddNTP	Dideoxyribonucleoside triphosphate
DEPC	Diethyl pyrocarbonate
dH_2O	Distillated water
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclase
dNTP	Deoxyribonucleoside triphosphate
dpf	Days post fertilization
dph	Days post hatching
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
H_2O_2	Hydrogen peroxide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISH	*In situ* hybridization
kDa	kDalton, unified atomic mass unit
MAB	Monoclonal antibody
MARVEL	MAL and Related proteins for VEsicle trafficking and membrane Link
MeOH	Methanol
mRNA	Messenger RNA

NBT/BCIP	Nitro blue tetrazolium-/-5-bromo-4-chloro-3-indolyl phosphate
NCC	Norwegian coastal cod
NEAC	Northeast Arctic cod
Pan I	Pantophysin
PBST	Phosphate buffered saline
PFA	Paraformaldehyde
qPCR	Real-time quantitative PCR
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per minute
RT	Reverse transcriptase
SCAMP	Secretory carrier-associated membrane protein
SSC	Saline-sodium citrate
Syp	Synaptopysin
Sypl	Synaptophysin-like
TAE	Tris/Acetate/EDTA
U	Uracil
VAMP	Vesicle associated membrane proteins
WISH	Whole mount *in situ* hybridization
X-Gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
α	Alpha
β	Beta
Δ	Delta
ω	Omega

1. **Introduction**

Pantophysin (*Pan* I) locus is a highly informative non-neutral genetic marker used for differentiation between cod populations in the north-east Atlantic. Differences observed in *Pan* I allele frequencies between Norwegian coastal cod (NCC) and Northeast Arctic cod (NEAC) are highly significant. The homozygous genotype *Pan* I^{AA} is mainly found in the more stationary NCC, while *Pan* I^{BB} genotype is predominate in the migratory NEAC. The *Pan* I alleles differ at both nucleotide and amino acid level. This variation appears to be caused by positive Darwinian selection that favors different alleles in different environments. Different selection pressures are suggested to be linked to water temperature, salinity and depth (Pogson et al., 1995; Fevolden and Pogson, 1997; Pogson, 2001; Pogson and Fevolden, 2003; Pogson and Mesa, 2004; Sarvas and Fevolden, 2005a and 2005b). Pantophysin is a cellular isoform of the synaptic vesicle protein synaptophysin. These two proteins share many sequence features, exhibiting characteristic MARVEL domain (MAL and Related proteins for VEsicle trafficking and membrane Link) (Sanchez-Pulido et al., 2002; Arthur and Stowell, 2007). Pantophysin is ubiquitously expressed in various tissues and its role is poorly understood (Leube, 1994; Haass et al., 1996; Windoffer et al., 1999). Although nothing is known about the functional role of pantophysin in fishes, strong evidence of selection suggests that the observed polymorphism at *Pan* I locus might be linked to different expression and/or function of the different *Pan* I genotypes. Pampoulie et al., (2008) proposed that different pantophysin variants might be related to the migratory behaving in cod. The characterization of the expression patterns of synaptophysin family members may help to reveal their functions and the biochemical basis for the selection.

This quantitative and qualitative study was performed in order to gain knowledge about evolution and spatio-temporal expression patterns of *Pan* I and synaptophysin-like2 (*Sypl*2) gene, with emphasis on differences between NCC and NEAC. Three major goals in this study were:

1. To study evolution of pantophysin, synaptophysin and synaptophysin-like genes in various vertebrates, including Atlantic cod by performing phylogenetic and syntenic analyses,
2. To examine and quantify the expression levels of *Pan* I and *Sypl*2 in early-life stages and in various tissues in adult NCC and NEAC using real-time quantitative PCR (qPCR),
3. To localize the expression of *Pan* I and *Sypl*2 in Atlantic cod embryos and larvae using whole mount *in vitro* hybridization (WISH).

2. Literature review

2.1 Atlantic cod, a commercial fish species

Atlantic cod (*Gadus morhua* L.) is a teleost belonging to the family Gadidae and inhabiting the North Atlantic Ocean. It is a cold acclimated species, widely distributed along the eastern and northern coasts of North America, along the coasts of Greenland and from the Bay of Biscay north to the Arctic Ocean, including the Barents Sea, the North Sea and the Baltic Sea. Within the ocean, cod is found in a variety of habitats from the shoreline down to the continental shelf, at depths up to 600 meters.

Adult cod can grow to approximately 130 cm in length, weighing around 25-30 kg. However larger fish with length of 150 to 200 cm, weighing up to 100 kg have been recorded. Cod can reach a lifespan up to 20 years (Fahay et al., 1999).

The Atlantic cod is a seasonal multiple spawner and comprises multiple migratory or stationary populations. The migratory behavior is associated with reproduction season and seasonal variations in water temperature. During winter cod migrate into warmer waters and gather in spawning aggregations. Spawning occurs in several batches throughout the spawning season, between January and April (Falk-Petersen 2005). Fecundity is high and a large female can release between 3 and 9 million eggs (Fahay et al., 1999). Due to a wide geographical distribution, cod eggs can be found over a range of temperatures from -1.5 °C in the north-west Atlantic to 9 °C in the north-east Atlantic. Duration of the cod embryo development is mainly affected by temperature (Gorodilov et al., 2008). In average hatching occurs after 2-3 weeks in spring conditions with water temperature of 6°C (Lough et al., 1989). The larval phase is initiated by the first exogenous feeding and it lasts about 3 months at 8 °C. The juvenile period is reached at 20-30 mm length when fins are fully developed (Falk-Petersen 2005). The adult period starts with maturation of gonad (Jobling, 1995).

Age of sexual maturity varies between different cod populations. Northeastern populations mature around 5 to 7 years while southern populations mature between 2 to 3 years. However, both size and age at sexual maturity have been reduced most likely due to intense exploitation and harvesting of larger and older fish (Fahay et al., 1999). The Atlantic cod is one of the most exploited, commercially highly valuable fish species in the North Atlantic. The Norwegian commercial fishery harvested in total 341 000 tons (live weight) of wild Atlantic cod in 2011, with a landed value of 3.9 billion NOK. Measured by this value of landed catch, cod fishing was the most important for fisheries in Norway (http://www.ssb.no/fiskeri/).

Due to intense exploitation and size-selective fishing of older and larger fish, many spawning stocks have been seriously reduced and even collapsed. In 2000, cod was placed on the list of endangered species by World Wide Fund for Nature (WWF) (http://www.wwf.org/). Fishing has become restricted in many areas and quotas and licenses have been introduced. Regulated fishing that replaced unlimited exploitation may help to maintain cod populations on sustainable level. However, wide geographical distribution of cod and international waters shared by different countries make difficult to impose regulations. In addition, changes in the food chain, global warming and increased sea temperatures are directly related to the reduction of cod populations (Portner et al., 2001).

Limited supply of wild-caught cod, caused by the decline of cod stocks and seasonal variations in catch created an opportunity for the commercial farming of cod. Thus, cod has become an emerging aquaculture species, offering a solution to cover the market's demands for fresh white fish. Farming programs are being developed in Norway, Iceland, Faroe Islands, Scotland, Canada and United States (Pampoulie et al., 2006). Production of farmed cod in Norway reached a volume of 21 000 tons of slaughtered fish for food in 2010 (http://www.ssb.no/fiskeri/). Although this production is still moderate compared to production of Atlantic salmon in Norway, there is a pronounced will to develop the cod farming into a similar sized industry as salmon.

Cod aquaculture industry meets big challenges such as early sexual maturation, juvenile quality, mortality in female cod at spawning, diseases, escaping from net pens and spawning in sea cages that might lead to unwanted genetic impact on wild cod (Taranger et al., 2010; Jorstad et al., 2008). During sexual maturation the fish invest energy in gonad development, resulting in decreased somatic growth and muscle weight. In addition to longer production cycle and higher mortality, increased filet water content reduces market value (Trippel et al., 2008). Development of breeding programs and applying molecular tools may contribute to increased growth, higher disease resistance and delayed maturation in farmed cod.

2.2 Atlantic cod population structure

Across wide geographical distribution, cod is divided into subpopulations that differ significantly in growth and reproductive characteristics (Brander, 2005). Knowledge of population structure is important for optimal management of wild stocks and for the development of cod aquaculture. Thus, different techniques such as morphometric analyses, conventional taq-recapture programs and genetics have been used to characterize cod populations (Pampoulie et al., 2008).

Based on life-history characteristics, habitats and migration patterns Atlantic cod in Norwegian and adjacent waters is divided into two main groups: the Northeast Arctic cod (NEAC) and the Norwegian coastal cod (NCC) (Sarvas and Fevolden, 2005b). The NEAC, also called the Arcto-Norwegian stock is the largest population of Atlantic cod in the world. This is a migratory stock that inhabits the Barents Sea and waters of northern Norway. In December to January mature NEAC starts its spawning migration from the Barents Sea down to the Norwegian coast. Due to this migrating behavior NEAC is often referred to as “skrei”, a Norwegian word meaning “the wanderer”. The entire spawning season lasts from February to May. The main spawning in March and April takes place mostly in the Lofoten and Vesterålen area in Norway (Sarvas and Fevolden, 2005a).

The NCC is the more stationary coastal population found in northern Norwegian fjords. During the feeding season some of the NCC leaves the fjords for the outer coastal waters but the majority returns back to fjord for spawning. Spawning occurs in locations inside fjords. However, there are some coastal areas in northern Norway where both NEAC and NCC spawn (Godø, 2000; Sarvas and Fevolden, 2005a).

Several studies have been made to investigate whether these two stocks are genetically distinct populations or they belong to one common gene pool. The otolith structure, vertebral numbers, frequencies of blood types, frequencies of hemoglobin (Hb) alleles, several isozyme alleles, mitochondrial cytochrome b (cyt *b*) locus and pantophysin (*Pan* I) have been examined (Nordeide et al., 2011). It was shown that NEAC and NCC exhibit some overlap in otolith structure and vertebral numbers. Based on the isozymes and the cyt *b* locus it was not possible to distinguish these two stocks. However, the blood type E, the Hbl-1 allele and especially the *Pan* I locus have shown distinct genetic differentiation between NEAC and NCC (Pogson et al., 1995; Fevolden and Pogson, 1997; Pogson and Fevolden, 2003; Sarvas and Fevolden, 2005a).

2.3 Pantophysin (*Pan* I) locus

Pogson et al. (1995) discovered an anonymous cDNA clone (GM798) that revealed highly significant variation in the allele frequencies of three tightly linked polymorphic restriction sites between between NEAC and NCC populations. This cDNA clone was sequenced and initially identified as cod synaptophysin (*Syp* I) (Fevolden and Pogson, 1997). However, cDNA clone more likely represent pantophysin, a cellular isoform of the neuroendocrine integral membrane protein synaptophysin (Leube, 1994; Haass et al., 1996).

By aligning nucleotide sequences of human and murine pantophysin with rat synaptophysin Haass et al. (1996) identified 6 exons interrupted by 5 introns in the pantophysin gene. This intron/exon structure was identical to the synaptophysin gene,

except for the last, sixth intron that is present in the synaptophysin 3'-noncoding region but is absent in the pantophysin gene.

2.3.1 Polymorphism at *Pan* I locus

Three polymorphic restriction sites (*Bst*EII, *Dra*I and *Pst*I) were described in the *Pan* I gene region by Pogson et al. (1995). The *Dra*I restriction site is located in the fourth intron of the gene (Figure 1). Depending on the absence or presence of *Dra*I restriction site, the *Pan* I possesses two alleles called *Pan* I^A and *Pan* I^B (Pogson, 2001). These two alleles appeared to have evolved before the speciation event separating Atlantic cod from its sister taxon, the Pacific Alaska pollock (*Theragra chalcogramma*). Based on a standard mtDNA clock it is estimated that the alleles are at least 2 million years old (Pogson and Mesa, 2004).

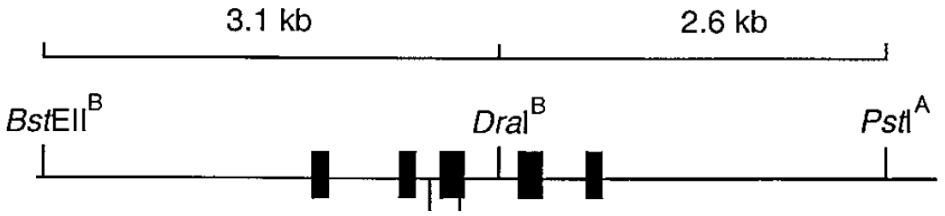


Figure 1 Restriction map of the Pan I gene region showing the locations of polymorphic restriction sites. Exons are presented by solid boxes. The BstEII^B, DraI^B and PstI^B alleles refer to the presence of sites for each restriction enzyme and alternate alleles refer to their absence (Pogson, 2001).

By examining nucleotide sequences of *Pan* I alleles in 5 different populations of Atlantic cod Pogson (2001) showed that nucleotide variation between the *Pan* I^A and *Pan* I^B alleles was much higher compared to variation within either allelic group. Furthermore, *Pan* I^A exhibited much higher nucleotide diversity than *Pan* I^B. In total 15 nucleotide mutations and a 6 bp insertion were fixed between these two alleles, which differentiated mostly in a 30 bp region of the second intron and in a 54 bp region in the fourth exon. These two regions present two peaks of polymorphism across the *Pan* I gene.

2.3.2 Differentiation among cod populations at *Pan* I locus

Fevolden and Pogson (1997) reported significantly higher frequencies of the *Pan* I^A (i.e. *Syp* I^A) allele in NCC (mean = 0.806) compared to NEAC (mean = 0.098). The differences observed in *Pan* I allele frequencies were considerably larger than reported for any other locus, including hemoglobin locus Hb-I. These results suggested that the *Pan* I is a promising genetic marker for differentiating cod populations.

Sarvas and Fevolden (2005b) confirmed that the *Pan* I^A allele and the homozygous genotype *Pan* I^{AA} predominate in NCC, while the *Pan* I^B allele and *Pan* I^{BB} genotype are mainly found in NEAC. The observed difference in allele frequency was stable through years and was found in all age groups. However, intermediate *Pan* I allele frequencies are found in some areas, i.e. outer coastal regions where NEAC and NCC may co-occur. Intermediate allele frequencies could have arisen by a mechanical mixture of two genetically different populations, by interbreeding or through some form of selection. Distinct populations with intermediate allele frequencies may also exist (Sarvas and Fevolden, 2005b).

Because NEAC and NCC occasionally intermingle at spawning sites in northern Norway (i.e Lofoten) it seems that there is an opportunity for gene exchange between these two stocks. However, observed difference in the Pan I allele frequencies challenge the view that extensive gene flow does occur and indicate that two stocks do in fact represent different breeding units. The reasons why NEAC and NCC not interbreed may be different, such as behavioral differences, differences in mating choice, spawning at different depths, different temperatures or at different times (Sarvas and Fevolden, 2005a).

Variation in *Pan* I allele frequencies has been used to investigate existence of genetically distinct populations within a single fjord (Sarvas and Fevolden, 2005b). The results revealed that during the spawning season immature cod with *Pan* I^{BB} genotype inhabits the outer part of the fjord. Thus, this part of the fjord probably serves as a nursery and feeding ground for young NEAC. The inner part of the fjord is mainly inhabited with mature cod

with high frequencies of the *Pan* I^{A} allele, suggesting that the coastal cod is a local breeding stock in the fjord (Sarvas and Fevolden, 2005b).

Data Storage Tags (DSTs) have been used to investigate both horizontal and vertical migration of Atlantic cod, related to habitat choice in feeding migrations in Icelandic waters (Pampoulie et al., 2008). The DSTs results in combination with variation at *Pan* I locus have shown that cod populations are not evenly distributed and indicated the presence of two distinct behavior types. Most of the individuals carrying the *Pan* I^{AA} genotype were likely to remain in relatively shallow waters (<125 m), while individuals carrying the *Pan* I^{BB} genotype preferred deeper and colder waters (>125 m). The heterozygotes showed both type of behaviors. These results showed that Atlantic cod carrying different *Pan* I genotypes exhibited different migrations during the feeding season. Therefore, the *Pan* I expression might be linked to the migratory behavior of cod (Pampoulie et al., 2008).

2.3.3 Natural selection at *Pan* I locus

The *Pan* I locus is used for molecular characterization of cod populations in the north-east Atlantic because it is a highly informative marker. The differences between coastal and Northeast Arctic cod populations revealed by the *Pan* I locus are highly significant, temporally stable and larger than for any other genetic marker (Pogson et al., 1995; Fevolden and Pogson, 1997; Pogson and Fevolden, 2003; Sarvas and Fevolden, 2005a and 2005b). Moreover, the *Pan* I exhibits distinct differentiation between populations at large and small geographic scales and does not show a relationship between levels of gene flow and geographic distance (Pogson et al., 2001). This locus is also unique in exhibiting nearly complete linkage disequilibrium among three restriction sites in the gene region (Pogson and Fevolden, 1998).

Loci that exhibit unusually high levels of variation might indicate the possible action of natural selection that favors different alleles in different environments. Therefore, it was suspected that the *Pan* I is undergoing different selection pressures between coastal and

Northeast Arctic cod populations (Fevolden and Pogson, 1995, 1997; Pogson and Fevolden, 1998, 2003). Nucleotide and amino acid sequence variation among *Pan* I alleles have provided evidence of strong natural selection at this locus. The *Pan* I^A and *Pan* I^B alleles differ at the nucleotide level on average by 19 mutations and on the protein level each allele has undergone three amino acid substitutions since diverging from a common ancestral allele. This variation appears to be caused by unusual mixture of balancing and directional selection. However, the type of balancing selection that might be acting at the *Pan* I locus is presently unknown (Pogson, 2001). It is also confirmed that positive Darwinian selection at the *Pan* I locus is widespread among marine fishes belonging to the family Gadidae (Pogson and Mesa, 2004).

Observed polymorphism and strong footprint of selection at *Pan* I in Atlantic cod suggest that it might perform some critical function. Since nothing is known about the functioning of pantophysin in fishes, the biochemical basis for the selection is also poorly understood. It was suggested that water temperature, salinity and depth might be linked to selection at this locus. It is also possible that selection acts on other genes associated to *Pan* I (Sarvas and Fevolden 2005a; Pampoulie et al., 2008).

2.4 Structure of pantophysin protein

Pantophysin is a cellular isoform of the neuroendocrine-specific protein synaptophysin (Haass et al., 1996). Both proteins belong to a growing family of integral membrane proteins found in synaptic or cytoplasmic vesicles. The family consists of three distinct groups:

1. Physins – a group that includes two neuronal isoforms synaptophysin (*Syp* 1) and synaptoporin (synaptophysin II), pantophysin (synaptophysin-like protein 1) and mitsugumin, an isoform identified in striated skeletal muscle (Haass et al., 1996; Leube 1994; Arthur and Stowell, 2007)

2. Gyrins - a group with neuronal synaptogyrin and ubiquitous cellugyrin (Stenius et al., 1995; Janz and Südhof, 1998);
3. SCAMPs (secretory carrier-associated membrane proteins) with several related isoforms (Singleton et al., 1997).

Among these three groups, the physins and especially synaptophysin have been studied the most. Synaptophysin is a cell-type restricted protein, detected in synaptic vesicles in neurons and neuroendocrine cells only. It was the first synaptic vesicle protein to be cloned and identified (Jahn et al., 1985; Wiedenmann and Franke, 1985). This 38 kDa protein is the most abundant synaptic vesicle protein by mass, accounting for 10% of total vesicle protein (Takamori et al., 2006).

Synaptophysin and pantophysin are members of the MARVEL (MAL and Related proteins for VEsicle trafficking and membrane Link) domain family. The MARVEL domain is characterized by four transmembrane-helix architecture and it is found in vertebrate proteins such as myelin and lymphocyte proteins, involved in processes of tight junction formation, vesicle trafficking and regulation of membrane fusion events (Sanchez-Pulido et al., 2002; Arthur and Stowell, 2007). Even though the precise role of the MARVEL domain is still unclear, the high degree of conservation of transmembrane domains suggests that they are essential for basic properties common to all members of the MARVEL family.

Secondary structure of synaptophysin consists of four transmembrane hydrophobic regions with α helical conformation, two intravesicular loops and two tails, short amino- and long carboxy–terminal tail. Both tails are located on the outer cytoplasmic surface of the vesicles. Pantophysin is a homolog of the synaptophysin and these two proteins share many sequence features. The alignment of amino acid sequences of pantophysin, synaptophysin and other members of the synaptophysin gene family in different species (Figure 2) shows that the highest level of homology is across transmembrane domains, the intravesicular loops show the lower degree of conservation, while amino- and carboxyl tails exhibit the highest degree of sequence variability (Leube, 1994; Pogson, 2001).

Synaptophysin is N-glycosylated on the first intravesicular loop and phosphorylated on the carboxyl tail which also contains a Ca^{2+} binding site and proline- and tyrosine-repeated motifs (Wiedenmann and Franke, 1985; Leube, 1994; Pogson 2001; Arthur and Stowell, 2007). Similar to synaptophysin, the first intravesicular loop in pantophysin contains a potential N-glycosylation site in rat and human pantophysin, although in shifted position and conserved YPFRL region with phenylalanines on both sides. Paired cysteine residues that form intramolecular disulfide bonds are found in both loops. However, pantophysin sequences are lacking the C terminus, characteristic for synaptophysins (Leube, 1994; Pogson, 2001).

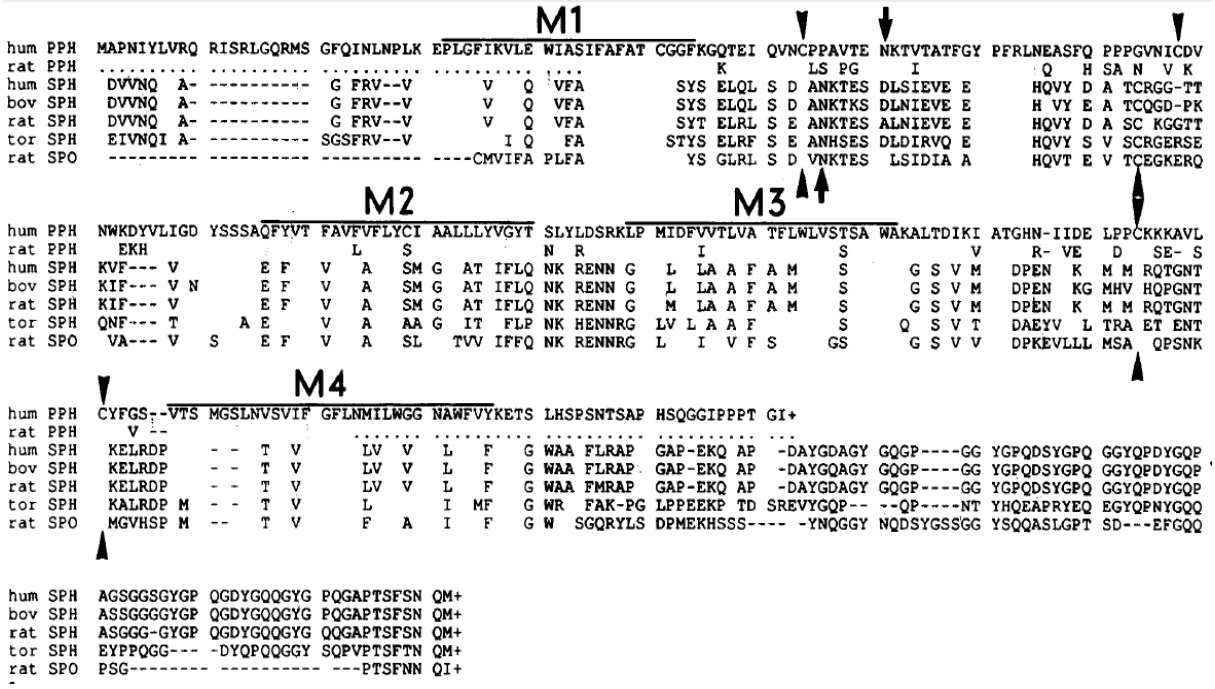


Figure 2 Comparison of cDNA derived amino acid sequence from human pantophysin (hum PPH) to rat pantophysin (rat PPH), synaptophysin from human (hum SPH), cow (bov SPH) and the ray fish Torpedo californica (tor SPH) as well as to rat synaptoporyn (rat SPO). The locations of the four transmembrane domains are labeled with M1 to M4. The presence of cysteine in intravesicular loops is marked with arrow heads and a potential N-glycosylation site with an arrow (Leube, 1994).

The highly conserved structure of pantophysin allows all mutations, identified between *Pan* I alleles to be localized at distinct domains (Table 1) (Pogson 2001). One-quarter of the identified polymorphisms were found within coding DNA and nine involved amino acid substitutions. Six of these replacement mutations were fixed between the two *Pan* I alleles (three within each allele) and all occurred in the 56 amino acid region in the first intravesicular loop. This suggests that the first intravesicular loop might be an important domain of the protein, performing some critical functions (Pogson 2001).

Table 1 Polymorphism at Pan I locus and amino acid replacement mutations. The first intravesicular loop is marked with IV1, the second transmembrane domain with M2 and the cytoplasmatic carboxy-terminus with CYT2 (Pogson, 2001).

Allele	Codon position	Nucleotide change	Amino acid change	Location	Classification	Distribution in sample
Pan I^A	61	AAA → CAA	K → Q	IV1	Radical	Fixed
	64	AAC → ACC	N → T	IV1	Radical	Fixed
	71	GAG → AAG	E → K	IV1	Very radical	Polymorphic
	79	TCT → ACT	S → T	IV1	Radical	Fixed
	214	GAC → TAC	D → Y	CYT2	Radical	Polymorphic
Pan I^B	43	GAG → GTG	E → V	IV1	Radical	Fixed
	61	AAA → AAT	K → N	IV1	Radical	Fixed
	64	AAC → GAC	N → D	IV1	Radical	Fixed
	92	TCC → ACC	S → T	M2	Radical	Polymorphic

Arthur and Stowell (2007) presented the first three dimensional structure of the synaptophysin (Figure 3). The structure reveals that synaptophysin forms a basket-like formation with six individual spokes. This hexameric complex has a closed conformation on the cytosolic side of the membrane and an extended open conformation on the vesicle lumen side and a large pore within the membrane (Figure 4).

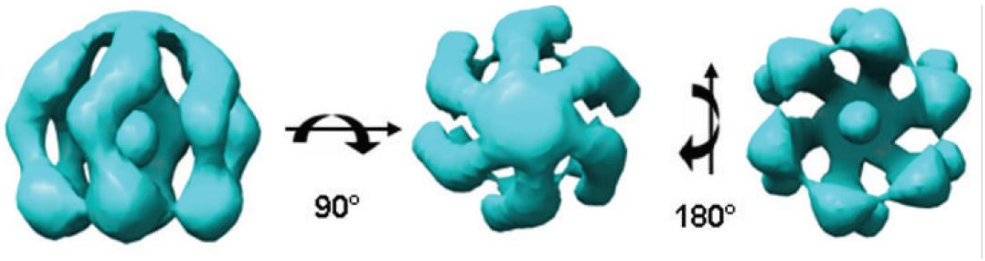


Figure 3 Three views of synaptophysin complex three dimensional reconstruction, rendered at 20 Å resolution. The structure shows an overall diameter of 70 Å with an inner diameter of 30 Å (Arthur and Stowell, 2007).

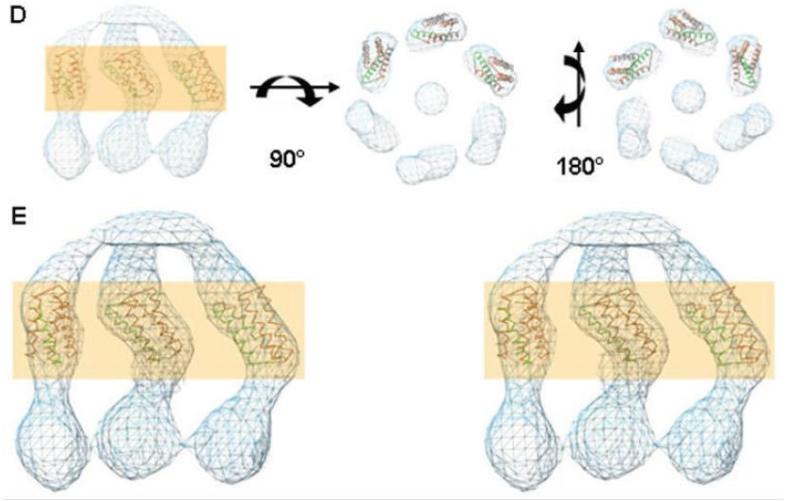


Figure 4 (D) Mesh view of synaptophysin complex. (E) Stereo mesh view of synaptophysin complex that shows the inner wall of the complex. The lipid bilayer is represented by highlighted area (Arthur and Stowell, 2007).

The described structure suggests that synaptophysin may form a pore complex during synaptic vesicle fusion and supports the role played by synaptophysin in vesicle membrane interactions and may help to understand the role of other synaptophysin family proteins, including the pantophysin (Arthur and Stowell, 2007).

2.5 Expression of pantophysin

Pantophysin displays similarities in gene structure and amino acid sequence with synaptophysin. In neurons and neuroendocrine tissues synaptophysin and pantophysin colocalize in the same cytoplasmic vesicles (Haass et al., 1996; Windoffer et al., 1999). However, expression of pantophysin mRNA and protein is not restricted to neurons and neuroendocrine cells only. Using antibodies against the unique and cytoplasmically exposed carboxy terminus, pantophysin was shown to be a widely distributed integral membrane protein found in cytoplasmic vesicles in both neuroendocrine and non-neuroendocrine tissues. Its expression was not restricted to a particular cell type but to the vesicles of similar size and appearance (Leube, 1994; Haass et al., 1996).

Pantophysin has been localized to small vesicles with a diameter <100 nm (average 40-70 nm), a smooth surface and an electron-translucent lumen (Figure 5). Independent of their cargo, pantophysin is present in different transport vesicles that function in intracellular trafficking pathways between various membrane compartments. It was shown that pantophysin vesicles also contain vesicle-associated membrane protein 3 (VAMP3)/cellubrevin and secretory associated membrane proteins (SCAMPs).

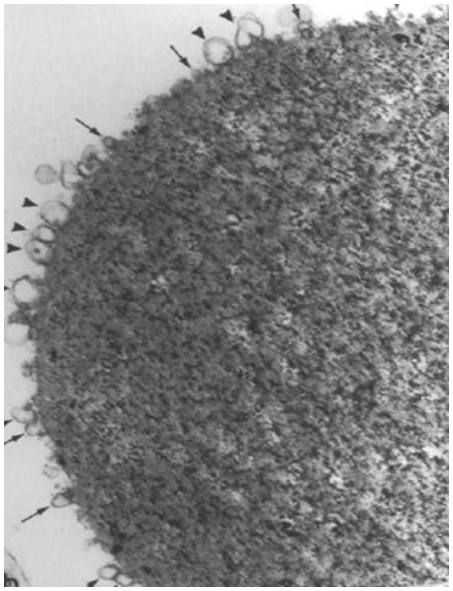


Figure 5 Electron micro-graph of vesicles immunoisolated with pantophysin. The pantophysin vesicles are marked with arrows (Haass et al., 1996).

The intracellular distribution revealed by immunoelectron microscopy of human cultured cells has shown that multiple dotlike structures-containing pantophysin were abundant in the peripheral flat cytoplasm regions near the plasma membrane of neighboring cells. The small clusters of pantophysin particles were present in peripheral regions of endoplasmic reticulum, Golgi apparatus, multivesicular bodies and plasma membrane. However, pantophysin was not detected in coated vesicles, mitochondria, nucleus and cytoplasmic filaments (Haass et al., 1996).

Leube (1994) demonstrated by extensive PCR analyses and Northern blot hybridization experiments that pantophysin gene is ubiquitously expressed *in vitro* and *in vivo*. The pantophysin mRNA was found in all cell lines from normal tissues as well in cultured cells derived from epithelia, fibroblasts, muscle cells or neuronal tissues.

Haass et al. (1996) reported that pantophysin immunofluorescence was strongest in the exocrine portion of human pancreas. Expression in the neuroendocrine cells of Langerhans islets was weaker but still significant detectable. In addition, pantophysin was detected in blood vessel walls, including endothelial and myoid cells and in connective tissue cells. The immunological detection in bovine tissues has shown that pantophysin expression was heterogeneous and differed significantly between various tissues (Windoffer et al., 1999). Pantophysin was clearly detectable in testis, parotid gland, pancreas and intestines. In striated skeletal muscle and liver pantophysin was expressed in trace amounts only. In liver pantophysin expression was restricted to cells surrounding sinusoids and was not detectable in hepatocytes. Remarkably, it was shown that hepatocytes lack pantophysin, VAMP and SCAMP epitopes even though they are highly active cells in terms of membrane trafficking. Observed pantophysin expression in different tissues was not parallel with expression of other vesicle proteins (cellubrevin and SCAMPs) that colocalize in the same cytoplasmic vesicles (Windoffer et al., 1999).

Brooks et al (2000) have reported the first characterization of pantophysin in adipocytes, the cells that have a large volume of intracellular vesicle traffic. It was shown that pantophysin mRNA was abundant in adipose tissue and increased during adipogenesis.

Pantophysin was detected in a variety of adipocyte transport vesicles. Furthermore, pantophysin was present in the same vesicles that contain an insulin-sensitive glucose transporter GLUT4. These results suggest that this subpopulation of pantophysin-GLUT4-containing vesicles exhibits insulin stimulated regulation.

2.6 The role of pantophysin

Pantophysin is associated with small cytoplasmic transport vesicles that function in various intracellular trafficking pathways. However, very little is known about the role of pantophysin in these pathways and fusion events. Ubiquitous expression of pantophysin and its highly conserved molecular structure suggest that it may perform basic structural and housekeeping cellular functions.

It has been shown that pantophysin colocalizes with synaptophysin and other related vesicle proteins in the same vesicles. This mutually not exclusive co-expression suggests that all these molecules may perform similar overlapping functions, including regulation of vesicle fusion (Leube, 1994; Haass et al., 1996; Windoffer et al., 1999).

Existence of functional redundancy might explain why the deletion of synaptophysin did not cause major phenotypic defects in knock-out mice (Eshkind and Leube, 1995). A number of diverse functions have been attributed to synaptophysin and synaptophysin related proteins, including exocytosis, synapse formation, formation of pores for vesicle fusion, biogenesis and endocytosis of synaptic vesicles (Leube et al., 1987). The morphology, shape, characteristic dimension and polypeptide composition of synaptic vesicles as well as synaptic transmission were not altered in the absence of synaptophysin. In addition, no drastic changes in mRNA levels of synaptophysin isoforms, including pantophysin were noted. One possible explanation for the absence of detectable functional impairments in mice may be molecular redundancy that allows compensation of synaptophysin by pantophysin or other related membrane vesicle proteins (Eshkind and Leube 1995).

Synaptophysin-deficient mice may have defects in vesicle recycling and synaptic transmission that were not detected by previous studies (Eshkind and Leube 1995). Also, some of defects may be not crucial for vital functions but might be important for specific functions. Support for this came from the behavioral study in mice by Schmitt et al (2009), showing that synaptophysin depletion does have functional consequences and affects higher brain functions. Detailed behavioral analyses, including tests focusing on novelty, anxiety and learning/memory and visual/physiological analyses were performed on synaptophysin knockout mice (Schmitt et al, 2009). The results showed that synaptophysin knockout mice were more exploratory than wild type control mice and they examined novel objects more closely and intensely. Furthermore, knockout mice had a significant deficit in object recognition despite their enhanced object exploration. In addition, deficits in learning and memory and reduced spatial learning were observed (Schmitt et al, 2009).

A recent study by Kwon and Chapman (2011) based on a quantitative analysis of synaptic vesicles recycling in cultured neurons revealed that distinct structural elements of synaptophysin differentially regulate the vesicle retrieval-endocytosis during and after neuronal stimulation. Truncation of the C-terminal tail of synaptophysin caused slower endocytosis during neuronal activity but had no effect on endocytosis after neuronal activity. Therefore, the loss of synaptophysin causes synaptic depression and delay the replacement of usable synaptic vesicles during sustained neuronal activity. These findings are also supported by previously reported alterations of synaptophysin in neurons in case of nerve injury, increased age, Alzheimer's disease or chronic restraint stress. Based on the observed results it was proposed that synaptophysin has the role to maintain endocytic capacity, a number of synaptic vesicles that can be retrieved in synapses and it is required to ensure synaptic vesicle availability at release sites (Kwon and Chapman, 2011).

Discovering the role of synaptophysin may help to understand the role of other synaptophysin family proteins, including pantophysin. However, previous studies about molecular structure, expression and function of pantophysin and synaptophysin-like proteins have been focused on mammalian species. Even though the *Pan* I locus is a highly

informative marker used for molecular characterization of cod populations, nothing is known about tissue-specific expression, intracellular distribution or functioning of pantophysin in fishes. The differences detected between the *Pan* I alleles in cod populations in the north-east Atlantic suggest that observed polymorphism might be related to the differential expression and/or functioning of the protein in different tissues.

3. Material and methods

3.1 Sample collection

cDNA samples from Atlantic cod representing the NEAC population were kindly provided by Dr. Hanne Johnsen, University in Tromsø, Norway. The samples from early life stages included cDNAs from late gastrula at 8 somite stage (6 days post fertilization, dpf), 10-somite stage (7 dpf), 45-somite stage (14 dpf), hatching (21 dpf), feeding larvae of 5 mm body length (12 days post hatching, dph) and larvae of 10-15 mm body length (35 dph). The cDNAs from tissues in adult NEAC included brain, eye, head kidney, gill, heart, muscle, pyloric ceca, spleen and liver sampled from one male and one female fish. Tissue dissection, total RNA isolation and cDNA synthesis were performed as described in Johnsen et al. (2010). The cDNA was diluted 1:10 and stored at -20 °C for further use in stage and tissue-specific expression analysis.

Atlantic cod embryos of 60 d° and larvae of 90 d° (days after fertilization multiplied with water temperature of 6 °C) were provided by Dr. Jacob Torgersen, Nofima in Ås, Norway. After dechorionation, 80 embryos and 80 larvae were submerged in 100 % methanol, stored at -20 °C and used for WISH.

Tissue samples from adult CC were a gift from Dr. Sissel Jentoft, Centre of Ecological and Evolutionary Synthesis (CEES), University of Oslo, Norway. The tissues dissected from two females and one male fish included telencephalon, optic tectum, mesencephalon, cerebellum, hypophysis, medulla oblongata, epithelial mucus, heart, head kidney, liver, gonads, anterior part and distal part of the intestines (colon). Tissue dissection was performed as described in Star et al (2011). The tissues were submerged in RNAlater and stored at -20 °C before total RNA isolation.

3.2 Preparation of RNA

3.2.1 Total RNA isolation

Total RNA was isolated from the NCC tissues by using Isol-RNA Lysis Reagent (5 PRIME) and PureLink™ Pro 96 RNA Purification Kit (Invitrogen, Life Technologies). The Isol reagent lyses cells, removes proteins and DNA, inactivates RNases and protects the RNA during extraction.

The frozen tissue samples, submerged in RNAlater were transferred from -20 °C and placed on ice until they thawed. Small pieces of tissues were cut with a clean scalpel and transferred into 2 ml sample tubes containing 800 µl Isol-RNA Lysis Reagent and 6-7 ceramic beads (1.4 mm zirconium oxide beads). The tissues were homogenized using a tissue homogenizer (Precellys®24, Bertin Technologies) with the speed of 5 500 rpm during two cycles of 20 sec with 5 sec between the cycles. After homogenization 200 µl of chloroform was added in each sample. The tubes were manually shaken vigorously for 15 sec and incubated for 2-3 min at room temperature. By centrifugation at 12 000 x g for 15 min at 4 °C the homogenate was separated into an upper aqueous phase and a lower organic phase. The aqueous phase contains RNA while proteins remain in the organic phase and DNA in the inter-phase. Thus, 300 µl of the aqueous phase was transferred into a 1.5 ml eppendorf tube and an equal volume of lysis buffer was added. To remove any contaminating phenol remained after lysis with the Isol and to clean-up the RNA 300 µl 96 % ethanol was added. The solution was mixed well and transferred into a 96-well RNA filter plate which was placed on top of a receiver plate. The stacked plates were centrifuged at 2 100 x g for 2 min at room temperature. During centrifugation the RNA binds to a silica-based membrane in the filter plate. The flow-through was discarded. To each well of the filter plate 500 µl 1X Wash Buffer II was added. After centrifugation at 2 100 x g for 2 min and discarding the flow-through, the filter plate was prepared for DNase digestion.

3.2.2 DNase treatment

On-Column DNase digestion was performed in order to remove any genomic DNA contamination from the RNA preparations. DNase I is an enzyme that digests DNA but does not cause RNA degradation.

DNase I solution was prepared by pipetting 8 µl 10X DNase I Buffer, 9.9 µl DNase I (2 700 units) and 62 µl RNase-free water in a sterile RNase-free tube for each reaction. A high concentration of the enzyme was recommended in order to provide its full activity in the presence of salts and chelating agents. A solution of 80 µl DNase I was added to each well of the filter plate and incubated 30 min at room temperature.

3.2.3 Washing RNA

After DNase digestion, 500 µl PureLink™ Pro 96 Wash Buffer I (Invitrogen, Life Technologies) was added to each well of the filter plate and centrifuged at 2 100 x g for 2 min. The flow-through from the receiver plate was removed. Then 750 µl PureLink™ Pro 96 Wash Buffer II (Invitrogen, Life Technologies) was added per well and centrifuged at 2 100 x g for 2 min. The washing step with Wash Buffer II and subsequent centrifugation were repeated twice. Centrifugation at 2 100 x g for 10 min was performed to dry the silica membrane.

3.2.4 RNA elution

50 µl RNase free water was added to each sample in the filter plate, placed on top of a new elution plate. After incubation for 1 min at room temperature, stacked plates were centrifuged at 2 100 x g for 2 min. Eluted RNA was transferred from the elution plate into a new 1.5 ml eppendorf tube and stored at –80 °C for cDNA synthesis.

The quality and concentration of the purified RNA was measured by Nanodrop spectrophotometer (NanoDrop®, Saaven Werner AB).

3.3 Complementary DNA (cDNA) synthesis

Total RNA was reverse transcribed into complementary DNA (cDNA) with TaqMan® Reverse Transcription Reagents Kit (Applied Biosystems, Roche). A single-stranded cDNA, complementary to the mRNA is generated by the enzyme reverse transcriptase (RT) in the process that includes three phases:

1. Incubation – hybridization of oligo $d(T)_{16}$ primers to 3' poly (A) tail of the mRNA or hybridization of random hexamers anywhere on the total RNA
2. Reverse transcription – single-stranded cDNA is synthesized by RT
3. RT inactivation – termination of reaction

The following cDNA synthesis mix was prepared for each reaction:

10x RT Buffer	3.0 µl
25 mM $MgCl_2$	6.6 µl
10 mM dNTP Mix	6.0 µl
50 µM oligo $d(T)_{16}$	0.75 µl
50 µM Random Hexamers	0.75 µl
RNase Inhibitor 20 U/ µl	1.0 µl
MultiScribe™ Reverse Transcriptase 50U/ µl	0.7 µl
RNA 400 ng	x µl
RNase-free water up to final volume	30.0 µl

Prepared cDNA synthesis mix was kept on ice. The RT-PCR program used was:

1. 25°C for 10 min
2. 48 °C for 60 min

3. 95 °C for 5 min
4. 4 °C for infinite

Synthesized cDNA was stored at -20 °C for further use in tissue-specific expression analysis.

3.4 PCR amplification of cod *Pan* I and *Sypl*2

cDNA from various tissues in adult NEAC and CC was used as template in PCR reactions with gene-specific primers (Table 1, Appendix A). The forward and reverse primers were designed based on the *Pan* I (ATLCOD1Bc1522258, Figure A1 in Appendix A) and *Sypl*2 cDNA sequences (ATLCOD1Bc1511891, Figure A2 in Appendix A) using the Vector NTI® software and the primers were synthesized by Invitrogen (Life Technologies).

The following PCR mixture was prepared for each reaction:

10x Advantage® cDNA PCR Buffer	2.5 µl
2 mM dNTP	5.0 µl
Advantage® cDNA Polymerase Mix	0.5 µl
Forward primer (10 µM)	1.0 µl
Reverse primer (10 µM)	1.0 µl
cDNA 10x	1.0 µl
dH_2O up to final volume	25.0 µl

Prepared PCR mixture was kept on ice. The PCR program used was:

1. 95 °C for 1 min
2. 40 cycles of: 95 °C for 30 sec, 55-66°C for 30 sec and 68 °C for 1 min
3. 68 °C for 6 min
4. 4°C for infinite

The amplified PCR products were confirmed by gel electrophoresis.

3.5 Visualization by agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by length and to estimate the size of DNA. During gel electrophoresis particles can be separated both by charge and size. By applying an electric field negatively charged nucleic acid molecules move through an agarose gel toward the positive electrode (anode). Molecules that are shorter migrate more easily through the pores of the gel, moving faster and migrating further than longer molecules. Size of the fragments is estimated by comparison to a ladder with fragments of known size.

Agarose gels of 1.0-2.0 % were used depending on the size of the fragments to be separated. Lower concentrations of agarose allow resolution of larger fragments while higher concentrations allow better separation of small fragments. The agarose gels were prepared by mixing agarose powder with 1 x TAE buffer (Tris/Acetate/EDTA) to the desired concentration (1.0-2.0 gr agarose in 100-200 ml 1 x TAE buffer) and heating it in a microwave until agarose was completely melted. Then RedSafe™ staining solution (Intron Biotechnology) was added (1 µl of RedSafe in 30 ml gel and 2 µl in 50 ml gel). The RedSafe emits green fluorescence when bound to DNA and allows visualization of DNA when electrophoresis is complete. The melted agarose was poured into a plastic tray with tape around the sides to prevent spillage and with a comb to form sample wells. After 30 min at room temperature the gel has solidified and the comb and the tape were removed. The gel in the plastic tray was placed in an electrophoresis chamber filled with 1 x TEA buffer. DNA ladder (100 bp DNA ladder, BioLabs) and samples containing DNA and loading dye solution (6 x Mass Ruler™, Fermentas) were carefully loaded into the sample wells. After covering the chamber with a lid, power leads were connected with the power source (PowerPac 300, Bio-Rad). The voltage applied to a gel was 75-120 volt for 15-30 min depending on the size of the fragments to be separated. After electrophoresis was complete, DNA fragments

were visualized with Gel Doc™ EZ Imager and analyzed with the software Image Lab (Bio-Rad).

DNA fragments confirmed by electrophoresis were purified from the agarose gels or from the PCR products and used for further analysis.

3.6 PCR product purification

PCR products were purified with Montage™ PCR Centrifugal Filter Devices (Millipore). This kit separates amplified DNA from primers, unincorporated dNTPs and other excess reaction components that can interfere with future cloning or sequencing.

dH_2O was added to 10-20 µl of PCR product in order to make 100 µl PCR reaction that was pipetted in a sample reservoir, inserted into filtrate collection tube. The volume in the sample reservoir was adjusted to a final volume of 400 µl by adding dH_2O and centrifuged at 1 000 x g for 15 min. The reservoir was removed from the filtrate collection tube, placed upright into a new clean tube and 20 µl dH_2O was added in the reservoir. The reservoir was then inverted and centrifuged at 1 000 x g for 2 min. The purified DNA was stored at -20 °C for further use in cloning and direct sequencing.

3.7 Extraction of DNA fragments from agarose gel

Zymoclean™ Gel DNA Recovery Kit (Zymo Research) was used to recover DNA fragments from agarose gel.

DNA fragments of interest were carefully excised from the agarose gel with a clean scalpel and transferred into 1.5 ml eppendorf tubes. To each volume of agarose excised from the gel 3 volumes of ADB-Buffer (Agarose Dissolving Buffer) were added. The samples were incubated at 50 °C for 5-10 min until the gel slice was dissolved completely. Melted agarose was transferred to a Zymo-Spin™ Column, placed into a collection tube and centrifuged at 10 000 x g for 60 sec. The flow-through was discarded and 200 µl of wash buffer was added and centrifuged at 10 000 x g for 30 sec. After discarding the flow-through, the wash step

was repeated. The column was placed into a new 1.5 ml eppendorf tube and 10 µl dH2O was added and centrifuged at 10 000 x g for 60 sec. Eluted DNA was stored at -20 °C for further use in cloning and sequencing.

3.8 Cloning of amplified fragments of cod *Pan* I and *Syp*l2

Cloning of DNA is based on the replication of a recombinant DNA molecule within a host organism in order to generate a large population of cells containing identical DNA molecules. First, recombinant DNA molecules are assembled by insertion of a fragment of interest into a cloning vector that is able to replicate inside a host cell. The recombinant DNA is then introduced into a host organism and replicated along with the host DNA. In this way a large number of copies of the original recombinant DNA molecule are generated.

3.8.1 Ligation reaction

DNA purified from PCR products or from agarose gels was cloned by using the pDrive Cloning Vector (Qiagen). The vector is supplied in a linear form with size of 3851 bp and it contains a large number of unique restriction enzyme recognition sites, universal sequencing primer sites and promoters for *in vitro* transcription (SP6 and T7). It also allows ampicillin and kanamycin selection and blue-white screening of recombinant colonies because contains a fragment of the *lacZ* gene (*lacZα*).

Cloning with pDrive vector is based on UA hybridization. The pDrive vector has at each 3' end an U overhang that hybridizes to the A overhang of PCR products. Therefore, when using the pDrive vector PCR products should be created by Taq DNA polymerase or any other polymerase that lacks 3' to 5' proofreading activity and adds a single A at each 3' end of the PCR product. The cloning efficiency of the pDrive Cloning Vector is improved by lower tolerance of U for a nonspecific base pairing compared to TA cloning vectors. Optimal hybridization conditions are provided by Ligation Master Mix containing DNA

ligase which synthetizes phosphodiestar bonds between nucleotides at the ends of vector insert.

The following ligation mixture was prepared:

Ligation Master Mix	5.0 µl
pDrive Cloning Vector	1.0 µl
PCR product	4.0 µl
Total volume	10.0 µl

After short spinning and gently vortexing, the ligation mixture was incubated at 4 °C over night.

3.8.2 Transformation reaction

The ligation mixture was transformed into *E. coli* XL1-Blue Competent Cells (Agilent Technologies, Stratagene). This host strain is resistant to tetracycline and contains a fragment of the *lacZ* gene (*lacZΔM15*), allowing blue-white color screening for recombinant colonies.

XL1-Blue Competent Cells were transferred from - 80 °C and placed on ice. When thawed, the tubes with cells were gently swirled and 50 µl cells were pipetted into pre-chilled 1.5 ml eppendorf tubes. To each aliquot of cells 0.75 µl β-mercaptoethanol was added. The tubes were incubated on ice for 10 min and mixed gently every 2 min. Then 2 µl ligation reaction was added and the samples were incubated on ice for 30 min. The mixture was heat-shocked in water bath at 42°C for 45 sec. After incubation on ice for 2 min, 0.9 ml Lysogeny broth (LB) medium pre-heated at 42°C was added. The tubes were incubated at 37 °C for 1 hr with shaking (225-250 rpm).

LB agar plates for cultivation of *E. coli* colonies were prepared by mixing 5 gr yeast extract, 10 gr tryptone, 10 gr NaCl and 15 gr agarose and adding the mixed powder into 1 l dH_2O. The solution was vigorously stirred and autoclaved at 121 °C for 20 min. When the liquefied agar has cooled (about 55 °C), it was poured into a sterile plastic petry plate and kept at room temperature until it was solidified. The plates with solid agar were flipped upside down to avoid condensation on the agar.

The mixture containing 40 µl ampicillin (50 mg/ml), 40 µl X-Gal and 1 µl 20 % IPTG was prepared and spread evenly on each LB agar plate, with a sterile L-shaped plastic rod. The plates were kept at room temperature for 15 min to absorb the ingredients. Then 100-150 µl cell culture was plated onto prepared plates. The plates were kept at room temperature for 10 min, wrapped in aluminum foil and incubated upside down at 37 °C over night.

3.8.3 Selection and analysis of transformants

Blue-white color screening was used to select for *E. coli* colonies carrying recombinant plasmids. This method relies on the principle of α-complementation of the β-galactosidase gene (*lacZ* gene).

The pDrive vector contains a fragment of the *lacZ* gene called *lacZα* which encodes the α-peptide of the β-galactosidase. The *lacZα* is complement to the *lacZΔM15* which is present in the *E. coli* XL1-Blue Competent Cells and encodes the ω-peptide. These peptides are not functional by themselves. However, when the vector is transformed into competent cells and *lacZα* and *lacZΔM15* are expressed together, they produce a functional β-galactosidase. Disruption of this process allows the blue-white color screening. The cloning site in the pDrive vector is located within the *lacZα* fragment. Thus, when an insert is successfully ligated it disrupts the *lacZα* fragment. In this way α-complementation is interrupted and in cells with the vector containing an insert no functional β-galactosidase can form.

An active β-galactosidase hydrolyses substrate X-Gal into a blue-colored product. Therefore, if expression of the *lacZ* gene is induced by IPTG in the presence of X-Gal, the presence or absence of a functional β-galactosidase can be detected. In colonies containing a vector with an insert, the functional β-galactosidase is not produced and the X-Gal is not hydrolyzed. Therefore, these colonies will be white while blue colonies indicate the presence of an active β-galactosidase so they contain a vector without an insert.

White colonies were picked from the plates and each colony was transferred into 4-5 ml LB medium with 5 µl ampicillin (50 mg/ml). The colonies were incubated at 37 °C over night with shaking (225 rpm). Incubation longer than 16 hrs was avoided in order to prevent cell lysis and loss of the recombinant plasmid in cell debris. After incubation, the tubes were transferred on ice until plasmid isolation.

3.8.4 Isolation of recombinant plasmid from *E. coli*

Plasmids were isolated from *E. coli* cells using NucleoSpin® Plasmid Kit (Macherey-Nagel). The purification procedure involves alkaline lysis of bacterial cells. Alkaline pH denatures chromosomal DNA but not the circular plasmid DNA. After removing proteins and chromosomal DNA, the plasmid DNA binds to a silica membrane at high salt buffer conditions. Under low salt conditions, the plasmid DNA is washed and eluted from the silica membrane.

Bacterial cells were harvested by centrifugation at 11 000 x g for 30 sec. The supernatant was discarded and the rest of liquid was removed by turning the tube upside-down and drying it on a paper towel. The cell pellet was resuspended in 250 µl Buffer A1, by pipetting up and down until no cell clumps remained and transferred into a 1.5 ml eppendorf tube. 250 µl Buffer A2 was added and the tube was gently inverted 6-8 times and incubated for 5 min at room temperature. When the lysate was clear 300 µl Buffer A3 was added and mixed thoroughly by inverting 6-8 times. The lysate was centrifuged at 11 000 x g for 5 min at room temperature, until a white pellet with cellular debris was formed. 750 µl of supernatant was pipetted into a NucleoSpin® Plasmid Column, placed in a collection tube

and centrifuged at 11 000 x g for 1 min. During centrifugation the plasmid DNA was absorbed onto the silica membrane. The flow-through was discarded. The silica membrane was washed by adding 600 µl Buffer A4 and centrifuged at 11 000 x g for 1 min. The flow-through was discarded and the silica membrane was dried by centrifugation at 11 000 x g for 3 min. The NucleoSpin® Plasmid Column was placed in a 1.5 ml eppendorf tube and the plasmid DNA was eluted by adding 50 µl Buffer AE, incubated for 1 min at room temperature and centrifuged at 11 000 x g for 1 min.

The quality and concentration of the purified DNA was measured by Nanodrop spectrophotometer (NanoDrop®, Saaven Werner AB). The transformation was confirmed by PCR reaction performed with plasmid DNA as a template and gene-specific primers (Table 1, Appendix A). The amplified products were confirmed by gel electrophoresis on 1 % agarose gels. Purified DNA was stored at -20 °C for further use in sequencing and WISH.

3.9 DNA sequencing analysis

DNA sequencing was performed on ABI 3730 DNA analyzer automated sequencer with Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

The sequencing kit is based on the Sanger dideoxy-mediated chain termination method. This method uses advantage of the ability of DNA polymerase to incorporate dideoxynucleotides (ddNTPs) as analogues of deoxynucleotides (dNTPs) into the growing chain. The ddNTPs lack a hydroxyl OH group at the 3' position. This position is normally where the free OH group in one nucleotide binds to the phosphate group of a new nucleotide in order to form a chain. When the polymerase at random inserts a ddNTP, with H-atom instead of an OH group at 3' position, additional nucleotides cannot be added. Thus, chain elongation is terminated selectively at A, C, G or T position.

The Ready Reaction Premix in the sequencing kit includes a mixture of dNTPs and fluorescent-labeled ddNTPs and AmpliTaq DNA Polymerase. Four different fluorescent dyes (yellow, red, blue, green) are used to label ddNTPs. Different dyes can be detected

simultaneously allowing a single-tube reaction for each primer, instead of four reactions. The ratio of dNTPs/ddNTPs is calculated so termination occurs at least once infinitely position in the template. In this way fragments with different length are produced and each fragment is terminated and labeled at a different base.

3.9.1 Cycle sequencing reaction

During cycle sequencing reaction cycles of thermal denaturation of template DNA, primer annealing and polymerization are repeated. Cycle sequencing uses only a single primer to increase the amount of sequencing products with ddNTP at the 3' position, exponentially with each repeated cycle.

The following sequencing mixture was prepared for each reaction:

5x Big Dye Terminator (v3.1) Sequencing Buffer	3.5 µl
Big Dye Terminator Ready Reaction Premix (v3.1)	1.0 µl
Primer SP6/T7 3.2 pMol/µl	1.0 µl
Template (150 ng)	0.5 µl
dH_2O	14.0 µl
Total volume	20.0 µl

SP6 and T7 primers, targeting the end of the vector were used when a template was obtained by cloning. For direct sequencing, when a template is purified directly from the PCR products, gene-specific primers were used as the sequencing primers (Table 1, Appendix A).

After short spin, the sequencing mixture was kept on ice protected with aluminum foil because the Reaction Premix is light sensitive.

The PCR program used was:

1. 96 °C for 1 min
2. 25 cycles: 95 °C for 10 sec, 50 °C for 5 sec and 60 °C for 4 min
3. 10 °C for infinite

Sequencing reaction products were stored at - 20 °C until they were purified and prepared for sequencing.

3.9.2 Sequencing reaction cleanup

Purification of sequencing reaction products was performed using a vacuum based filtration with Montage-SEQ$_{96}$ Cleanup Kit (Millipore). Unincorporated dye terminators, salts, enzymes, excess primers and primer dimers can interfere with the signal strength and reduce the sequencing quality. Therefore, they must be removed from the sequencing reaction prior to sequencing. Montage-SEQ$_{96}$ filter plate contains a membrane that retains DNA while contaminants are filtered to flow-through. The purified sequencing reaction products are resuspended and recovered from the membrane surface.

The sequencing mixture was transferred from - 20 °C and placed on ice until thawed. After a short spin, the mixture was pipetted into the Montage-SEQ $_{96}$ filter plate and 20 µl Injection Solution was added to each reaction. The filter plate was placed on the vacuum manifold, covered on the top with aluminum foil and vacuum filtered at 15 “Hg for 5 min or until wells were empty. The excess of fluid from the bottom of plate was removed with a paper towel. The same filtration procedure was repeated two more times, but with 25 µl Injection Solution. After adding 20 µl Injection Solution, the filter plate was placed on a micro-plate shaker and shaken at 1000 rpm for 10 min. To avoid absorption on the filter plate membrane, the resuspended and purified DNA was quickly transferred into a 96-well sequencing plate. Before sequencing was performed, the sequencing plates were centrifuged at 1500-2000 rpm for 1 min at room temperature.

3.9.3 DNA sequencing

DNA sequencing was performed on ABI 3730 DNA analyzer (Applied Biosystems). This is an automated sequencer based on capillary electrophoresis. Capillaries are filled with a polymer (POP-7™ Polymer) that separates DNA fragments. The fluorescently labeled DNA is loaded into capillaries by a short period of electrophoresis called electro-kinetic injection. As DNA fragments travel through the polymer during electrophoresis, they are separated by size. When the fragments reach detection point, a laser in the sequencing instrument excites dye molecules and causes them to fluoresce. In the Big Dye® Sequencing Kit each of four ddNTPs is labeled with two dyes. One of them is a fluorescein donor dye which is linked to one of four different dichlororhodamine acceptor dyes. The fluorescein donor dye collects energy from the laser light and transfers it to one of four acceptor dyes. Each acceptor dye emits light at characteristic wavelength. These fluorescence emissions are collected simultaneously and separated by a spectrograph. If energy transfer between the two dyes is more efficient, the brighter sequencing signals are and lower background noise is. In this way it is possible to identify the ddNTP that caused termination event.

The fluorescence data were read and interpreted with the Sequencing analysis software 5.2 (Applied Biosystems). The software Sequence Scanner™ v1.0 (Applied Biosystems) was used to display the raw view, analyzed view with chromatograms showing peaks in four different colors representing A, C, G and T and sequence nucleotide view. The obtained sequences of *Pan* I and *Sypl*2 were imported to the Vector NTI® software which allows assembly of overlapping sequences into contigs and comparison with the published sequences.

3.10. Real-time quantitative PCR (qPCR)

Real-time qPCR was used to quantify the expression of *Pan* I and *Sypl*2 in early life stages in NEAC and in various tissues in adult NEAC and NCC.

qPCR allows amplification and simultaneously quantification of a targeted DNA sequence. The amplified product is quantified in each PCR cycle, as the reaction progress in real time. A successful qPCR requires equal input amounts of DNA for all samples and technical duplicates or triplicates for each reaction on the plate. Variations in input amounts and efficiency of reverse transcription can be compensated with an endogenous ‘‘housekeeping’’ gene which is present at constant amounts in all samples. The products are quantified during the log phase of the reaction, as absolute number of copies or relative amount when normalized to DNA input or to endogenous genes. The real-time PCR instrument has a fluorescence detection threshold. The difference between amplification generated signal and background noise cannot be discriminated below this threshold. The cycle at which generated fluorescence is greater than the threshold is called the threshold cycle (Ct), which depends on the initial concentration of the target DNA sequence. If initial DNA concentration is higher, than the lower its Ct will be and vice versa. The Ct values are essential for real time PCR analysis in order to produce accurate and reproducible data.

The qPCR was performed in duplicates in 96-well optical plates on Light® Cycler 480 (Roche Diagnostics). The cDNA diluted 1:10 was used as a template in qPCR reaction and SYBR green was used as a fluorescent dye for the quantification of double stranded DNA. The forward and reverse qPCR primers were designed based on the *Pan* I (ATLCOD1Bc1522258, Figure A1 in Appendix A) and *Sypl*2 (ATLCOD1Bc1511891, Figure A2 in Appendix A) using the Vector NTI® software and synthesized by Invitrogen (Life Technologies) (Table 1, Appendix A). A 110 nt long fragment of *Pan* I gene was PCR amplified using the primers 5’-AGCCTCGTTGGTCCTCTA-3’ and 5’-CACTTGACATCAGTCAGACCTT -3’, while a 136 nt region of *Sypl*2 gene was amplified using the primers 5’-GGAGCACCCACTTTACGGTCAA-3’ and 5’-CCCCACGCTCACAAAGAACTC-3’. Expression of *ubiquitin* gene as a reference gene was

analyzed using gene-specific primer set (Table 1, Appendix A). The specificity of PCR amplification was confirmed by melting curve analysis (Tm calling, Light® Cycler 480, Roche Diagnostics). All primer pairs gave single distinctive melting peaks implying that no primer dimer and unspecific amplification products were present. Control samples with no cDNA template were included in qPCR analysis in order to test DNA contamination.

The following mixture was prepared for each reaction:

SYBR Green Master Mix	6.0 µl
Forward primer (10 µM)	0.5 µl
Reverse primer (10 µM)	0.5 µl
cDNA 10x	5.0 µl
Total volume	12.0 µl

Dilution series starting from 1X to 16X were prepared in triplicates in order to calculate standard curves and amplification efficiency for each primer set. The mix cDNA, containing cDNAs from all tissues included in the analysis was used as the template in dilution series.

Prepared mixtures were kept on ice. Prior to qPCR analysis, the 96-well plate was sealed with an optical foil and centrifuged at 2 000 rpm for 1 min. The following 3 step program used was:

1. Preincubation: 95 °C for 5 min
2. Amplification with 40 cycles of: 95 °C for 5 sec, 60 °C for 30 sec, 72 °C for 1 sec
3. Cooling: 40 °C for 30 sec

Cycle threshold (Ct) values and amplification efficiency values were calculated using Light® Cycler 480 Software release 1.5.0 (Roche Diagnostics) and the second derivate max method. Because qPCR was performed in duplicates, for each sample the mean of

duplicates and standard deviation were calculated. Samples with standard deviations > 0.5 were reanalyzed. All measured efficiencies were between 1.95 and 2.0. The Ct values of target and reference genes were exported to Microsoft Excel (Microsoft Corporation) and used to calculate differential gene expression by the comparative Ct method (ΔΔCt method). This method compares the number of cycles passed until the fluorescent signal of the target cDNA amplification reaches a certain threshold. The amount of target gene, normalized to the reference gene (*ubiquitin)* and relative to a calibrator (sample with the lowest gene expression) was calculated using the following formula:

$$2^{-\Delta\Delta Ct}$$

The first Δ ("delta") presents the difference between expression of target and reference gene (normalized expression). The second Δ describes the difference between sample and a calibrator (relative expression). The relative expression is presented as fold-difference in expression of target gene relative to the calibrator. However, statistical analyses and tests for significant differences were not possible to perform due to small number of tested individuals.

3.11. *In situ* hybridization (ISH)

ISH is a technique that allows precise localization of a specific segment of nucleic acid within tissue sections or whole mounts (W-ISH). It is based on hybridization of a complementary DNA or RNA probe to a target sequence.

Single-stranded RNA probes (riboprobes) have several advantages: they are very sensitive and RNA-RNA hybrids are more stable than DNA-RNA hybrids. Furthermore, RNA-RNA hybrids are resistant to degradation by RNase A so non-specific bindings can be removed after hybridization by RNase A. A DNA fragment to be transcribed is first amplified by PCR reaction and then cloned into the multiple cloning site of a transcription vector between the SP6 and T7 promoters for bacteriophage SP6 and T7 polymerase, respectively. After isolation and purification of the vector, the sequence of an insert is confirmed by

sequencing and the orientation of an insert relative to the promoter is verified. Depending on the orientation, the subsequently generated DNA template will produce sense or antisense strand RNA. For *in situ* hybridization complementary antisense transcripts are required. After sequence confirmation, the plasmid DNA is used as template in probe PCR with SP6 and T7 primer. In this way the DNA template, containing the correct RNA polymerase promoter in its sequence is generated and it is used to produce sense or antisense riboprobes. The riboprobes are synthetized by *in vitro* transcription during which the RNA polymerase first binds to its double-stranded DNA promoter and then separates the two DNA strands and uses the 3' - 5' strand as template for the synthesis of a complementary 5' - 3' RNA strand.

In vitro transcription is performed in the presence of a mixture of ribonucleotides and radioactive-, fluorescent- or antigen-labeled ribonucleotides. The ribonucleotides can be labeled with digoxigenin (DIG), a steroid isolated from digitalis plants (*Digitalis purpurea* and *Digitalis lanata*) that conjugates with uracil nucleotides. The DIG-labeled RNA probe is detected with high affinity anti-digoxigenin (anti-DIG) antibodies. Because DIG is synthetized only in the plants, the anti-DIG antibodies bind only to DIG and not to other biological substrates. This makes the DIG method very sensitive and efficient nonradioactive labeling and detection method. In order to visually detect the DIG-labeled ribobrobe hybridized to its target mRNA, the anti-DIG antibodies are conjugated to the enzyme alkaline phosphatase (AP). In the presence of the substrate nitro blue tetrazolium-/-5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP) AP gives a characteristic dark blue color precipitate.

3.11.1 Preparation of riboprobes for cod *Pan* I and *Sypl*2- Probe template PCR

A 511 nt long fragment of *Pan* I gene was PCR amplified using the primers 5’-CATTCATCCGAGTTCTGGAA-3’ and 5’- TGAGGTTGAGGAAGCCAA -3’, while the PCR amplification of a 547 nt region of *Sypl*2 gene was carried out using the primers 5’-GGTCCTTGTGGAAGTGCGTCT-3’ and 5’- GGTTCATCCGTGTCCTGGAA-3’. The amplified fragments were cloned into the pDrive Cloning Vector between SP6 and T7 promoter

required for *in vitro* transcription. After isolation, purification and sequence confirmation by sequencing, the plasmid DNA was used as a template in PCR reaction with SP6 and T7 primer to generate the DNA template with correct RNA polymerase promoter sequences.

Plasmid DNA diluted 1:10 was used as a template in PCR reaction. The following PCR mixture was prepared for each reaction:

5x Herculase® II Reaction Buffer	5.0 µl
2 mM dNTP	5.0 µl
Herculase® II Fusion DNA Polymerase	0.5 µl
SP6 primer (10 µM)	0.6 µl
T7 primer (10 µM)	0.6 µl
Plasmid DNA	1.0 µl
dH2O up to final volume	25.0 µl

Prepared PCR mixture was kept on ice. The PCR program used was:

1. 95 °C for 2 min
2. 5 cycles of: 95 °C for 30 seconds, 53 °C to 49 °C (-1 °C per cycle) for 30 seconds and 72 °C for 1 minute
3. 35 cycles of: 95 °C for 30 seconds, 49 °C for 30 seconds and 72 °C for 1 minute
4. 72 °C for 7 min
5. 4 °C for infinite

Amplified PCR products were visualized on a 1 % agarose gel. The PCR products that showed distinct bands were purified directly from the PCR mixture.

3.11.2 PCR product purification

PCR products were purified with Amicon® Ultra-0.5 Centrifugal Filter Units (Millipore). The purity of the DNA template is important for labeling efficiency during riboprobe synthesis.

50 µl PCR mixture and 250 µl DEPC H_2O were pipetted in an Amicon Ultra-0.5 filter device inserted into a micro-centrifuge tube. After centrifugation at 14 000 x g for 15 min the filtrate was discarded and the filter device was placed upside down in a clean tube and centrifuged at 1 000 x g for 2 min. In this way the concentrated DNA was transferred from the filter device into the tube. Purified DNA was stored at -20 °C for further use in riboprobe synthesis.

The quality and concentration of the purified DNA was measured by Nanodrop spectrophotometer (NanoDrop®, Saaven Werner AB).

3.11.3 Synthesis of digoxigenin (DIG) labeled riboprobe by *in vitro* transcription

Both sense and antisense DIG-labeled riboprobes were generated from the probe DNA template by *in vitro* transcription using bacteriophage SP6 and T7 RNA polymerase.

The following transcription mixture was prepared for each reaction:

10x Transcription Buffer	2.0 µl
2.5 mM DIG RNA Labeling Mix	2.0 µl
SP6/T7 RNA polymerase	2.0 µl
DNA template (300 ng)	x µl
DEPC dH_2O up to final volume	20.0 µl

Transcription mixture was incubated at 37 °C over night. After incubation 30 µl of DEPC dH_2O was added to each reaction. The synthetized RNA probe was visualized on a 1 % agarose gel. Appearance of a bit smeary RNA band with greater intensity than plasmid DNA

band indicated that RNA probe has been synthesized. The RNA probes were stored at -80 °C until precipitation was performed.

3.11.4 Purification of DIG-labeled riboprobe by ethanol precipitation

DIG-labeled RNA probes were purified from unincorporated ribonucleotides by precipitation with ethanol, lithium chloride (LiCl) and EDTA.

The RNA precipitates out of aqueous solutions by adding the ethanol and salt such as LiCl. The LiCl has an advantage of not precipitating carbohydrate, protein or DNA. The RNA precipitation is possible because salt neutralizes negatively charged phosphate (PO3-) groups along the sugar phosphate backbone, making RNA less hydrophilic. In addition, the ethanol has a much lower dielectric constant than water, causing the RNA to precipitates out of solution. The precipitated RNA can be separated from the rest of solution by centrifugation.

The precipitation mixture was prepared by adding 2 µl 4 M LiCl, 2 µl 0.5 M EDTA and 150 µl 96 % EtOH into 50 µl of the synthetized riboprobe. The mixture was kept at -20 °C over night. The precipitated RNA was pelleted by centrifugation at 13 000 rpm for 60 min at 4°C. The RNA pellet is very fragile and it may easily detach from the tube wall. Therefore, the supernatant was carefully removed while ensuring that the RNA pellet remained in the tube. The RNA pellet was washed by adding 500 µl 70 % DEPC EtOH in order to remove any residual LiCl. The 70 % DEPC EtOH was then removed by centrifugation at 13 000 rpm for 30 min at 4 °C. The supernatant was carefully removed and the pellet was resuspended in 1 ml hybridization buffer. After gently mixing, the precipitated DIG-labeled RNA probe was stored at -20 °C for further use in WISH.

3.11.5 Preparation of Atlantic cod embryos and larvae

Atlantic cod embryos of 60 d° and larvae of 90 d° were prepared for WISH analysis by removing the chorion and yolk. The embryos and larvae submerged in 100 % methanol (MeOH) were transferred from -20 °C and kept on ice. With a Pasteur pipette the embryos and larvae were transferred into a petri plate, filled with 100 % MeOH and observed with a

microscope (Leica Wild M3B) with 16X magnification. The chorion and yolk were removed very carefully with forceps, cleaned with RNase Away Reagent to prevent RNase contamination.

Prepared embryos and larvae (80 of each stage) were submerged in 100 % MeOH and stored at -20 °C for further use in WISH.

3.11.6 WISH analysis

3.11.6.1 Rehydration

Prepared embryos and larvae were transferred from 100 % MeOH into wells in a Nunclon™ Surface plate using a Pasteur pippete. The excess of MeOH was carefully removed and discarded, leaving the embryos and larvae intact in the wells. Then rehydration was performed by the successive washes as follows:

75 % MeOH / 25 % DEPC dH_2O	5 min
50 % MeOH / 50 % DEPC dH_2O	5 min
25 % MeOH / 75 % DEPC dH_2O	5 min
1 x DEPC PBST	5-10 min

Prior to each of the following steps, previously added solution was carefully removed and discarded without disrupting the embryos and larvae. After the last rehydration, the embryos were kept in 1 x DEPC PBST while larvae were prepared for bleaching by carefully removing 1 x DEPC PBST.

3.11.6.2 Bleaching

Bleaching of the larvae was required in order to remove the pigmentation that may interfere with staining and visualization of the hybridized DIG-labeled riboprobe. The following bleaching mixture was prepared:

Formamide	100.0 µl
30 % H_2O_2	660.0 µl
20 x SSC	50.0 µl
DEPC dH_2O	1.2 ml

0.5-1 ml of bleaching mixture was added into the well with larvae and exposed to UV light for 1-3 min or until pigmentation disappeared completely. After the bleaching, the larvae were rinsed with 1 x DEPC PBST for 3 x 10 min.

3.11.6.3 Permeabilization

1 x DEPC PBST was carefully removed from the wells and a mixture containing 1.5 ml 1 x DEPC PBST and 1.5 ml of Proteinase K (10 µg/ml) was added to both embryos and larvae. The embryos were incubated for 6 min while larvae were incubated for 9 min at room temperature. After incubation, the permeabilization mixture was replaced with 200 µl paraformaldehyde (PFA) and incubated for 20 min at room temperature. The PFA was carefully removed and the embryos and larvae were washed with 1 x DEPC PBST for 5 x 10 min.

3.11.6.4 Pre-hybridization

After permeabilization the embryos and larvae were transferred into wells for hybridization to sense (control) and antisense RNA probes in the Nunclon™ Surface plate. The excess of 1 x DEPC PBST was carefully removed and 0.5 ml of the hybridization buffer was added very slowly to each sample. Then the entire plate was covered with a lid and incubated at 65 °C for 3 hrs.

After transfer from -20 °C and thawing on the ice, the sense and antisense DIG-labeled RNA probes were denatured by heating at 70 °C for 1-5 min. 100 µl of denatured DIG-labeled riboprobe was added to 900 µl hybridization buffer in order to make a probe solution.

3.11.6.5 Hybridization

The hybridization buffer was carefully removed and replaced with 1 ml probe solution. After pipetting DEPC dH_2O between the wells to assure sufficient humidity during the incubation, the plate was covered with a lid and wrapped with a parafilm. After short incubation at 80 °C for 5 min the plate was incubated at 52 °C over night.

3.11.6.6 Post hybridization washing

After hybridization overnight, the probe solution was carefully removed and discarded. Fresh formamide solutions with pH=7-8 were prepared and incubated at 65 °C for 15-20 min. After incubation, the successive washes were performed as follows:

50 % formamide / 1 x SSC / drop of Tween20	2 x 30 min at 65 °C
25 % formamide / 0.5 x SSC /drop of Tween20	1 x 30 min at 65 °C
0.2 x SSC / drop of Tween20	2 x 30 min at 65 °C
1 x PBST	3 x 5 min at room temperature

3.11.6.7 Blocking

Blocking was performed to prevent unspecific antibody binding. Blocking solution was prepared as follows:

Blocking Reagent	300.0 mg
5 x MAB	3.0 ml
5 % inactivated Lamb serum	3.0 ml
dH_2O up to final volume	15.0 ml

The blocking solution was heated in a water bath at 60 °C until dissolved completely. After vigorous vortex, the blocking solution was cooled to 4 °C. After removing 1 x PBST, the embryos and larvae were rinsed with 1 x MAB for 5 min at room temperature. Then 1 x

MAB was replaced with 1 ml blocking solution and incubated for 1.5 hrs by shaking on a bench shaker.

3.11.6.8 Detection using BCIP/NBT

Blocking solution was carefully removed and replaced with 1 ml of a fresh blocking solution containing anti-digoxigenin-alkaline phosphatase conjugated Fab Fragments diluted 1:2000 (Roche Diagnostics). After overnight incubation at 4 °C with shaking on a bench shaker, the blocking solution with Fab Fragments was carefully removed and the successive washes were performed at room temperature as follows:

1 x MAB	5 x 30 min
Alkaline phosphatase buffer	1 x 5 min
Alkaline phosphatase buffer with 0.6 mg/ml levamisole	3 x 5 min

Alkaline phosphatase buffer contained: 5 ml 1M MTRIS HCL pH=9.5, 2.5 ml 1M MgCl2, 1 ml 5M NaCl, 100 µl Tween-20 and DEPC dH2O up to final volume 50 µl.

After last washing step, alkaline phosphatase buffer containing levamisole was replaced with 1 ml staining solution prepared by adding 20 µl NBT/BCIP (Roche Diagnostics) to fresh alkaline phosphatase buffer with 0.6 mg/ml levamisole. The samples were incubated for 15-30 min at room temperature. To detect staining pattern, the embryos and larvae were observed with a microscope (Leica Wild M3B) every 30 min. A clear staining was observed in the samples with the *Sypl*2 probe only, and further staining was stopped by removing the staining solution and fixating the embryos and larvae in glycerol.

3.12 Sequence analysis

Nucleotide and amino acid sequences of pantophysin, synaptophysin and synaptophysin-like genes in Atlantic cod and other vertebrates were identified using Ensembl at http://www.ensembl.org/ and NCBI BLAST (Basic Local Alignment Search Tool) at http://www.ncbi.nlm.nih.gov/. In addition, BLAST search through the reference cod genome based on the sequences from the North East Artic cod (NEAC) population (http://codgenome.no) and the GAFFA database (Genome Annotation Framework for Flexible Analysis) based on the Norwegian coastal cod (NCC) population (http://genofisk.cbu.uib.no/blast) was performed. The obtained sequences were aligned using the on-line algorithm ClustalW2 at the European Bioinformatics Institute (EBI) at http://www.ebi.ac.uk/Tools/msa/clustalw2/ applying the default values. The multiple alignments were exported to GenDoc software version 2.7 for further analysis. Identities and similarities were calculated using the on-line algorithm LALIGN at http://fasta.bioch.virginia.edu/fasta/lalign.htm.The sequences were screened for the conserved domains using NCBI BLAST server at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

A phylogenetic relationship between orthologous sequences in various vertebrate species was conducted using ClustalW in the software *Mega* version 5 (Tamura et al., 2011). The final alignment consisted of 45 amino acid sequences from 13 taxa and included 358 characters. These multiple alignments were used to construct a phylogenetic tree using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances between amino acid sequences were calculated using the p-distance method that calculates the proportion of amino acid differences between each pair of sequences. All positions containing gaps and missing data were removed by complete deletion. Tree topological stability and reliability was evaluated by the bootstrap test that uses the bootstrap re-sampling strategy. The number of replicates was 10 000 and the percentage of replicate trees in which associated taxa clustered together was shown next to the branches. Branches reproduced in less than 65 % bootstrap replicates were collapsed. All options

were provided by the software. The *Sypl* sequence of the tunicate *Ciona intestinalis* was used as outgroup to root the tree.

The synteny analysis was performed by searching for genes flanking cod *Pan* I, *Sypl*2 and *Syp*1 and the orthologous genes in medaka, zebrafish, *Xenopus* and human in Ensembl at http://www.ensembl.org/.

4. Results

4.1 Sequence analysis

Ensembl searching for pantophysin, synaptophysin and synaptophysin-like genes in the genome of Atlantic cod resulted in five distinct genes designated pantophysin (*Pan* ENSGMOG00000001154 at scaffold 1597:87015-88958), synaptophysin1 (*Syp*1 ENSGMOG00000019650 at scaffold 1609:446,748-456,190), synaptophysin2 (*Syp*2 ENSGMOG00000018842 at scaffold 98:292-12,350), synaptophysin-like1 (*Sypl*1 ENSGMOG00000011737 at scaffold 2630:36309-43358) and synaptophysin-like2 (*Sypl*2 ENSGMOG00000010980 at scaffold 2301:77,045-89,991). Cod *Pan* consists of five exons and the predicted transcript of 705 nucleotides (nt) encodes a protein of 234 amino acids (aa). However, the Celera assembly of the Atlantic cod genome revealed the sequence ATLCOD1Bc1522258 with additional nucleotides ATGTATATCTTTCATCATCATCAG in the first exon coding for the MYIFHHHQ aa (Figure A1 in Appendix A).

The cod *Sypl*2 gene predicted in Ensembl consists of eight exons, with transcript length of 657 nt and translation length of 218 aa. The coding sequences contain a 12 nt region TTCAG**GG**TGGAA with a single GG insertion (bold) at 204-205 nucleotide position, coding for the FRVE aa (Figure A3 in Appendix A). However, ViroBLAST search through Celera assembly in the cod genome resulted in the genomic sequence ATLCOD1Bc1511891. The exon-intron boundaries were identified using the software GenScan and the coding sequence with six exons was identified. The predicted sequence of 705 nt contained a 47 nt long insertion instead of the GG insertion (Figure A2 in Appendix A). A 179 nt long fragment including the 47 nt insertion was PCR amplified using the primers *Sypl*2s2-*Sypl*2a and *Sypl*2ins-as-Sypl2a (Table 1 Appendix A) in both NEAC and NCC (Figure 6). The sequence of amplified products, including the 47 nt insertion was confirmed by sequencing.

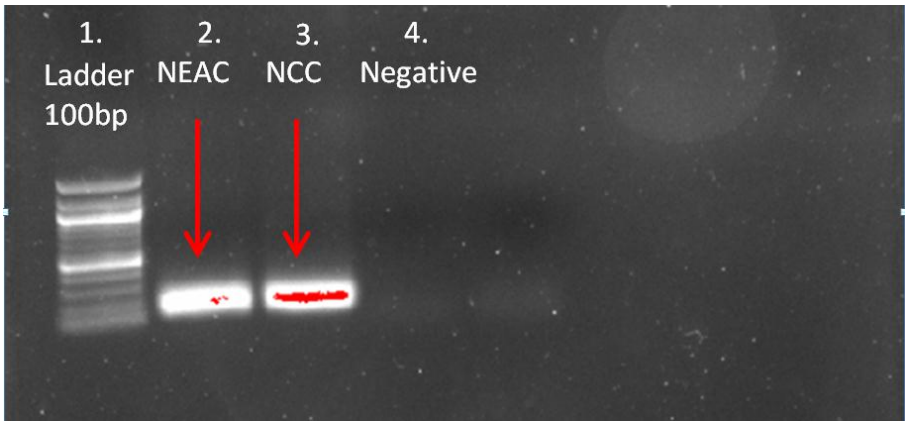
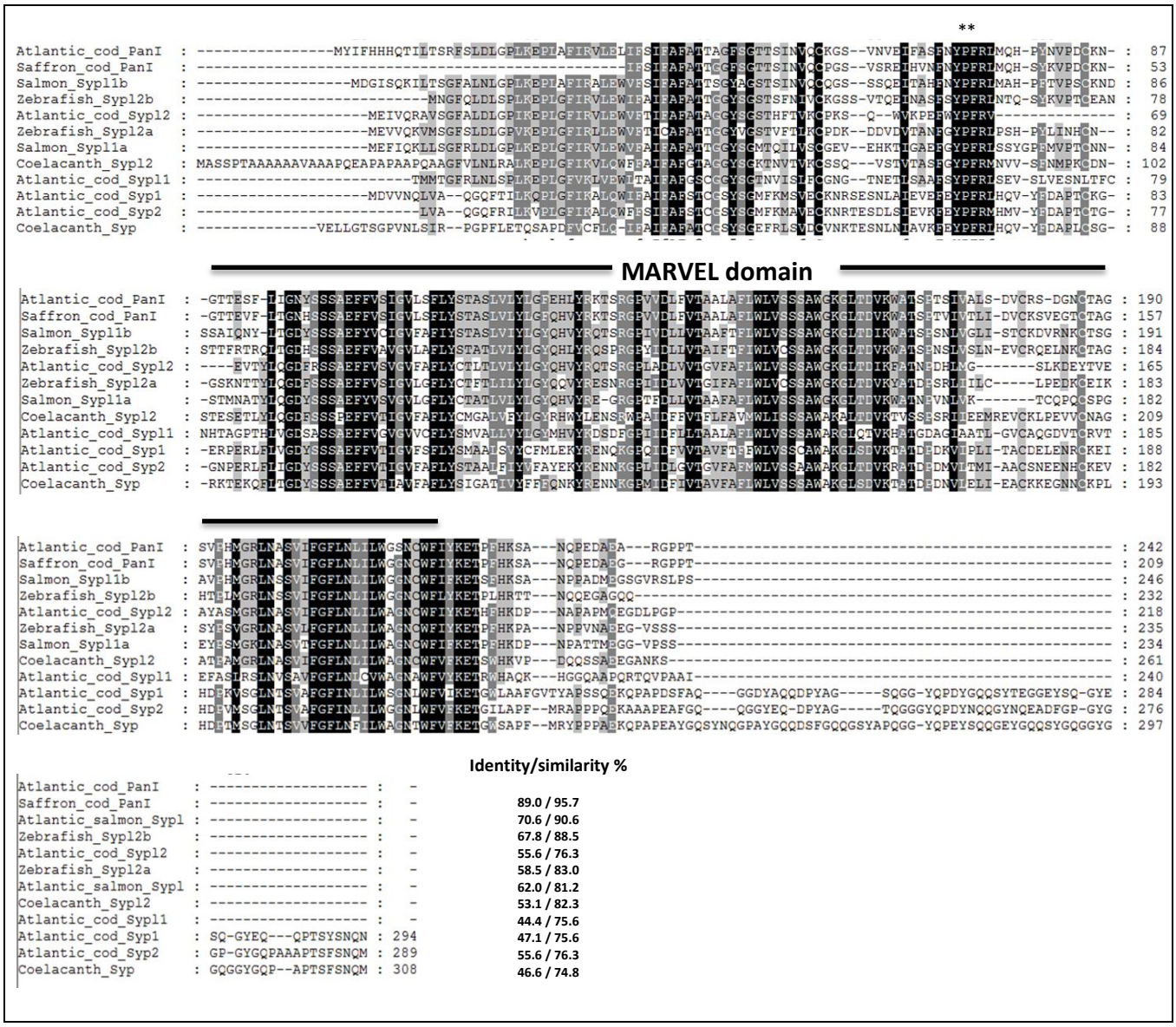


Figure 6 PCR analysis of cod Sypl2 expression in North East Artic cod (NEAC, lane 2) and Norwegian coastal cod (NCC, lane 3). 100 bp DNA ladder (lane 1) and negative control with no template (lane 4) are shown.

4.2 Multiple alignment

The multiple alignment of predicted pantophysin, synaptophysin and synaptophysin-like amino acid sequences showed that all examined teleosts contain a highly conserved MARVEL domain and block of five amino acids (YPFRL). The sequence identity between Atlantic cod *Pan* I and other examined teleosts was 44.4 - 89.0 % (Figure 7).

Figure 7 see next page: Multiple alignment of pantophysin, synaptophysin and synaptophysin-like amino acid sequences from diverse teleosts. Hyphens (-) are inserted to optimize the alignment. The MARVEL domain is indicated and the block of YPFRL amino acids is marked (asterisks). Sequence identities and similarities (%) relative to Atlantic cod Pan I are presented. The degree of conservation in each column in the alignment is marked with different shade grading. Black color corresponds to 100 percent conserved, dark grey to 80 percent and light gray to 60 percent conserved. Ensembl accession numbers from the top down are: Atlantic cod Pan I (ENSGMOP00000001198), Saffron cod Pan I (AAQ62764.1), Atlantic salmon Sypl1b (ACI32929.1), zebrafish Sypl2b (ENSDARP00000001793), Atlantic cod Sypl2 (ENSGMOP00000011744), zebrafish Sypl2a (ENSDARP00000115067), Atlantic salmon Sypl1a (ACI33252.1), Coelacanth Sypl2 (ENSLACG00000016146), Atlantic cod Sypl1 (ENSGMOP00000012549), Atlantic cod Syp1 (ENSGMOP00000021137), Atlantic cod Syp2 (ENSGMOP00000020229), Coelacanth Syp (ENSLACP00000004791).


MARVEL domain
Identity/similarity %
89.0 / 95.7
70.6 / 90.6
67.8 / 88.5
55.6 / 76.3
58.5 / 83.0
62.0 / 81.2
53.1 / 82.3
44.4 / 75.6
47.1 / 75.6
55.6 / 76.3
46.6 / 74.8

4.3 Phylogenetic tree

The phylogenetic tree was constructed using identified pantophysin, synaptophysin and synaptophysin-like amino acid sequences from various vertebrate species. The urochordata *Ciona intestinalis* was used to root the tree (Figure 8). Phylogenetic analysis showed that a number of examined vertebrates classified into clade of teleosts and tetrapods, segregating with high bootstrap confidence values. However, the teleosts synaptophysins (*Syp*) clustered with the tetrapod orthologs with high bootstrap value and suggests that this protein has been evolutionary conserved. Teleosts with synaptophysin-like1 (*Sypl*1) sequences clustered together with 100 percent bootstrap value and formed a separate clade. The teleosts clade, separated from tetrapods was furthermore clustered into two clusters. *Pan* I clustered together with *Sypl*1b and *Sypl*2b, while *Sypl1 and Sypl*2 together with *Sypl*1a and *Sypl*2a formed a second cluster with high bootstrap value. The gadoids were closely grouped together within the first cluster, with Atlantic cod segregating with its sister taxon Alaska pollock. The *Gadiformes* shared the node with Atlantic salmon (*Salmoniformes*), zebrafish (*Cypriniformes*), medaka (*Beloniformes*), tilapia (*Perciformes*), stickleback (*Gasterosteiformes*) and fugu (*Tetradontiformes*). The branching within the second cluster followed similar pattern, with *Salmoniformes* branching off together with the *Esociformes.* Tetrapods with *Sypl*1 clustered together with high bootstrap value, except *Xenopus Sypl*1 that formed a separate branch. The clade of tetrapods with *Sypl*2 followed the conventional phylogeny and gathered amphibians, mammals and a bird with 100 percent bootstrap confidence value.

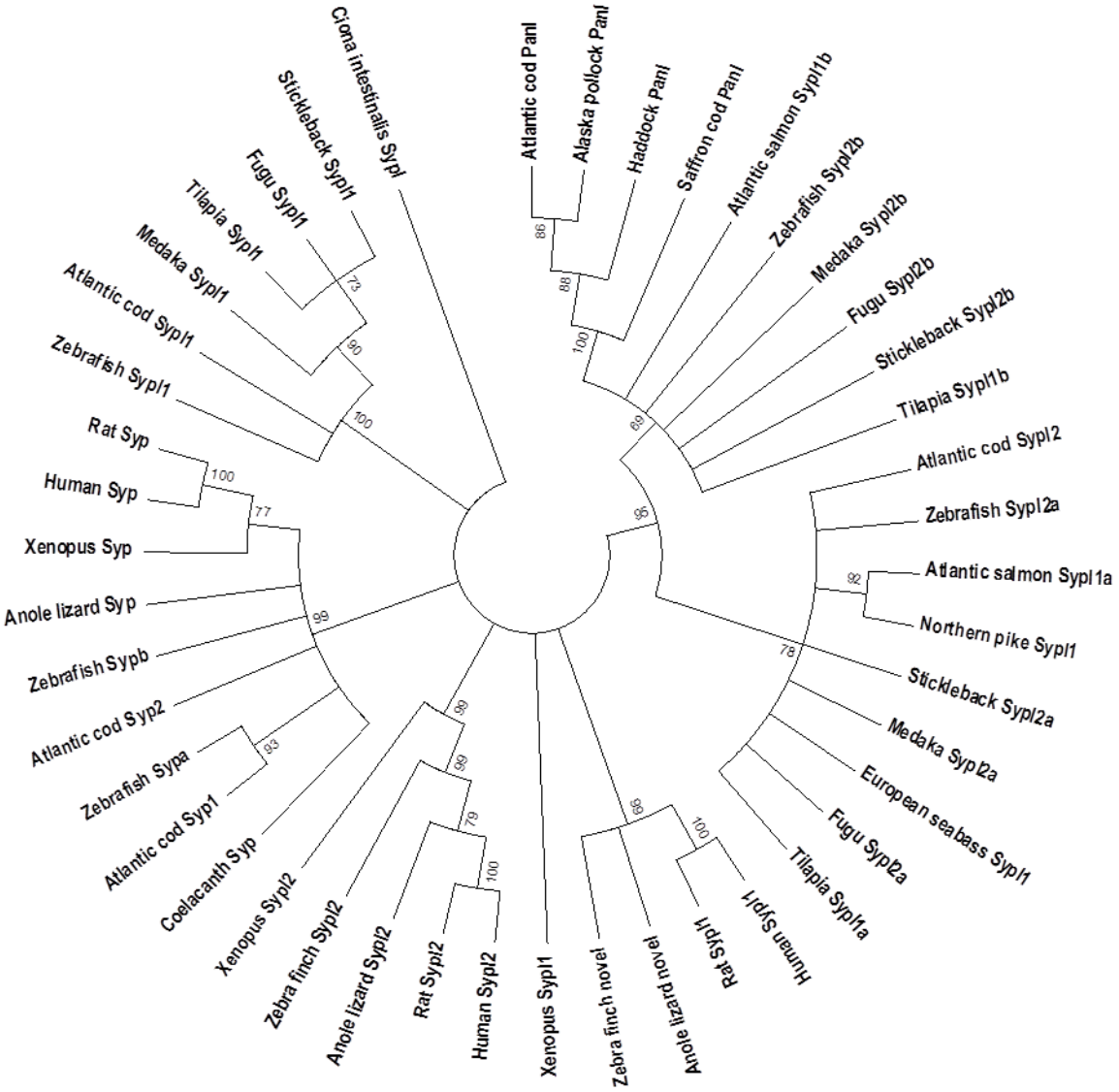


Figure 8 Phylogenetic tree showing the evolutionary relationship between pantophysin, synaptophysin and synaptophysin-like sequences in various vertebrate species. The tree was constructed using the Neighbor-Joining method in MEGA5. The evolutionary distances were calculated using the p-distance method and all gaps were removed by complete deletion. The bootstrap confidence values giving the percent frequency each clade appears in 10 000 replicates are shown next to the branches. Branches reproduced in less than 65 percent bootstrap replicates were collapsed. Ensembl and GeneBank accession numbers are: Atlantic cod Pan I (ENSGMOP00000001198), Alaska pollock Pan I (AAQ62772.1), Haddock Pan I (AAQ62771.1), Saffron cod Pan I (AAQ62764.1), Atlantic salmon Sypl1b (ACI32929.1), zebrafish Sypl2b (ENSDARP00000001793), medaka Sypl2b (ENSORLP00000007593), tilapia Sypl1b (XP_003444868.1), stickleback Sypl2b (ENSGACP00000010388), Fugu Sypl2b (ENSTRUP00000009992), Atlantic cod Sypl2 (ENSGMOP00000011744),

zebrafish Sypl2a (ENSDARP00000115067), Atlantic salmon Sypl1a (ACI33252.1), Northern pike Sypl1 (ACO14231.1), stickleback Sypl2a (ENSGACP00000015267), medaka Sypl2a (ENSORLP00000013031), European seabass Sypl1 (CBN80944.1), Fugu Syp2a (ENSTRUP00000000568), tilapia Sypl1a (XP_003459384.1), Human Sypl1 (ENST00000011473), Rat Sypl1 (ENSRNOP00000013091), Anole lizard novel gene (ENSACAP00000016430), Zebra finch novel gene (ENSTGUP00000003131), Xenopus Sypl1 (ENSXETP00000054574), human Sypl2 (ENSP00000358888), rat Sypl2 (ENSRNOP00000026883), Anole lizard Sypl2 (ENSACAP00000010685), zebra finch Sypl2 (ENSTGUP00000014338), xenopus Sypl2 (ENSXETP00000025929), Atlantic cod Syp2 (ENSGMOP00000020229), Coelacanth Syp (ENSLACP00000004791), Atlantic cod Syp1 (ENSGMOP00000021137), zebrafish Sypa (ENSDARP00000063095), zebrafish Sypb (ENSDARP00000082673), Anole lizard Syp (ENSACAP00000012707), xenopus Syp (ENSXETP00000016833), human Syp (ENSP00000263233), rat Syp (ENSRNOP00000013724), zebrafish Sypl1 (ENSDARP00000098590), Atlantic cod Sypl1 (ENSGMOP00000012549), medaka Sypl1 (ENSORLP00000016550), tilapia Sypl1 (ENSONIP00000021266), Fugu Sypl1 (ENSTRUP00000042689), stickleback Sypl1 (ENSGACP00000024908) and Ciona intestinalis Sypl (ENSCINP000000336878) used as outgrup to root the tree.

4.4 Synteny analysis of pantophysin, synaptophysin and synaptophysin-like genes

In silico analysis showed that cod *Pan* I, *Sypl*2 and *Syp*1 and orthologous genes are flanked by segments that show highly conserved synteny in the genome of teleosts, *Xenopus* and human (Figure 9).

Cod scaf 1597 GPR61 2 ATXN7L2 2 **Pan I** SORT1 1 NGRN NFYA AHCYL1 1 KCNC4	**Medaka chr 7** GPR61 1 **Sypl2b** SORT1 1 NGRN NFYA AHCYL1 2 KCNC4	**Zebrafish chr 8** KCNC4 AHCYL1 NFYAL SORT1b **Sypl2b** ATXN7L2b	**Xenopus scaf GL172681.1** ATXN7L1 **Sypl1**	**Human chr 7** ATXN7L1 **Sypl1**
Cod scaf 2301 **Sypl2** ATXN7L2 1 CYB561D1 ZBTB40 1 ZBTB40 2 USP48	**Medaka chr 5** USP48 ZBTB40 1 ZBTB40 2 CYB561D1 ATXN7L2 **Sypl2a**	**Zebrafish chr 11** SORT1a **Sypl2a** ATXN7L2 CYB561D1 ZBTB40 USP48	**Xenopus scaf GL172677.1** NFYA SORT1 **Sypl2**	**Human chr 1** **Sypl2** ATXN7L2 CYB561D1 GPR61 SORT1
Cod scaf 1609 UCKL1 EMILIN3 PRICKLE3 PLP2 **Syp1** PPP1R3A	**Zebrafish chr 8** **Sypa** PLP2 2 EMILIN3a UCKL1	**Medaka chr 7** **Syp1** PLP2 PRICKLE3 EMILIN3 UCKL1	**Xenopus scaf GL172790.1** **Syp** PRICKLE3 PLP2	**Human chr X** PLP2 PRICKLE3 **Syp** PPP1R3A

Figure 9 Conserved syntenic regions associated with pantophysin (grey), synaptophysin (yellow) and synaptophysin-like (rose) genes in Atlantic cod, medaka, zebrafish, Xenopus and human. Chromosomal and scaffold positions of cod Pan I, Sypl2 and Syp1 orthologous genes (marked bold) and flanking genes were identified in Ensembl. Full length names of the genes are given in the Table 2 (Appendix B).

4.5 Real-time quantitative PCR (qPCR) analysis

4.5.1 *Pan* I and *Sypl*2 expression during early life stages in NEAC

Embryonic and larval cDNAs were used in qPCR analysis in order to quantify mRNA levels of *Pan* I and *Sypl*2 during early life stages in NEAC (Figure 10).

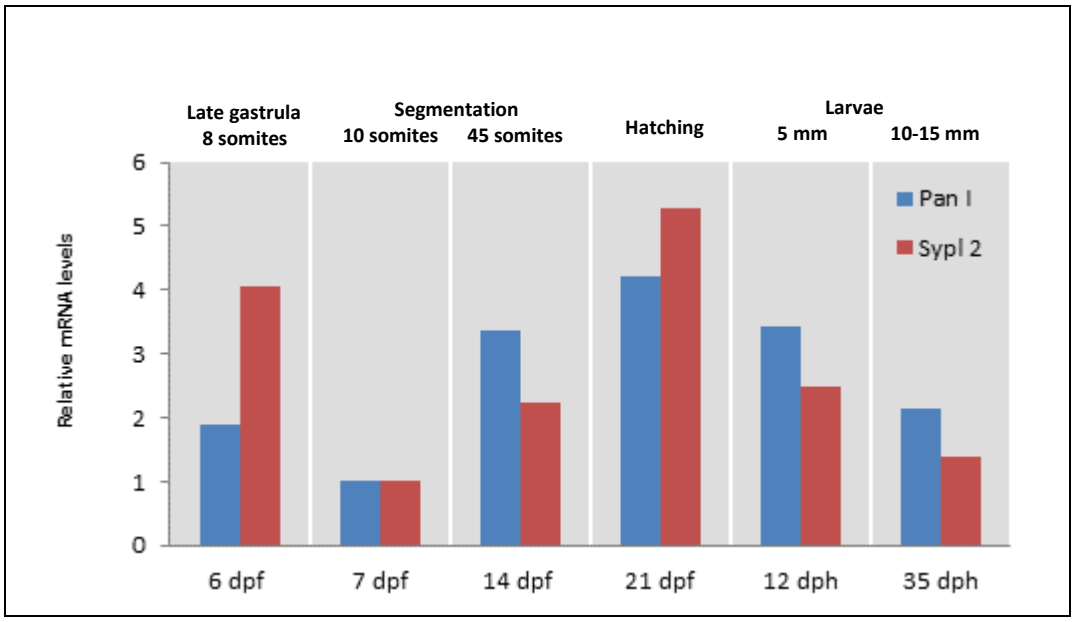


Figure 10 Real-time quantitative PCR analysis of Pan I and Sypl2 genes during early life stages in North East Artic cod (NEAC). The timeline includes 6 days post fertilization (dpf), 7 dpf, 14 dpf, 21 dpf, 12 days post hatching (dph) and 35 dph. Both target genes were normalized to ubiquitin. Relative mRNA expression was calculated using the comparative Ct method and represents fold change compared to a calibrator sample (sample with the lowest gene expression). Each bar represents the relative expression calculated in a sample based on cDNAs pooled from 10 individuals.

The expression of *Pan* I and *Sypl*2 was examined from late gastrula with 8 somites (6 dpf) until the larval size of 10-15 mm body length (35 dph). Both target genes were expressed in all examined stages, showing very similar expression patterns. The highest *Pan* I and *Sypl*2 expression was detected at hatching (21 dpf) while the lowest expression was at 10-somite stage (7dpf). After reaching a peak, the expression decreased until the stage of 35 dph. In comparison to *Pan* I, the *Sypl*2 mRNA levels were 2.1 times higher at 6 dpf.

4.5.2 *Pan* I and *Sypl*2 expression in adult NEAC and NCC

The cDNAs from various tissues in adult NEAC and NCC were used for qPCR analysis. The expression of *Pan* I and *Sypl*2 was examined in eight tissues in the female NEAC1 (Figure 11A), nine tissues in the male NEAC2 (Figure 11B), ten tissues in the male NCC1 (Figure 11C) and eleven tissues in two females NCC2 (Figure 11D) and NCC3 (Figure 11E). The expression of *Sypl*2 was quantified in the same individuals and tissues, except in the NEAC2 brain where *Sypl*2 expression was not analyzed.

The qPCR analysis revealed that both genes were transcriptionally active in all examined tissues. The *Pan* I transcript was detected at the highest level in the head kidney (NEAC1 and NCC3), spleen (NEAC2), epithelial mucus (NCC1) and ovaries (NCC2). The highest *Sypl*2 mRNA levels were detected in the muscle (NEAC1 and NEAC2), testicles (NCC1) and epithelial mucus (NCC2 and NCC3). Interestingly, when *Pan* I was compared to the *Sypl*2 expression between NEAC1 and NEAC2 an inverse expression pattern was revealed. The *Sypl*2 peak levels observed in the muscle were opposite to the *Pan* I levels. Furthermore, the head kidney and spleen showed the highest *Pan* I expression, while exhibiting the lowest *Sypl*2 expression. Similar pattern was observed in the head kidney in NCC3 where the *Sypl*2 mRNA level was 2.7 times lower than the *Pan* I expression, highest in the same tissue. In NCC1 and NCC2 both genes showed similar expression patterns with the highest gonadal and epithelial expression. However, compared to *Pan* I the *Sypl*2 expression was higher in these tissues in all three NCC individuals.

Pan I and *Sypl*2 exhibited individual and tissue-specific expression differences. The *Pan* I mRNA levels were higher in NEAC2 compared to NEAC1 in almost all examined tissues, especially in the liver, spleen and heart. The *Sypl*2 tissue-specific differences between NEA1 and NEA2 were less variable than *Pan* I, with the highest difference observed in the liver. Both *Pan* I and *Sypl*2 mRNA levels were higher in gonads and epithelial mucus in NCC1 compared to the same tissues in NCC2 and NCC3.

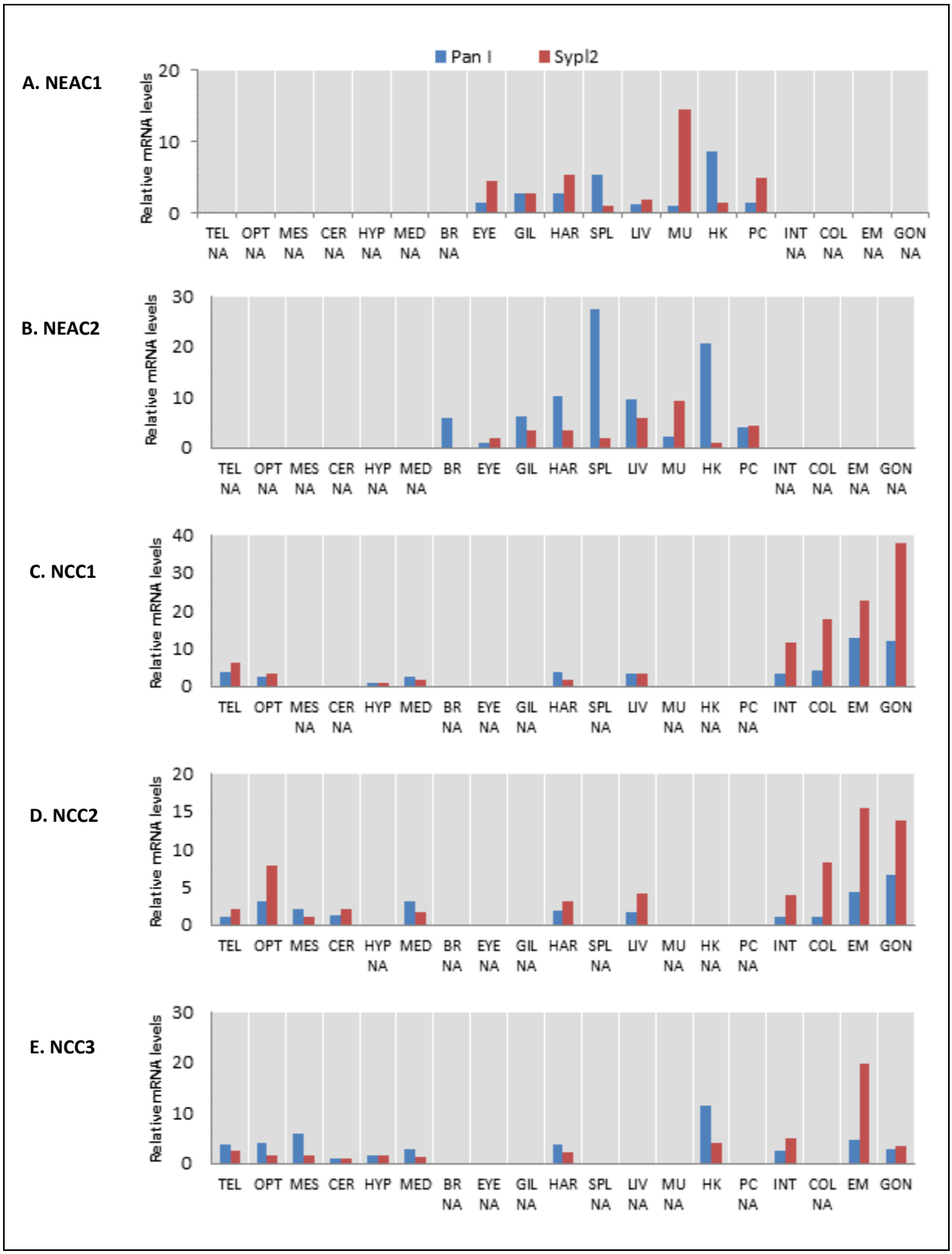

A. NEAC1
Pan I
Sypl2
Relative mRNA levels
20
10
0
TEL NA
OPT NA
MES NA
CER NA
HYP NA
MED NA
BR NA
EYE
GIL
HAR
SPL
LIV
MU
HK
PC
INT NA
COL NA
EM NA
GON NA
B. NEAC2
Relative mRNA levels
30
20
10
0
TEL NA
OPT NA
MES NA
CER NA
HYP NA
MED NA
BR
EYE
GIL
HAR
SPL
LIV
MU
HK
PC
INT NA
COL NA
EM NA
GON NA
C. NCC1
Relative mRNA levels
40
30
20
10
0
TEL
OPT
MES NA
CER NA
HYP
MED
BR NA
EYE NA
GIL NA
HAR
SPL NA
LIV
MU NA
HK NA
PC NA
INT
COL
EM
GON
D. NCC2
Relative mRNA levels
20
15
10
5
0
TEL
OPT
MES
CER
HYP NA
MED
BR NA
EYE NA
GIL NA
HAR
SPL NA
LIV
MU NA
HK NA
PC NA
INT
COL
EM
GON
E. NCC3
Relative mRNA levels
30
20
10
0
TEL
OPT
MES
CER
HYP
MED
BR NA
EYE NA
GIL NA
HAR
SPL NA
LIV NA
MU NA
HK
PC NA
INT
COL NA
EM
GON

Figure 11 see previous page: Real-time quantitative PCR analysis of Pan I and Sypl2 genes in various tissues in NEAC1(A), NEAC2 (B), NCC1(C), NCC2(D) and NCC3 (E). The following tissues are examined: telencephalon (TEL), optic tectum (OPT), mesencephalon (MES), cerebellum (CER), hypophysis (HYP), medulla oblongata (MED), brain (BR), eye (EYE), gills (GIL), heart (HAR), spleen (SPL), liver (LIV), muscle (MU), head kidney (HK), pyloric ceca (PC), anterior part of the intestines (INT) and distal part or colon (COL), epithelial mucus (EM) and gonads (GON). Both target genes were normalized to ubiquitin. Relative mRNA expression was calculated using the comparative Ct method and represents fold change compared to a calibrator (sample with the lowest gene expression). The calculated values are given in Table 3 and 4 (Appendix C). Absence of bars marked with NA represents samples that are not analyzed.

4.6 Whole mount *in situ* hybridization (WISH) analysis of cod *Sypl*2

WISH analysis of the *Sypl*2 expression was carried out on Atlantic cod embryos of 60 d° and larvae of 90 d° using synthetized antisense and sense RNA riboprobe (Figure 12).

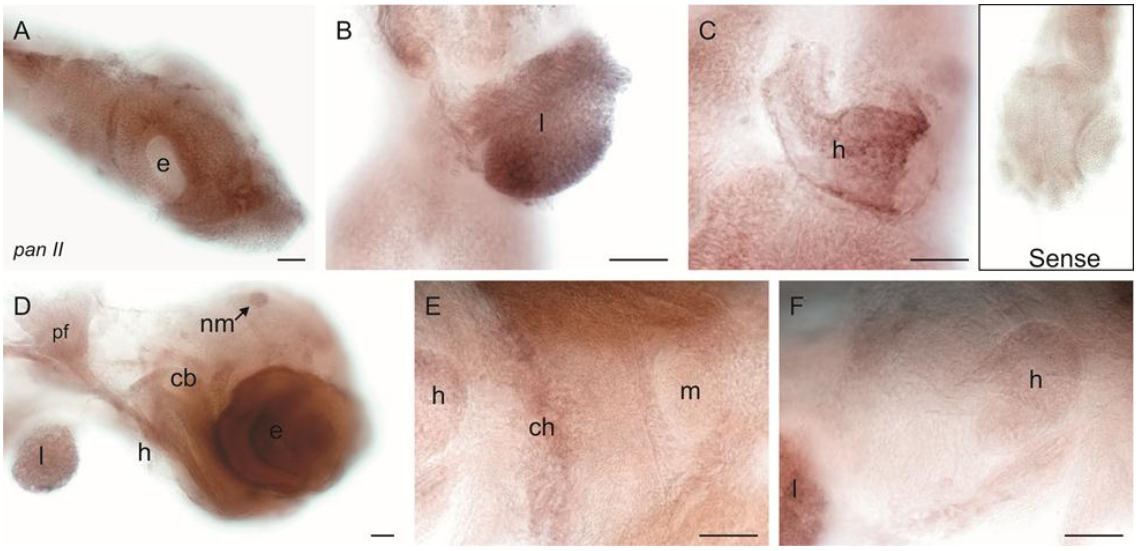


Figure 12 WISH analysis of Sypl2 expression in Atlantic cod embryos of 60 d° (A-C) and larvae of 90 d° (D-F). A) Staining in the embryonic head, lateral view. B) Staining in the embryonic liver (l), ventral view. C) Staining in the embryonic heart (h), ventral view. D) Staining in the larval eye (e), neuromast (nm), ceratobranchial arches (cb), heart (h), liver (l) and pectoral fin (pf), lateral view. E) Staining in the larval heart (h), ceratohyal arch (ch) and mandible (m), ventral view. F) Staining in the larval heart (h) and liver (l), ventral view. Sense control with no staining. The scale bar correspondents to 50 µm.

WISH revealed that *Sypl*2 mRNA was identified in the embryonic head but no staining was visible in the eye. The *Sypl*2 signal was detected in the larval eye even though the visualization of the signal was somewhat masked by the eye's pigmentation. Furthermore, the *Sypl*2 expression was identified in the embryonic and larval liver and heart as well as in the larval neuromast, ceratohyal and ceratobranchial arches, arch, mandible and pectoral fin.

5. Discussion

In silico analyses of the Atlantic cod genome identified five distinct genes in synaptophysin family, designated pantophysin (*Pan*), synaptophysin1 (*Syp*1), synaptophysin2 (*Syp*2), synaptophysin-like1 (*Sypl*1) and synaptophysin-like2 (*Sypl*2). Cod *Pan* predicted in Ensembl consists of five exons and the transcript of 705 nt encodes a protein of 234 aa. The *Pan* I sequence (ATLCOD1Bc1522258) found in the Celera assembly of the Atlantic cod genome has transcript length of 729 nt, with additional nucleotides in the first exon and encodes a protein of 342 aa (Figure A1 in Appendix A). A 511 nt long fragment of this sequence was PCR amplified using the primers 5'-CATTCATCCGAGTTCTGGAA-3' and 5'-TGAGGTTGAGGAAGCCAA -3' (Table 1 Appendix A). The cod *Sypl*2 gene predicted in Ensembl consists of eight exons, with transcript length of 657 nt and translation length of 218 aa (Figure A3 in Appendix A). The *Sypl*2 sequence (ATLCOD1Bc1511891) found in the Celera assembly consists of six exons with 705 nt transcript length (Figure A2 in Appendix A). A 547 nt region of this sequence was PCR amplified using the primers 5'-GGTCCTTGTGGAAGTGCGTCT-3' and 5'- GGTTCATCCGTGTCCTGGAA-3' (Table 1 Appendinx A). The amplified fragments were cloned and sequences were confirmed by sequencing. Even though the presence of the 47 nt long insertion in the *Sypl*2 sequence (ATLCOD1Bc1511891) was confirmed in both NEAC and NCC, one should not preclude that different transcripts produced by alternative splicing may be present in both populations.

An alignment of the identified cod amino acid sequences and ortologs in various teleosts indicated a high level of evolutionary conservation. All examined proteins exhibited the characteristic MARVEL domain and the block of YPFRL aa, flanked with conserved phenylalanines on both sides (Figure 7). The structure with four transmembrane hydrophobic domains characteristic for the MARVEL domain, two intravesicular loops and two cytoplasmic tails, amino- and carboxy–terminal tail is shared by all members of synaptophysin gene family (Sanchez-Pulido et al., 2002; Pogson and Mesa, 2004; Arthur and Stowell, 2007; Estrada et al., 2007). Previous studies showed that this structure is

conserved across mammalian species (Leube, 1994), indicating its essential role for the protein basic properties. The YPFRL aa block located in the first intravesicular loop (IV1) is conserved among gadoid species, human and mouse physins (Leube, 1994; Pogson, 2001; Pogson and Mesa, 2004). Furthermore, all amino acid replacement mutations detected at *Pan* I locus were located in the IV1, suggesting that it might have the essential significance for the protein function (Pogson, 2001). The selection detected in IV1 might be driven by many different factors, such as physicochemical characteristics of cargo in the pantophysin vesicles or interactions with other vesicle membrane proteins. In addition, all pantophysisns contain potential N-glycolysation sites in the intravesicular loops. Many of surface glycoproteins serve as the receptors for cell invasion, being targeted by various pathogens. Thus, pathogen attack might be a reason for the selection observed in the intravesicular loops. However, it is possible that selection is unrelated to typical function of pantophysisn or that it acts on genes associated to *Pan* I (Pogson and Mesa, 2004).

Calculated sequence identities showed that Atlantic cod *Pan* I and orthologous sequences in Saffron cod (*Pan* I), Salmon (*Sypl1*b) and zebrafish (*Sypl*2b) exhibited a high identity level of 67.8-89.0 %. However, the identity between Atlantic cod *Pan* I and the rest of proteins examined was only 44.4-62.0 % (Figure 7). The lower homology is most probably due to lacking of the long carboxy–terminal tail in pantophysins, with the proline- and tyrosine-repeated motifs characteristic for synaptophysins (Leube, 1994; Pogson, 2001).

Phylogenetic analysis confirmed the evolutionary conservation between pantophysin, synaptophysin and synaptophysin-like proteins in various vertebrates (Figure 8). The tree was constructed using Ensembl predicted amino acid sequences with *Ciona intestinalis* as an outgroup. The mammalian *Pan* I sequences such as human and rat pantophysin, previously presented in this study (Figure 2) were not included in the phylogenetic analysis. The reason is that these sequences are not annotated in Ensembl. Instead, tetrapods *Pan* I othologs marked with *Sypl*1 were included in the tree construction. The teleosts *Syp* clustered together with tetrapod orthologs, indicating a common origin. Tetrapods with *Sypl*1 clustered together, except *Xenopus Sypl*1 that formed a separate branch. Tetrapods

*Sypl*2 and teleosts *Sypl*1 and formed two separate clades, each clustering with 100 % bootstrap confidence value. In the teleost clade, separated from tetrapods, two clusters were formed. Various taxa from the class *Actinopterygii* with *Pan* I clustered together with *Sypl*1b and *Sypl*2b, while *Sypl*1 and *Sypl*2 clustered together with *Sypl*1a and *Sypl*2a. These two clusters may present duplicated copies derived from third round of whole genome duplication (3R WGD), which occurred in teleosts after their divergence from tetrapods (Sato et al., 2009).

Syntenic analysis revealed that Atlantic cod *Pan* I, *Sypl*2 and *Syp*1 and orthologous genes in medaka, zebrafish, *Xenopus* and human were flanked by highly conserved segments, supporting the common origin of vertebrates (Figure 9).

When Pogson et al. (1995) discovered an anonymous cDNA clone (GM798) it was first identified as cod synaptophysin (*Syp* I) (Fevolden and Pogson, 1997). Studies by Leube (1994) and Haass et al. (1996) showed that cDNA clone more likely represent a cellular isoform of synaptophysin, named pantophysin. However, phylogenetic and synteny analyses in our study revealed incorrect sequence naming in several species and in order to avoid misunderstandings the synonymous names should be given. The NCBI BLAST search confirmed that human pantophysin is synonymous to human *Sypl*1 on chromosome 7. (Figure 9). Therefore, the correct name for human *Pan* I should be *Sypl*1. Similar, Anole lizard and zebra finch novel genes should be marked with *Sypl*1. According to the phylogenetic analysis teleosts with *Pan* I and orthologous *Sypl1b* should be renamed into *Sypl*2b. Teleosts including European seabass *Sypl*1, Northern pike *Sypl*1, Atlantic cod *Sypl*2, Coelacanth *Sypl*2, Atlantic salmon *Sypl*1a and Tilapia *Sypl*1a should be marked with *Sypl*2a. Furthermore, teleost and tetrapods *Syp* should be renamed into *Syp*1 and/or *Syp*2.

Gene expression analyses showed that *Pan* I and *Sypl*2 were expressed in early life stages in NEAC, from late gastrula with 8 somites (6 dpf) until the larval size of 10-15 mm body length (35dpf). Both genes showed similar expression pattern at distinct stages and peaked simultaneously at hatching (21 dpf) (Figure 10). These results indicate *Pan* I and *Sypl*2 functional significance during early cod development. The *Sypl*2 transcript was localized in

the head, heart and liver of 60 d° (10 dpf) cod developing embryo. In the cod larvae of 90 d° (15 dpf) *Sypl*2 mRNA was detected in the eye, liver, heart, neuromast, ceratohyal and ceratobranchial arches, mandible and pectoral fin (Figure 12).

Ubiquitous expression of *Pan* I and *Sypl*2 revealed in various tissues in adult NEAC and NCC was consistent with the previously reported expression in mammalian species (Leube, 1994; Haass et al., 1996; Windoffer et al., 1999). The head kidney, spleen, epithelial mucus and ovaries exhibited the highest *Pan* I mRNA levels, while the highest *Sypl*2 expression was detected in the muscle, testicles and epithelial mucus (Figure 10). An inverse expression pattern between *Pan* I and *Sypl*2 was observed in the muscle, head kidney and spleen in NEAC1 and NEAC2. In NCC1 and NCC2 both genes showed similar expression patterns with the highest gonadal and epithelial expression. The *Pan* I mRNA levels were higher in male NEAC2 compared to female NEAC1 in almost all examined tissues, especially in the liver, spleen and heart. Both *Pan* I and *Sypl*2 mRNA levels were higher in testicles and epithelial mucus in NCC1 compared to the ovaries and epithelial mucus in NCC2 and NCC3. Even though observed differences might reflect functional significance of *Pan* I *Sypl*2 in the examined tissues, statistical significance of observed expression differences was not possible to assess due to small number of tested individuals. In addition, one should not preclude that detected expression levels might be affected by age, sex, maturity, spawning season or some other factors.

6. Concluding remarks

The results obtained in this study confirm the high degree of evolutionary conservation of pantophysin, synaptophysin and synaptophysin-like genes in various vertebrates, including Atlantic cod. The revealing of incorrect sequence naming in several species might help to avoid further misunderstandings and it has an additional phylogenetic perspective. The examination of spatio-temporal expression patterns, with emphasis on differences between NCC and NEAC, provides the first characterization of the *Pan* I and *Sypl*2 expression in fish. However, increased number of tested individuals is required to further investigate expression differences between NCC and NEAC. The expression of the highly conserved regions should be further explored, with the special emphasis on the first intravesicular loop that might have the essential significance for the protein function. Furthermore, an effect of the *Pan* I inhibition on numerous genes can be investigated by applying powerful techniques such as microarrays. Additional studies of the precise tissue and subcellular distribution might help to achieve an understanding of the biological functions of individual members in the synaptophysin gene family.

7. **Reference**

Arthur, C. P. and Stowell, M. H. B. 2007. Structure of Synaptophysin: A Hexameric MARVEL-Domain Channel Protein. Structure 15, 707–714.

Brander, K. 2005. Spawning and life history information for North-Atlantic cod stocks. ICES Cooperative Research report 205.

Brooks, C. C., Scherer, P. E., Cleveland K., Whittemore, Jennifer L., Lodish, H. F. and Cheatham, B. 2000. Pantophysin Is a Phosphoprotein Component of Adipocyte Transport Vesicles and Associates with GLUT4-containing Vesicles. The Journal of Biological Chemistry 275, 2029–2036.

Godø, O.R. and Michalsen, K. 2000. Migratory behaviour of north-east Arctic cod, studied by use of data storage tags. Fisheries Research 48, 127-140.

Gorodilov, Y. N., Terjesen, B. F., Krasnov, A. and Takle, H. 2008. Description of Embryogenesis of Atlantic Cod Gadus Morhua. The Open Marine Biology Journal 2, 43-53

Eshkind, L. G. and Leube, R. E. 1995. Mice lacking synaptophysin reproduce and form typical synaptic vesicles. Cell Tissue Res 282: 423-433.

Estrada, B., Maeland, A.D., Stephen, S., Gisselbrechta, J.W., Bloorb, N.H. and Brownb, A.M. 2007. Developmental Biology 328-339.

Fahay, M.P., Berrien, P.L., Johnson, D.L. and Morse, W.W. 1999. Atlantic Cod (Gadus Morhua), Life history and habitat characteristics. National Ocean and Atomospheric Administration Technical Memorandum 122-152, 1-35.

Falk-Petersen, I.B. 2005. Comparative organ differentiation during early life stages of marine fish. Fish Shellfish Immunology 19, 397-412.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791.

Fevolden, S.E. and Pogson, G.H. 1997. Genetic divergence at the synaptophysin (Syp I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. Journal of Fish Biology 51, 895–908.

Haass, N. K., Kartenbeck, J. and Leube, R. E. 1996. Pantophysin Is a Ubiquitously Expressed Synaptophysin Homologue and Defines Constitutive Transport Vesicles. The Journal of Cell Biology, 134 (3) 731-746.

Jahn, R., Schiebler, W., Ouimet, C. and Greengard, P. 1985. A 38,000-dalton membrane protein (p38) present in synaptic vesicles. Proc Natl Acad Sci U S A. 82(12): 4137–4141.

Janz, R. and Südhof, T. C. 1998. Cellugyrin, a novel ubiquitous form of synaptogyrin that is phosphorylated by pp60 c-src. J Biol Chem 273: 2851-2857.

Jobling, M. 1995. Enviromental Biology of Fishes. Chapman & Hall.

Jorstad, K.E., Van Der Meeren, T., Paulsen, O.I., Thomsen, T., Thorsen., A. and Svasand, T. 2008. ”Escapes” of eggs from farmed cod spawning in net pens: Recruitment to wild stocks. Rev. Fish. Sci. 16, 285-295.

Kwon, S. E. and Chapman, E. R. 2011. Synaptophysin Regulates the Kinetics of Synaptic Vesicle Endocytosis in Central Neurons. Neuron 70, 847–854.

Leube, R.E. 1994. Expression of the synaptophysin gene family is not restricted to neuronal and neuroendocrine differentiation in rat and human. Differentiation 56: 163-171.

Leube, R. E., Kaiser, P., Seiter, A., Zimbelmann, R., Werner, W.F., Rehm, H., Knaus, P., Prior, P., Betz, H., Reinke, H., Beyreutherl K., Wiedenmann B. 1987. Synaptophysin: molecular organization and mRNA expression as determined from cloned cDNA. The EMBO Journal 6: 3261 -3268.

Lough, R. G., Valentine, P. C., Potter, D. C., Auditore, P. J., Bolz, G. R., Neilson, J. D. and Perry, R. I. 1989. Ecology and distribution of juvenile cod and haddock in relation to sediment type and bottom currents on eastern Georges Bank. Mar. Ecol. Prog. Ser., 56: 1–12.

Nei M. and Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nicholas, K.B. and Nicholas, H. B. Jr. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by authors.

Nordeide, T.J., Johansen, S.T., Jørgensen, T.E., Karlsen, B.O. and Moum, T. 2011. Population connectivity among migratory and stationary cod Gadus morhua in the Northeast Atlantic: A review of 80 years of study. Mar. Ecol. Prog. Ser. 435, 269–283.

Pampoulie, C., Jörundsdóttir, T.D., Steinarsson, A., Pétursdóttir, G., Stefánsson, M.Ö. and Daníelsdóttir, A.K. 2006. Genetic comparison of experimental farmed strains and wild Icelandic populations of Atlantic cod (Gadus morhua L.). Aquaculture 261 (2), 556–564.

Pampoulie, C., Jakobsdóttir, KB., Marteinsdóttir, G., Thorsteinsson, V. 2008. Are vertical behaviour patterns related to the pantophysin locus in the Atlantic cod (Gadus morhua L.)? Behav Genet. 38 (1), 76-81.

Pogson, G. H. 2001 Nucleotide Polymorphism and Natural Selection at the Pantophysin (Pan I) Locus in the Atlantic Cod, Gadus morhua (L.). Genetics 157: 317–330.

Pogson, G. H. and Fevolden, S. E. 1998. DNA heterozygosity and growth rate in Atlantic cod Gadus morhua (L.). Evolution 52: 915-920.

Pogson, G. H. and Fevolden, S. E. 2003. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (PanI) locus. Molecular Ecology 12, 63-74.

Pogson, G. H. and Mesa, K. A. 2004. Positive Darwinian Selection at the Pantophysin (Pan I) Locus in Marine Gadid Fishes. Mol. Biol. Evol. 21(1):65–75.

Pogson, G. H., Mesa, K. A. and Boutilier, G. 1995. Genetic Population Structure and Gene Flow in the Atlantic Cod Gadus morhua: A Comparison of Allozyme and Nuclear RFLP Loci. Genetics 139: 375-385.

Pogson, G.H., Taggart, C. T., Mesa K. A. and Boutilier, R. G. 2001. Isolation by distance in the Atlantic cod, Gadus morhua, at large and small geographic scales. Evolution 55 (1): 131-146.

Pörtner, HO., Berdal, B., Blust, R. 2001. Climate induced temperature effects on growth performance, fecundity and recruitment in marine fish: developing a hypothesis for a cause and effect relationships in Atlantic cod (Gadus Morhua) and common eelpout (Zoarces viviparous). Cont. Shelf Res 21, 1975-97.

Saitou, N. and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406-425.

Sanchez-Pulido, L., Belmonte, F. M., Valencia, A. and Alonso M. A. 2002. MARVEL: a conserved domain involved in membrane apposition events. Trends Biochem Sci 27: 599-601.

Sarvas, T. H. and Fevolden, S. E. 2005a. Pantophysin (Pan I) locus divergence between inshore v. offshore and northern v. southern populations of Atlantic cod in the north-east Atlantic. Journal of Fish Biology 67, 444—469.

Sarvas, T. H. and Fevolden, S. E. 2005b. The scnDNA locus Pan I reveals concurrent presence of different populations of Atlantic cod (Gadus morhua L.) within a single fjord. Fisheries Research 76, 307–316.

Schmitt, U., Tanimoto, N., Seeliger., M., Schaeffeld, F. and Leube, R. E. 2009. Detection of behavioral alterations and learning deficits in mice lacking synaptophysin. Neuroscience 162 234–243.

Singleton, DR., Wu, TT. and Castle, JD. 1997. Three mammalian SCAMPs (secretory carrier membrane proteins) are highly related products of distinct genes having similar subcellular distributions. J Cell Sci 110: 2099-107.

Stenius, K.., Janz, R., Südhov, T. C. and Jahn, R. 1995. Structure of synaptogyrin (p29) defines novel synaptic vesicle protein. J. Cell Biol. 131:1801-1809.

Takamori, S., Holt, M., Stenius, K., Lemke, EA., Grønborg, M., Riedel, D., Urlaub, H., Schenck, S., et all. 2006. Molecular anatomy of a trafficking organelle. Cell. 127 (4): 831-46.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution (In Press).

Taranger, G.L., Carrillo, M., Schulz, R.W., Fontaine, P., Zanuy, S., Felip, A., Weltzien, F.A., Dufour, S., Karlsen, O., Norberg, B., Andersson, E. and Hansen, T. 2010. Control of puberty in farmed fish. Gen. Comp. Endocrinol. 165, 483-515.

Trippel, E. A., Benfey, T. J., Neil, S. R. E., Cross, N., Blanchard, M. J. and Powell, F. 2008. Effects of continuous light and triploidy on growth and sexual maturation in Atlantic cod, Gadus morhua. Cybium 32 (2), 136-138.

Wiedenmann, B. and Franke, W. W. 1985. Identification and localization of synaptophysin, an integral membrane glycoprotein of Mr 38, 000 characteristic of presynaptic vesicles. Cell 41. 1017-1028.

Windoffer, R., Borchert-Stuhlträger, M., Haass, NK., Thomas, S., Hergt, M., Bulitta, CJ. and Leube, RE. 1999. Tissue expression of the vesicle protein pantophysin. Cell Tissue Res. 296(3): 499-510.

Appendix

A. Sequences

A1. Atlantic cod *Pan* I identified in Celera assembly

```
atgtatatctttcatcatcatcagacaattctgacgtcgagattttcgcttgacttgggg
M  Y  I  F  H  H  H  Q  T  I  L  T  S  R  F  S  L  D  L  G
ccattgaaggagccccttgcattcatccgagttctggaattgatattctccatatttgca
P  L  K  E  P  L  A  F  I  R  V  L  E  L  I  F  S  I  F  A
tttgcaacaactgcaggattttctggtaccacaagtatcaacgtgcaatgcaaaggaagc
F  A  T  T  A  G  F  S  G  T  T  S  I  N  V  Q  C  K  G  S
gtcaacgtggaaatatttgccagttttaactacccatttaggttgatgcagcatccttac
V  N  V  E  I  F  A  S  F  N  Y  P  F  R  L  M  Q  H  P  Y
aatgtccctgactgtaagaatggcaccactgagtctttccttattggcaactactcttcg
N  V  P  D  C  K  N  G  T  T  E  S  F  L  I  G  N  Y  S  S
tcagcagagttctttgtctccatcggggtgctgtccttcctctacagcacagcctcgttg
S  A  E  F  F  V  S  I  G  V  L  S  F  L  Y  S  T  A  S  L
gtcctctatctgggcttcgaacacctgtacaggaagacaagccgtggtcctgttgtggat
V  L  Y  L  G  F  E  H  L  Y  R  K  T  S  R  G  P  V  V  D
ctgtttgtgactgcagccctggccttcctgtggttggtttcgtcttcggcttggggcaaa
L  F  V  T  A  A  L  A  F  L  W  L  V  S  S  S  A  W  G  K
ggtctgactgatgtcaagtgggccactagccccacaagcatagtggcacttagtgatgtc
G  L  T  D  V  K  W  A  T  S  P  T  S  I  V  A  L  S  D  V
tgcaggtctgacggaaattgcacggcgggcagtgtgccccacatgggccgtctcaacgcc
C  R  S  D  G  N  C  T  A  G  S  V  P  H  M  G  R  L  N  A
tcggtgatttttggcttcctcaacctcattctgtggggaagcaactgttggttcatttat
S  V  I  F  G  F  L  N  L  I  L  W  G  S  N  C  W  F  I  Y
aaggaaacacctttccacaagtcagccaatcagccagaagacgcggaggcgagaggcccg
K  E  T  P  F  H  K  S  A  N  Q  P  E  D  A  E  A  R  G  P
Cccacttaa
P  T  -
```

Figure A1. The nucleotide and deduced amino acid sequence of Atlantic cod Pan I identified in Celera assembly (ATLCOD1Bc1522258). The coding sequence is shown in small black letters. The deduced amino acids are shown in capital black letters. The termination codon (taa) is underlined.

A2. Atlantic cod *Sypl2* identified in Celera assembly

```
atggagatcgttcagagagctgtgtcgggcttcgcacttgacctgggacccatcaaggag
 M  E  I  V  Q  R  A  V  S  G  F  A  L  D  L  G  P  I  K  E
cccctggggttcatccgtgtcctggaatgggtgttcaccatatttgcgttcgccacggca
 P  L  G  F  I  R  V  L  E  W  V  F  T  I  F  A  F  A  T  A
gggggctattcggggagcacccactttacggtcaaatgccacgaaaagccaatggaagtc
 G  G  Y  S  G  S  T  H  F  T  V  K  C  H  E  K  P  M  E  V
aagcccgagttctggtacccttttcaggttggccgacagatcctacctcatgccaacatgc
 K  P  E  F  W  Y  P  F  R  L  A  D  R  S  Y  L  M  P  T  C
aatagcaccgaagtggaagtaacgtacctccagggtgacttccgctcctctgccgagttc
 N  S  T  E  V  E  V  T  Y  L  Q  G  D  F  R  S  S  A  E  F
tttgtgagcgtgggggtgtttgccttcctgtactgcaccctcacgctggtcctctacctg
```

```
 F  V  S  V  G  V  F  A  F  L  Y  C  T  L  T  L  V  L  Y  L
ggctaccagcatgtgtaccggcagaccagtcgcggacccctcgccgacctggtggtcacc
 G  Y  Q  H  V  Y  R  Q  T  S  R  G  P  L  A  D  L  V  V  T
ggtgtgtttgccttcctttggctggtgtcctcgtcagcctggggcaaggggctcacagac
 G  V  F  A  F  L  W  L  V  S  S  S  A  W  G  K  G  L  T  D
atcaagtttgctaccaacccagaccacctgatgggcagcctcaaggacgagtataccgtg
 I  K  F  A  T  N  P  D  H  L  M  G  S  L  K  D  E  Y  T  V
gaggcatacgcctccatgggccgtctcaacgcttctgtgatatttggctttttgaacctg
 E  A  Y  A  S  M  G  R  L  N  A  S  V  I  F  G  F  L  N  L
attctgtgggccggtaactgctggttcatttacaaagagacgcacttccacaaggacccc
 I  L  W  A  G  N  C  W  F  I  Y  K  E  T  H  F  H  K  D  P
aacgctccggccccatgcaggagggagatctgcccgggccctag
 N  A  P  A  P  M  Q  E  G  D  L  P  G  P  -
```

Figure A2. The nucleotide and deduced amino acid sequence of Atlantic cod Sypl2 indetified in Celera assembly (ATLCOD1Bc1511891). The coding sequence is shown in small black letters. The deduced amino acids are shown in capital black letters. The region (yellow) with a 47 nt insertion (red) is indicated. The termination codon (tag) is underlined.

A3. Atlantic cod *Sypl*2 identified in Ensembl

```
atggagatcgttcagagagctgtgtcgggcttcgcacttgacctgggacccatcaaggag
 M  E  I  V  Q  R  A  V  S  G  F  A  L  D  L  G  P  I  K  E
cccctggggttcatccgtgtcctggaatgggtgttcaccatatttgcgttcgccacggca
 P  L  G  F  I  R  V  L  E  W  V  F  T  I  F  A  F  A  T  A
gggggctattcggggagcacccactttacggtcaaatgcccgaaaagccaatgggtcaag
 G  G  Y  S  G  S  T  H  F  T  V  K  C  P  K  S  Q  W  V  K
cccgagttctggtaccctttcaggggtggaagtaacgtacctccagggtgacttccgctcc
 P  E  F  W  Y  P  F  R  V  E  V  T  Y  L  Q  G  D  F  R  S
tctgccgagttctttgtgagcgtgggggtgtttgccttcctgtactgcaccctcacgctg
 S  A  E  F  F  V  S  V  G  V  F  A  F  L  Y  C  T  L  T  L
gtcctctacctgggctaccagcatgtgtaccggcagaccagtcgcggacccctcgccgac
 V  L  Y  L  G  Y  Q  H  V  Y  R  Q  T  S  R  G  P  L  A  D
ctggtggtcaccggtgtgtttgccttcctttggctggtgtcctcgtcagcctggggcaag
 L  V  V  T  G  V  F  A  F  L  W  L  V  S  S  S  A  W  G  K
gggctcacagacatcaagtttgctaccaacccagaccacctgatgggcagcctcaaggac
 G  L  T  D  I  K  F  A  T  N  P  D  H  L  M  G  S  L  K  D
gagtataccgtggaggcatacgcctccatgggccgtctcaacgcttctgtgatatttggc
 E  Y  T  V  E  A  Y  A  S  M  G  R  L  N  A  S  V  I  F  G
tttttgaacctgattctgtgggccggtaactgctggttcatttacaaagagacgcacttc
 F  L  N  L  I  L  W  A  G  N  C  W  F  I  Y  K  E  T  H  F
cacaaggaccccaacgctccggcccccatgcaggagggagatctgcccgggccctag
 H  K  D  P  N  A  P  A  P  M  Q  E  G  D  L  P  G  P  -
```

Figure A3. The nucleotide and deduced amino acid sequence of Atlantic cod Sypl2 identified in Ensembl (ENSGMOG00000010980). The coding sequence is shown in small black letters. The deduced amino acids are shown in capital black letters. The region (yellow) with a single GG insertion (red) is indicated. The termination codon (tag) is underlined.

Table 1 Primers used for PCR, real time qPCR and direct sequencing

Primer name	Sequence (5' – 3')	Application
Pan I s-qPCR	AGCCTCGTTGGTCCTCTA	qPCR
Pan I as-qPCR	CACTTGACATCAGTCAGACCTT	qPCR
Pan I s-ISH	CATTCATCCGAGTTCTGGAA	PCR
Pan I as-ISH	TGAGGTTGAGGAAGCCAA	PCR
Pan I s-ISH-2	CAGCATCCTTACAATGTCCCTGAC	PCR
Pan I as-ISH-2	TTCTGGCTGATTGGCTGACTTGT	PCR
*Sypl*2 s1-Syp2a	GGAGCACCCACTTTACGGTCAA	qPCR, direct sequencing
*Sypl*2 as1- Syp2a	CCCCACGCTCACAAAGAACTC	direct sequencing
*Sypl*2 as2-Syp2a	CCAGGTAGAGGACCAGCGTGA	qPCR, direct sequencing
*Sypl*2 s2-Syp2a	GGTTCATCCGTGTCCTGGAA	PCR
*Sypl*2 as3-Syp2a	GGTCCTTGTGGAAGTGCGTCT	PCR
*Sypl*2 ins-as-Syp2a	GGTGCTATTGCATGTTGGCAT	PCR
*Sypl*2 ins.skip-as-Syp2a	GGAGGTACGTTACTTCCACTGAAA	PCR
ubiqutin F-qPCR	GGCCGCAAAGATGCAGAT	qPCR
ubiqutin R-qPCR	CTGGGCTCGACCTCAAGAG	qPCR

B. Synteny analysis

Table 2 Abbreviations and corresponding full names of pantophysin, synaptophysin synaptophysin-like flanking genes included in the synteny analysis

Gene name abbreviation	Full gene name
GPR61	G protein-coupled receptor 61 (1 of 2)
GPR61	G protein-coupled receptor 61 (2 of 2)
ATXN7L1	ataxin 7-like 1
ATXN7L2	ataxin 7-like 2 (2 of 2)
ATXN7L2b	ataxin 7-like 2b
SORT1	sortilin (1 of 2)
SORT1a	sortilin 1a
SORT1b	sortilin 1b
NGRN	neugrin, neurite outgrowth associated
NFYA	nuclear transcription factor Y, alpha
NFYAL	nuclear transcription factor Y, alpha, like
AHCYL1	adenosylhomocysteinase-like 1 (1 of 2)
AHCYL1	adenosylhomocysteinase-like 1 (2 of 2)
KCNC4	potassium voltage-gated channel, member 4
CYB561D1	cytochrome b-561 domain containing 1
ZBTB40	zinc finger and BTB domain containing (1 of 2)
ZBTB40	zinc finger and BTB domain containing (2 of 2)
USP48	ubiquitin specific peptidase 48
UCKL1	uridine-cytidine kinase 1-like 1 (2 of 2)
EMILIN3	elastin microfibril interfacer 3 (2 of 2)
PRICKLE3	prickle homolog 3
PLP2	proteolipid protein 2
PPP1R3A	protein phosphatase 1

C. Real time qPCR analysis

Table 3 Relative expression of Pan I in NEAC1, NEAC2, NCC1, NCC2 and NCC3 calculated by using the comparative Ct method and presented as fold-difference in expression of target gene relative to the calibrator (sample with the lowest gene expression). Pan I was normalized to ubiquitin. The following tissues are examined: telencephalon (TEL), optic tectum (OPT), mesencephalon (MES), cerebellum (CER), hypophysis (HYP), medulla oblongata (MED), brain (BR), eye (EYE), gills (GIL), heart (HAR), spleen (SPL), liver (LIV), muscle (MU), head kidney (HK), pyloric ceca (PC), anterior part of the intestines (INT) and distal part or colon (COL), epithelial mucus (EM) and gonads (GON). Absence of bars marked with NA represents samples that are not analyzed.

Tissue	NEAC1 female	NEAC2 male	NCC1 male	NCC2 female	NCC3 female
TEL	NA	NA	3.94	1.11	3.86
OPT	NA	NA	2.62	3.09	4.18
MES	NA	NA	NA	1.99	5.82
CER	NA	NA	NA	1.19	1.00
HYP	NA	NA	1.00	NA	1.69
MED	NA	NA	2.70	3.04	2.69
BR	NA	6.02	NA	NA	NA
EYE	1.49	1.00	NA	NA	NA
GIL	2.73	6.17	NA	NA	NA
HAR	2.74	10.16	3.61	1.86	3.76
SPL	5.35	27.38	NA	NA	NA
LIV	1.11	9.45	3.39	1.60	NA
MU	1.00	2.34	NA	NA	NA
HK	8.51	20.75	NA	NA	11.43
PC	1.41	3.96	NA	NA	NA
INT	NA	NA	3.16	1.11	2.50
COL	NA	NA	4.08	1.00	NA
EM	NA	NA	12.68	4.36	4.76
GON	NA	NA	12.13	6.54	2.92

Table 4 Relative expression of Sypl2 in NEAC1, NEAC2, NCC1, NCC2 and NCC3 calculated by using the comparative Ct method and presented as fold-difference in expression of target gene relative to the calibrator (sample with the lowest gene expression). Pan I was normalized to ubiquitin. The following tissues are examined: telencephalon (TEL), optic tectum (OPT), mesencephalon (MES), cerebellum (CER), hypophysis (HYP), medulla oblongata (MED), brain (BR), eye (EYE), gills (GIL), heart (HAR), spleen (SPL), liver (LIV), muscle (MU), head kidney (HK), pyloric ceca (PC), anterior part of the intestines (INT) and distal part or colon (COL), epithelial mucus (EM) and gonads (GON). Absence of bars marked with NA represents samples that are not analyzed.

Tissue	NEAC1 female	NEAC2 male	NCC1 male	NCC2 female	NCC3 female
TEL	NA	NA	6.30	2.02	2.35
OPT	NA	NA	3.24	7.86	1.61
MES	NA	NA	NA	1.00	1.61
CER	NA	NA	NA	2.15	1.00
HYP	NA	NA	1.00	NA	1.65
MED	NA	NA	1.64	1.71	1.43
BR	NA	NA	NA	NA	NA
EYE	4.35	1.74	NA	NA	NA
GIL	2.75	3.33	NA	NA	NA
HAR	5.28	3.40	1.64	3.16	2.18
SPL	1.00	1.90	NA	NA	NA
LIV	1.94	6.02	3.38	4.17	NA
MU	14.43	9.38	NA	NA	NA
HK	1.39	1.00	NA	NA	4.18
PC	4.91	4.23	NA	NA	NA
INT	NA	NA	11.71	3.85	4.91
COL	NA	NA	17.66	8.22	NA
EM	NA	NA	22.71	15.35	19.84
GON	NA	NA	37.79	13.74	3.36