Molecular Evolution of the def6/swap70 Gene Family and Functional Analysis of swap70a in Zebrafish Embryogenesis

A thesis presented in partial fulfilment of the requirement for the degree of

Master of Research in Molecular Biology

in the

Institute of Genetics
School of Biology
Faculty of Medicine and Health Sciences
University of Nottingham

by

Wai Ho Shuen, BSc (1st Class Hon)

September 2010

Supervisor: Professor Fred Sablitzky

Abstract

Rho GTPases including Rac1, RhoA, and Cdc42 are molecular switches as for signal transduction. Cycling between the GTP-bound active state and GDP-bound inactive state is tightly controlled by regulatory proteins. The exchange of GDP for GTP is catalysed by guanine nucleotide exchange factors (GEFs). Upon the activation of Rho GTPases through GEFs, downstream effector molecules are activated and thus trigger cellular responses such as actin cytoskeletal reorganisation, membrane ruffling, cell migration, and gene expression. Based on homology, there are three main families of Rho GEFs, Dbl family, Dock family, and def6/swap70 family. The large Dbl family is characterised through an invariable domain arrangement of an N-terminal catalytic dbl homology (DH) and a C-terminal regulatory pleckstrin homology (PH) domain whereas Dock family members lack the DH domain but instead contain a Dock homology region 2 (DHR2) domain. The def6/swap70 GEFs on the other hand contain an atypical and unique PH-DH-like domain arrangement. Mammalian DEF6 and SWAP70 that exhibit a high similarity in their N-terminal ends containing a putative Ca^{2+} - binding EF hand are crucial mediators of signal transduction in T and B cells, respectively. Phylogenetic sequence analysis revealed that the atypical domain structure as well as the primary amino acid sequences of def6 and swap70 family members has been highly conserved in vertebrates and invertebrates. Whereas invertebrates have only one *def6/swap70* gene, two genes have been identified in tetrapod species, and four to five genes have been identified in teleosts species. In zebrafish, five paralogous genes were identified: *def6a, def6b, swap70a, swap70b and def6-like*. Remarkably, the predicted secondary and tertiary structure of all *def6/swap70* family members including the five proteins identified in zebrafish are very similar; most of them folding into a 'donut-shape' structure. The expression profile of the *def6/swap70* genes during zebrafish development indicated that *def6a* and *swap70a* are expressed maternally as well as zygotically during early development whereas expression of the other three genes was restricted to later stages. Morpholino-mediated knockdown of *swap70a* using two different splice morpholinos resulted in a delay of zebrafish development likely to be due to impaired convergence and extension cell movements during gastrulation. In addition, development of brain, eyes, ears number of otoliths and tail formation was affected. Preliminary data using the AUG

morpholino to target maternal and zygotic expression of *swap70a* indicate a more severe phenotype and high mortality of the morphants. Co-injection of GFP-tagged *swap70a* mRNA in low dose resulted in a partial rescue of the splice morpholino-mediated phenotype. Over-expression of GFP-swap70a in high dose however, resulted also in developmental defects in eyes, number of otoliths and tail formation. The observed phenotype of *swap70a* morphants described here is reminiscent of the phenotype of *def6a* morphants that was shown to be downstream of Wnt5b in the non-canonical Wnt/PCP signalling pathway (Goudevenou *et al*. in preparation) regulating convergence extension cell movement during gastrulation. It is therefore tempting to speculate that *swap70a* has a similar role acting either in conjunction with or in parallel to *def6a* in the non-canonical Wnt signalling pathway.

Acknowledgements

First, I would like to thanks my supervisor Professor Fred Sablitzky for giving me an opportunity to learn from him and his guidance during my MRes research. It is my honour to be his MRes student. He taught me quite a lot throughout the year and always allows me to try whatever I want. Big thanks go to Dr Peter Jones. He always gives me useful opinions and suggestions. Special thanks go to Professor John Brookfield and Dr Liz Bailes for their excellent discussions and opinions on bioinformatics analyses.

Thanks to PhD student Katerina Goudevenou for her teaching and experiment demonstration. She is always taking care of me and is so helpful. I am also grateful to PhD student Fiona Hey for the friendship and advices. Thanks also go to Lorraine for her help in our lab and the fun she gave us. I also appreciate Andrea for her help on zebrafish embryo collection.

Last but not least, I would like to sincerely thank Hong Kong Alistair Harvey Foundation Scholarship for the scholarship covering tuition fees during my studies in UK. Also, thanks to School of Biology, University of Nottingham for the MRes International Scholarship.

Table of Contents

List of Figures

List of Tables

Abbreviations

%	Percent
a.a.	Amino Acid
bp	Base pair
Cdc42	Cell division cycle 42
CE	Convergence and Extension
DEF6	Differentially expressed in FDCP 6
DH Domain	Dbl homology domain
DHL Domain	DH-like domain
DHR2	Dock homology region-2
DNA	Deoxyribosenucleic Acid
EDTA	Ethylenediaminetetraacetic acid
et al.	*et alteri*
Fz	Frizzled
GAPs	GTPase-activating proteins
GDIs	GDP dissociation inhibitors
GDP	Guanosine 5’-diphosphate
GEFs	Guanine nucleotide exchange factors
GFP	Green fluorescent protein
GTP	Guanosine 5’-triphosphate
hpf	hour post-fertilisation
MO	Morpholino
mRNA	Messeger ribonucleic acid
NES	Nuclear Exit Signal
NLS	Nuclear Localization Signal
PCP	planar cell polarity
PCR	Polymerase chain reaction
PH Domain	Pleckstrin homology domain
$PI(3,4)P_2$	Phosphatidylinositol (3,4)-triphosphate
$PI(3,4,5)P_3$	Phosphatidylinositol (3,4,5)-triphosphate
$PI(4,5)P_2$	Phosphatidylinositol (4,5)-diphosphate
PI3K	Phosphoinositide 3 Kinase
Rac	Ras-related C3 botulinum toxin substrate

Rac1	Ras-related C3 botulinum toxin substrate 1
Rho	Ras-homologue
RhoA	RAS homologue gene-family member A
RNA	Ribonucleic acid
SWAP70	Switch associated protein 70KDa
TBE	Tris-borate buffer
Tris	Tris(hydroxymethyl)aminomethane

1. Introduction

1.1 Rho Family GTPase

Rho GTPases act as tightly controlled molecular switch as for signal transduction. Rho GTPases are in an active form when they bind to GTP, whereas they are in an inactive state when they bind to GDP. There are three groups of proteins involved in the molecular switch cycle for the molecular switch of Rho GTPases. At inactive state, GDP dissociation inhibitors (GDIs) inhibit the exchange of GDP for GTP by sequestering the GDP-bound form of Rho GTPases. Upon extracellular signals, the exchange of GDP for GTP is catalysed by guanine nucleotide exchange factors (GEFs). Once the Rho GTPases are activated, they bind to downstream effector molecules. As a result, several cellular responses, such as cytoskeletal reorganisation, membrane trafficking, cell migration, gene expression and so on, will be triggered. The last group of proteins are GTPase-activating proteins (GAPs) which increase the weak intrinsic GTPase activities of Rho GTPases and thus the conversion of GTP-bound to GDP-bound is promoted as well as terminating the signal (reviewed in Fukata *et al*., 2003; reviewed in Pernis, 2009, reviewed in Hall, 1998; reviewed in Rossman *et al*., 2005; reviewed in Raftopoulou & Hall, 2004; Figure 1.1).

1.2 Rac, Rho, and Cdc42 are the members of Rho Family GTPases

As shown in Figure 1.2A, Rac, Rho, and Cdc42 regulate different pathways to govern cell migration. Rac is mainly involved at the front of the cell for actin reorganisation and polymerisation. Cdc42 which contributes to cell migration direction is also involved at the front of the cell. In contrast, Rho is supposed to be involved in regulation of contraction and retraction in the cell body and the rear of the cell (reviewed in Raftopoulou & Hall, 2004).

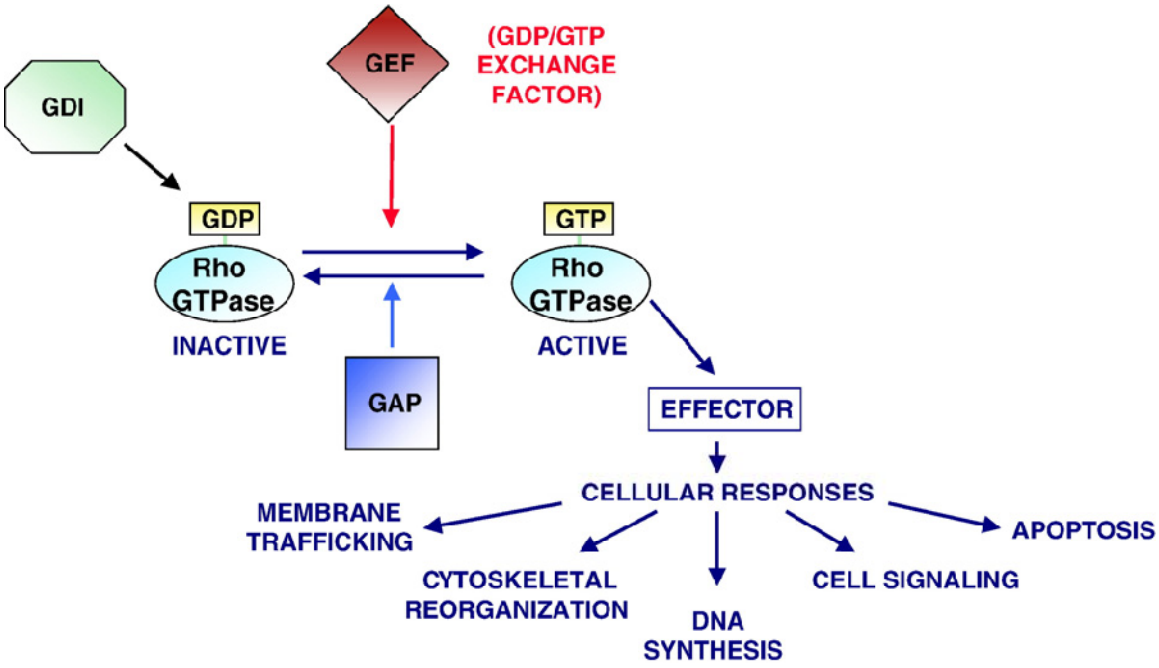


Figure 1-1 Molecular Switch of Rho GTPase. Molecular switch for signal transduction. Rho GTPases are in active form when they bind to GTP, whereas they are in inactive state when they bind to GDP. In the inactive state, GDP dissociation inhibitors (GDIs) inhibit the exchange of GDP for GTP. The exchange of GDP for GTP is catalysed by guanine nucleotide exchange factors (GEFs). Several cellular responses, such as cytoskeletal reorganisation, membrane trafficking, cell migration, and gene expression will be triggered. GTPase-activating proteins (GAPs) increase the conversion of GTP-bound to GDP-bound as a result of the signal termination. Adapted from Pernis, 2009.

1.3 Rac Signalling Pathways

Sub-family of Rac GTPase has Rac1, Rac2, and Rac3. But comparatively, Rac1 is well studied. As Raftopoulou & Hall (2004) mentioned, there are two pathways for Rac1 GTPase activation, phosphoinositide 3-kinase (PI3K)-dependent and PI3K-independent pathways.

1.4 PI3K-dependent Rac1 Signalling Pathway

According to the review by Welch *et al*. (2003) and Raftopoulou & Hall (2004), PI3K which is a lipid kinase is activated by cell surface receptors, such as receptor-tyrosine kinases and G-protein coupled receptors, depending on the types of PI3K. PI3K is known to be upstream of Rac1 GTPases. Activated PI3K will phosphorylate phosphatidylinositol (4,5)-diphosphate ($PI(4,5)P_2$) to be phosphatidylinositol (3,4,5)-triphosphate ($PI(3,4,5)P_3$) which is the lipid secondary messager. During the reaction, immediate product, phosphatidylinositol (3,4)-triphosphate ($PI(3,4)P_2$) is also formed. PIP_3 then binds to GEFs resulting in the activation of Rac1 GTPases. Also, Rac1 is proved to be upstream of PI3K as well to form the positive feedback loop (Figure 1.2B).

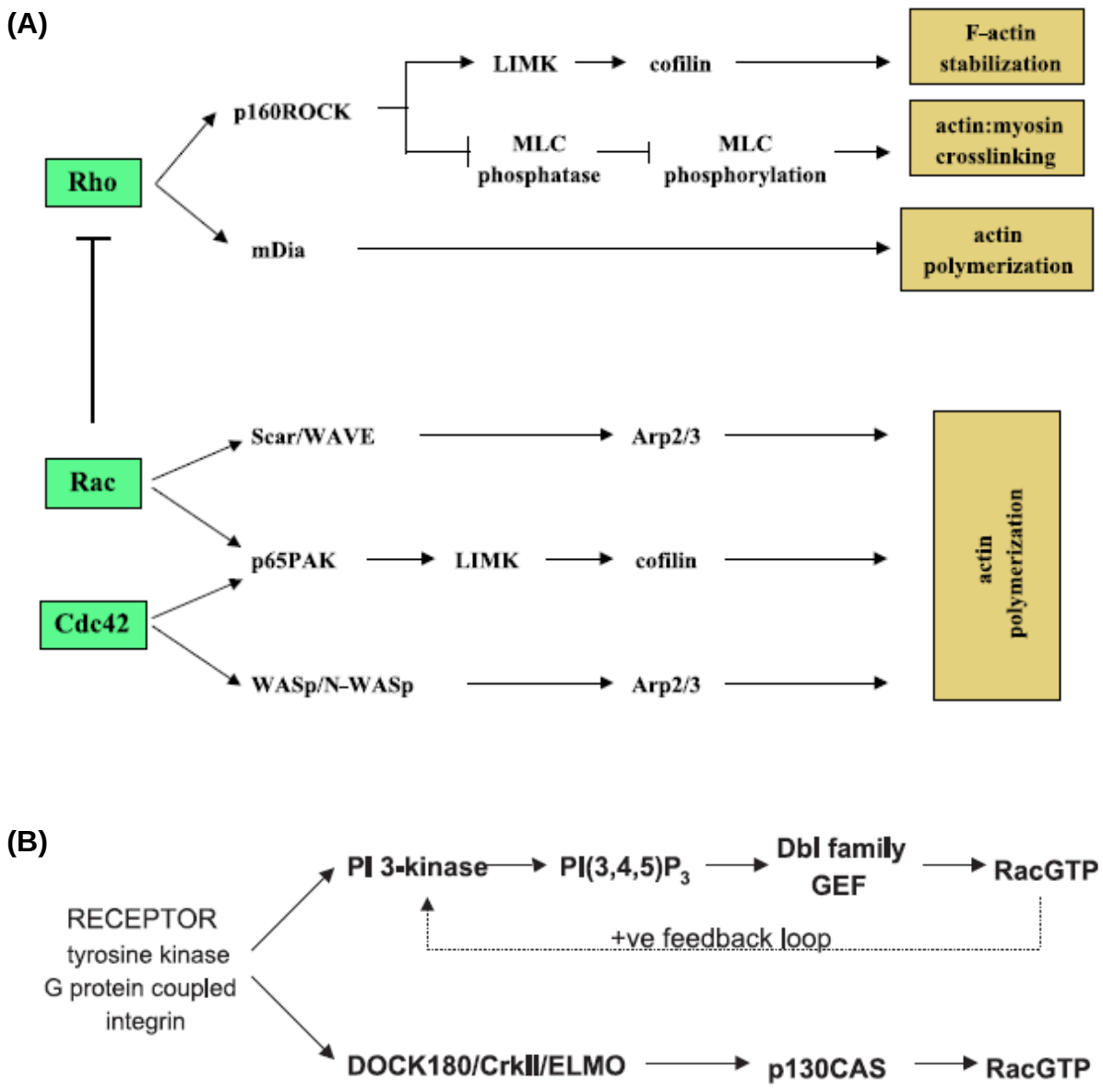


Figure 1-2 Signalling Pathways of Rho Family GTPases. (A) Rho, Rac, and Cdc42 and downstream effectors for cell migration. (B) Two Rac-GTPase Signalling Pathways. The upper one is PI3K-dependent pathway and the lower one is PI3K-independent pathway. Adapted from Raftopoulou & Hall (2004).

1.5 Guanine Nucleotide Exchange Factors (GEFs)

As mentioned above, GEFs catalyse the exchange of GDP for GTP and thus activate Rho GTPases. As of 2005, there are 84 GEFs, which can be grouped in three main families, identified in human. The major family is Dbl family. There are 69 GEFs which have multiple protein domains but all share the distinct feature of Dbl homology (DH) domain followed by pleckstrin homology (PH) domain. DH domain is essential for exchange of GDP for GTP whereas PH is important for PI(3,4,5)P_3 binding (reviewed in Zheng, 2001). The second family is called the Dock family containing 13 members, such as Dock180, Dock3, and Dock6. All members in this family have no DH domain and only few members have PH domain. The function of GTPase activation is carried out by Dock homology region-2 (DHR2) domain which is located at the C-terminal end of the amino acid sequences. The last family is def6/swap70 which contains DEF6 and SWAP70 only. Both DEF6 and SWAP70 have a special domain arrangement which is PH domain followed by DH domain. Dbl family members commonly contribute to the activation of Rho GTPases. In comparison, the Dock1, Dock2 and Dock3 from the Dock family are involved in the PI(3,4,5)P_3-independent Rac1 signalling pathway. SWAP70 in the def6/swap70 family is involved in the PI(3,4,5)P_3-dependent Rac1 signalling pathway. However, compared to the Dbl family, the Dock family, the def6/swap70 family are poorly characterised (reviewed in Rossman *et al*., 2005).

1.6 DEF6 and SWAP70 are Novel PH-DH-like Domain GEFs

DEF6 was discovered from mouse haemopoietic tissues (Hotfilder *et al*. 1999) and SWAP70 was isolated from B-cell nucleus complex (Borggrefe *et al*., 1998). DEF6 is mainly expressed in T cell whereas SWAP70 is mainly expressed in B cell. Both of them have putative EF-hand motif at the N-terminus, PH domain at the centre, and the region which shows limited homology to DH domain of classical GEFs is referred as DH-like (DHL) domain. The region between EF-hand motif and PH domain is referred as to DEF6-SWAP70 homology (DSH) domain (Mavrakis *et al*., 2004; reviewed in Tybulewicz & Henderson, 2009 and Biswas *et al*., 2010). Compared to typical GEFs (Figure 1.3), DEF6 and SWAP70 have an atypical domain arrangement which is PH domain C-terminal linked with DHL Domain (PH-DHL) (Figure 1.4).

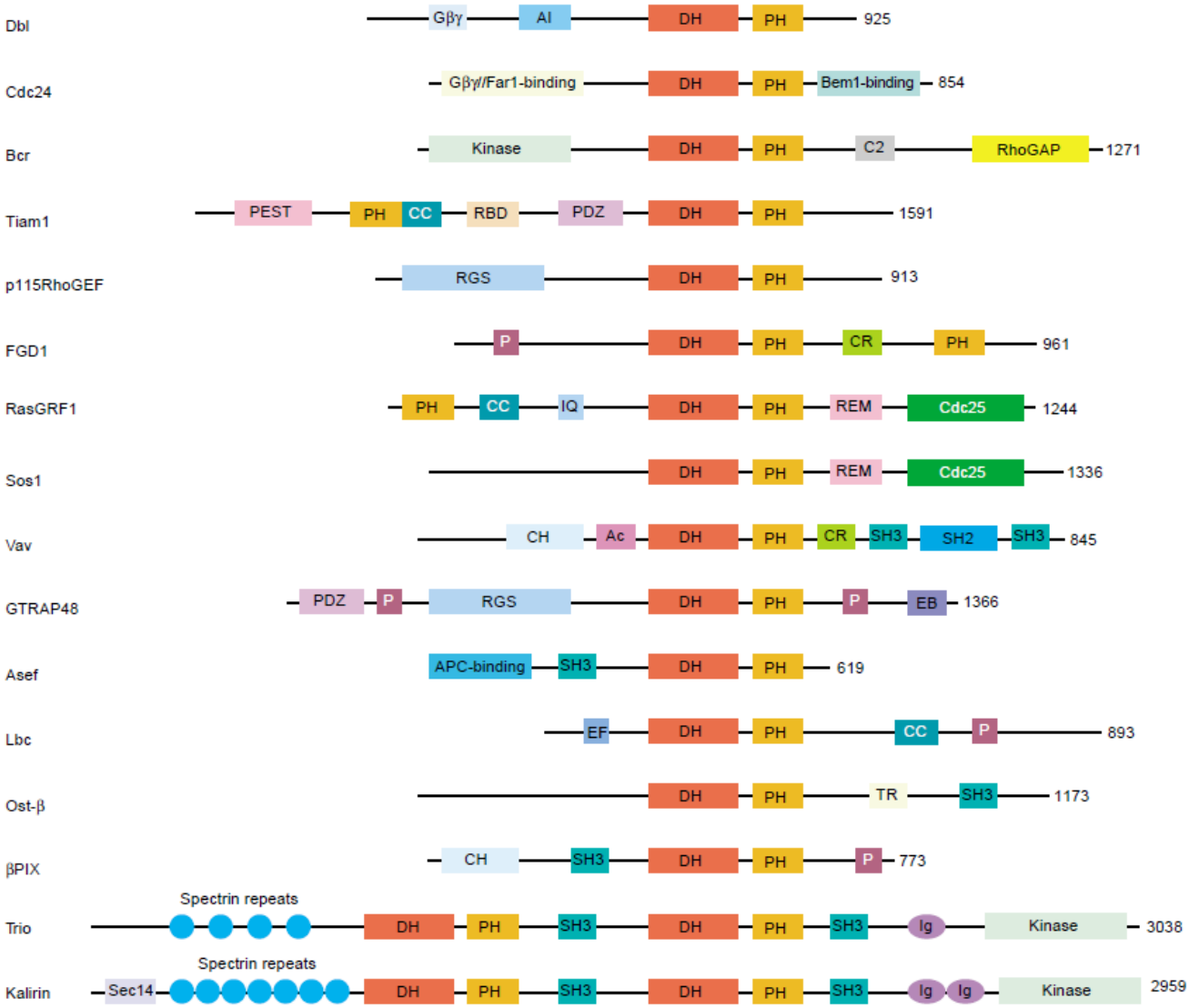


Figure 1-3 Typical Guanine Nucleotide Exchange Factors (GEFs). Typical GEFs have DH domain C-terminal linked with PH domain in the molecules. The DH-PH domain arrangement is found in most of the GEFs. The DH domain contributes to the GDP-GTP exchange for Rho GTPases and PH domain is involved in the binding with lipid secondary messager, PI(3,4,5)P_3. Adapted from Zheng (2001).

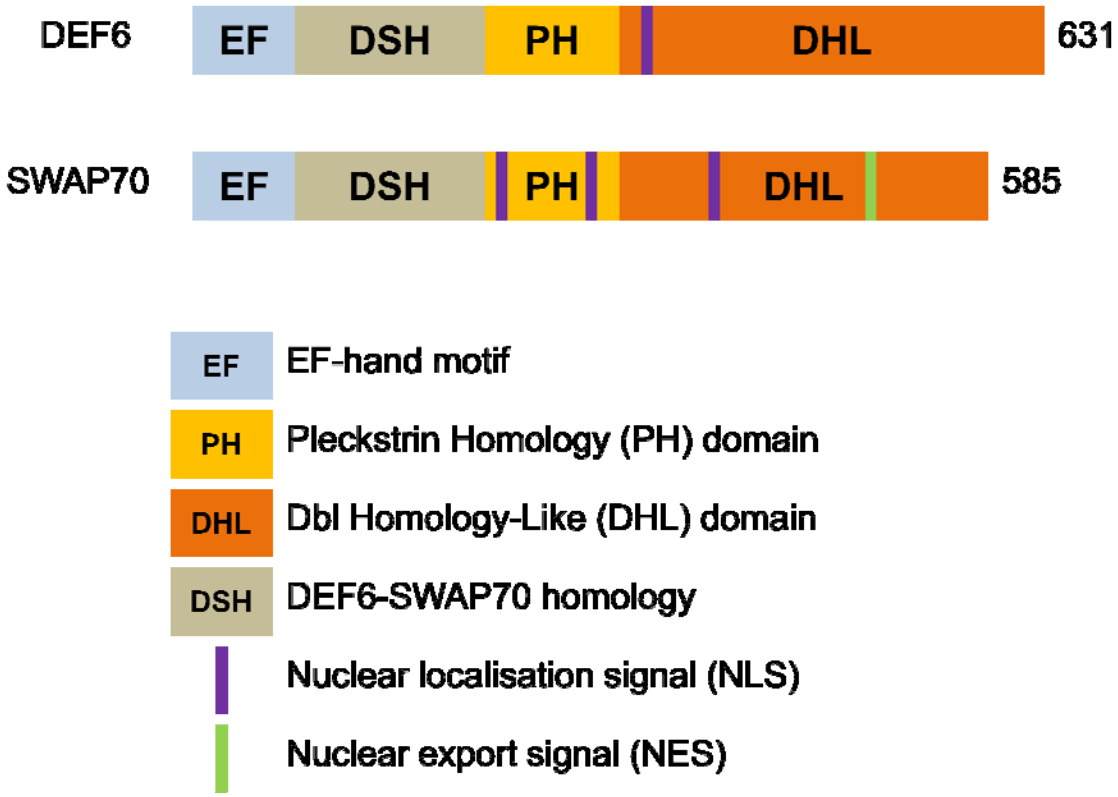


Figure 1-4 DEF6 and SWAP70 Atypical Domain Structures. DEF6 and SWAP70 have a putative EF-hand motif at the N-terminus, PH domain at the centre, and the region at the C-terminus which shows limited homology to DH domain of classical GEFs is referred to as DH-like (DHL) domain. The region between EF-hand motif and PH domain is referred as to DEF6-SWAP70 homology (DSH) domain. SWAP70 has 3 NLS and 1 NES whereas DEF6 has only 1 NLS (Mavrakis *et al.*, 2004; reviewed in Tybulewicz & Henderson, 2009 and Biswas *et al.*, 2010).

1.6.1 DEF6 and SWAP70 in Signal Transduction

DEF6 was shown to activate Rho GTPase family members, Rac1, Cdc42, and RhoA. In comparison, SWAP70 was shown to activate Rac1 and RhoA. Both can induce actin cytoskeleton rearrangement and membrane ruffling through PI3K-dependent signalling pathways (Shinohara *et al*., 2002; Mavrakis *et al*., 2004; Ocana-Morgner *et al*., 2009).

1.7 Functional Studies of SWAP70

1.7.1 SWAP70 is an atypical GEF for Rac1 contributing to membrane ruffling

SWAP70 acts as a Rac-GEF to mediate membrane ruffling (Shinohara *et al*., 2002). It binds to PI(3,4,5)P_3, which is a lipid second messager generated by PI3K, through a β1-β2 loop of the PH domain (Wakamatsu *et al*., 2006) and moves to the membrane by activity of the β3-β4 loop of PH domain (Fukui *et al*., 2007). Hilpela and co-workers (2003) found out that SWAP70 also bound to PI(3,4)P_2 and regulates the actin cytockeleton. SWAP70 is able to directly bind to F-actin through its C-terminal end, resulting in membrane ruffling (Ihara *et al*., 2006).

1.7.2 SWAP70 identified from B cell functions in both the nucleus and cytoplasm

SWAP70 which is a protein of 70kDa was found in a mouse B cell specific DNA recombination complex, SWAP (Borggrefe *et al*., 1998). The complex SWAP consists of four proteins, nucleophosmin, nucleotin, poly(ADP-ribose) polymerase, and SWAP70. Also, SWAP70 containing three nuclear localisation signals (NLSs) is involved in assembling switch recombinase for Ig heavy chain switching in nucleus. Borggrefe *et al*. (1999) further found out that SWAP70 is also present in cytoplasm. It is also highly expressed in activated B cells while no detection of SWAP70 was found in macrophages and T cells. Moreover, SWAP70 was mapped to mouse chromosome 7 and human chromosome 11p15 (Masat *et al*., 2000a). The human SWAP70 sequence was identified to have 3 NLSs and 1 nuclear export signal (NES) (Masat *et al*., 2000b). SWAP70 is mainly found in cytoplasm when a B cell is in rest and is translocated into the nucleus when a B cell is activated suggesting that SWAP70 has two roles. One is Ig heavy chain class switching in the nucleus and another one is in B cell activation signalling.

1.7.3 SWAP70 is involved in CD40 B cell signalling pathway and signal transduction in mast cell

According to Borggrefe *et al.* (2001), SWAP70-deficient B cells showed defects in CD40 signalling pathway and stimulation of autoantibody development resulting in autoimmune disease. The cells are also more sensitive to γ-irradiation than B cells of wild type. Importantly, besides B cells, SWAP70 is also expressed in mast cells. Immature mast cells mainly undergo FcεRI-mediated degranulation which is inhibited in SWAP70$^{-/-}$ mice, whereas mature mast cells which do not rely on FcεRI signalling showed normal degranulation in SWAP70$^{-/-}$ mice (Gross *et al*., 2002). Also, there is solid evidence to support SWAP70 acting downstream of FcεRI signalling pathway. First, activated c-Fos after the FcεRI is activated by cross-linking with IgE-antigen up-regulates *SWAP70* gene expression. As a result, degranulation is promoted (Lee *et al*., 2004). Second, in SWAP70$^{-/-}$ mice immature bone marrow mast cells results in anaphylaxis formed by alternating downstream FcεRI signalling such as gene expression of cytokines, PI3K activity, PIP_3 production (Rajeswari *et al*., 2008). A summary of SWAP70 function in mast cell FcεRI signalling is shown in Figure 1.5. Moreover, SWAP70 is involved not only in FcεRI signalling pathway, but also in growth factor receptor, c-kit signalling pathway in immature bone marrow mast cells. According to the works done by Rajeswari & Jessberger (2004), SWAP70 has a clear role in the c-kit signalling pathway contributing to mast cell activation, cell migration, and cell adhesion through activating Rac GTPase.

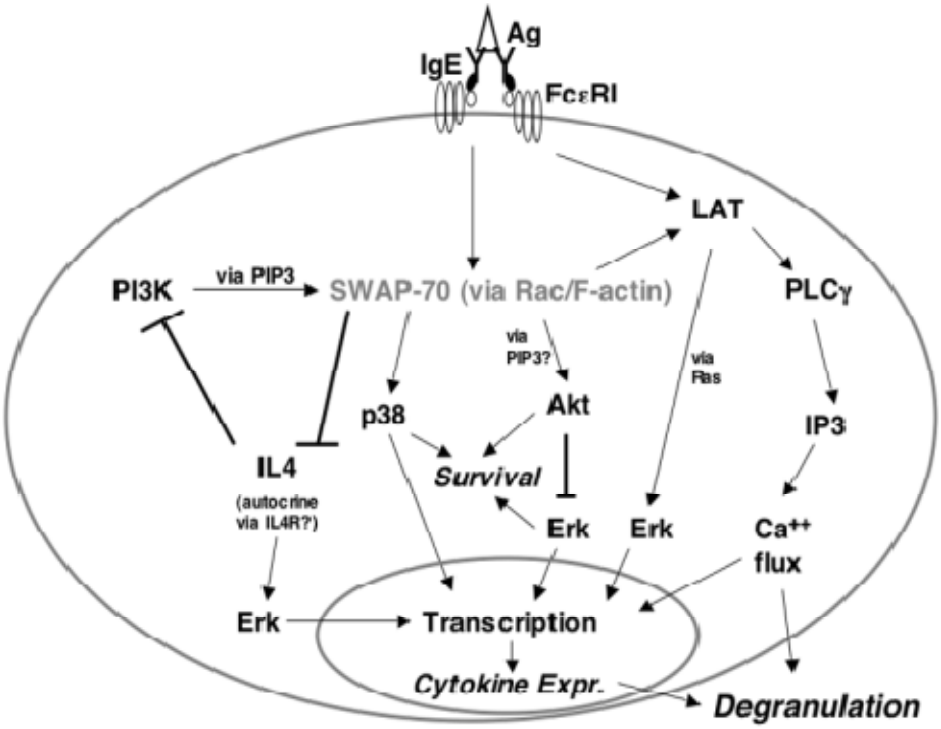


Figure 1-5 Summary of SWAP70 in Mast Cell Signalling. The FcεRI receptor is activated by cross-linking with IgE-antigen, *SWAP70* gene expression is up-regulated, followed by activation of Rac GTPase and F-actin binding to C-terminal region of SWAP70. As a result, downstream effector p38 is activated and gene expression is triggered, resulting in expression of cytokines, followed by degranulation. Adapted from Sivalenka *et al*. (2008).

1.7.4 SWAP70 is involved in early macropinocytosis and MHCII surface localisation regulation in dendritic cells

SWAP70 is transiently involved in early macropinosome formation in dendritic cells (Oberbanscheidt *et al*., 2007). Importantly, SWAP70 was firstly reported to bind to both active Rac1 and RhoA (Ocana-Morgner *et al*., 2009). However, SWAP70 negatively regulates RhoA and indirectly regulates RhoB whereas it shows positive regulation to Rac1. As a result, too early MHCII surface localisation is prevented through inhibition of RhoA/RhoB activation in dendritic cells.

1.7.5 SWAP70 is involved in signal transduction in cancer oncology

SWAP70^{-} deficient mouse embryo fibroblast showed poor tumor growth after v-Src virus transformation and had impared membrane ruffling. And thus the cells have poor cell movement and v-Src dependent invasiveness. Also, SWAP70 and v-Src act synergistically in the invasion. These suggest that SWAP70 is involved in the PI3K-dependent v-Src transformation pathway (Fukui *et al*., 2007; Murugan *et al*., 2008).

1.8 Functions of SWAP70 in Embryogenesis

1.8.1 Expression pattern of *swap70a* during zebrafish development by whole-mount *in situ* hybridisation

The expression pattern of *swap70a* was determined using whole-mount *in situ* hybridisation with an antisense *swap70a* probe. A sense *swap70a* probe served as control (Goudevenou unpublished work, 2007). As shown in Figure 1.6, *swap70a* is expressed from cleavage stage to hatching stage. Given that zygotic transcription starts at the mid-blastula stage at 2.45 hours post-fertilisation (hpf), the expression detected at cleavage stage indicates the presence of maternal *swap70a*. Expression of *swap70a* was still detected at bastula and gastrula stages. At early segmentation, *swap70a* expression is restricted to the optic primordium whereas at late segmentation, *swap70a* expression was mainly found in the optic primordium, otic vesicle, and ventral mesoderm. At pharyngula stage, *swap70a* is mainly expressed in the caudal vein, whereas at hatching stage, *swap70a* is mainly expressed in pectoral fins and pharyngeal arches. The expression pattern of *swap70a* reveals that *swap70a* is expressed throughout the zebrafish during early development and implies that *swap70a* function is temporally and spatially regulated.

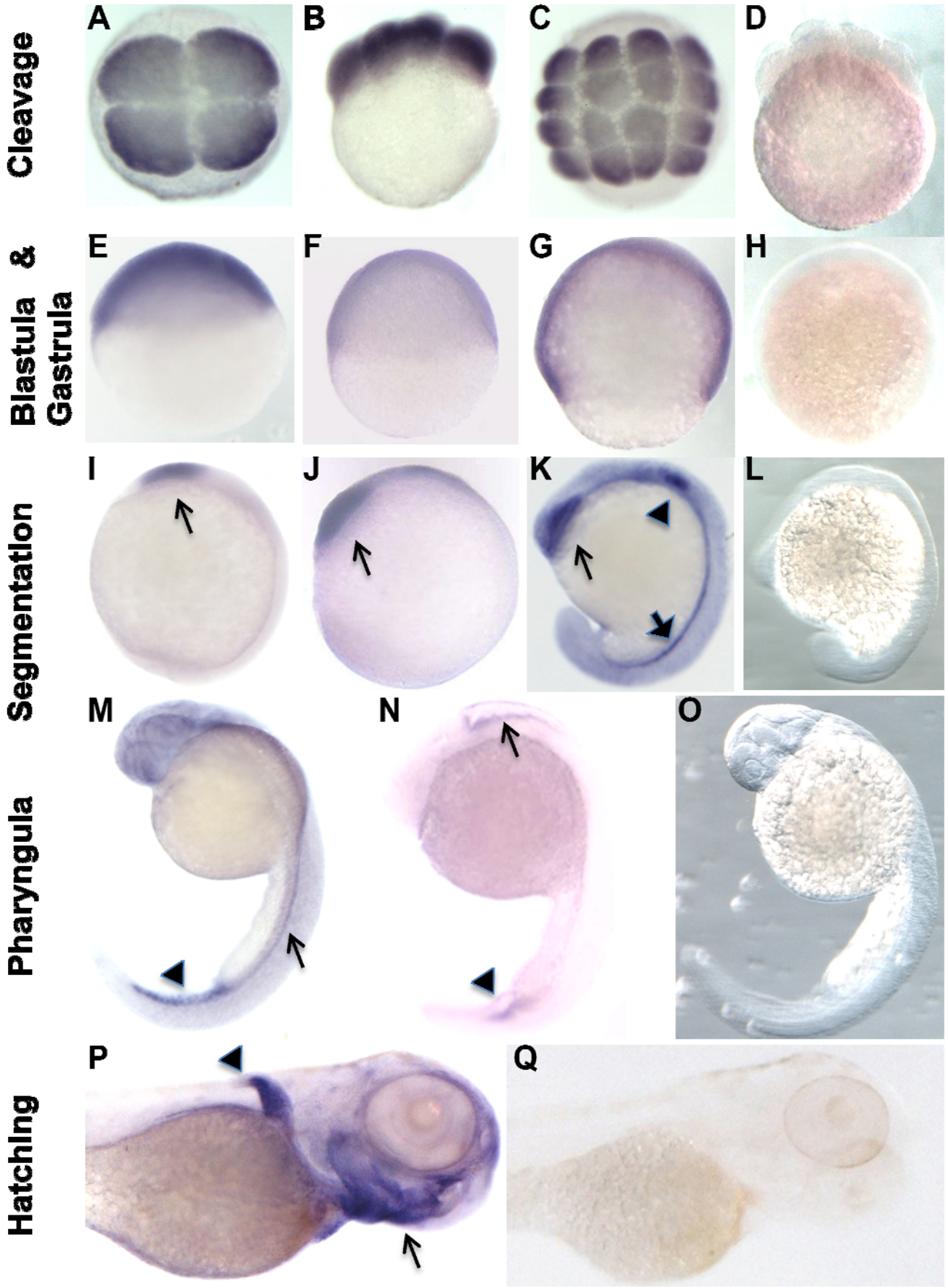

Cleavage
Blastula & Gastrula
Segmentation
Pharyngula
Hatching
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q

Figure 1-6 Expression pattern of *swap70a* during zebrafish embryogenesis. Embryos in A-C, E-G, I-K, M-N, and P were probed with anti-sense *swap70a*. D, H, L, O, and Q are negative controls of cleavage, blastula & gastrula, segmentation, and hatching stages, respectively, using sense *swap70a* probe. (A) 4-cell stage. (B) 8-cell stage. (C) 16-cell stage. (E) 30% epiboly. (F) 50% epidoly. (G) 70% epiboly. (M) prim-6. (N) prim-16. (P) high pec. (Goudevenou unpublished work, 2007)

1.8.2 Antisense Technology – Morpholino

Morpholinos (MOs) are modified anti-sense oligonucleotides that are used to knockdown gene expression in the zebrafish (*Danio rerio*) (Bill *et al*., 2009). There are two types of morpholinos (Bill *et al*., 2009; Heasman, 2002; Eisen & Smith, 2008). The first one (splice MO) binds to exon-intron boundaries of pre-mRNA and thus splicing is affected. Another one (AUG MO) target the initiation codon of the target mRNA, resulting in translation blocking. For the splice MO, it commonly causes retention of an intron which includes an in-frame stop codon or causes deletion of exon resulting in frame-shifting. Both events contribute to production of truncated proteins and thus the functions of the target proteins are knocked-down. In comparison, the protein synthesis is blocked as AUG MO binds to the start codon. Therefore, there is no target protein produced. As maternal mRNA are processed and do not undergo splicing, only zygotic transcripts are affected by splice MO. However, both maternal and zygotic transcripts, contain the same transcription start codon, can be affected by AUG MO.

1.8.3 Determination of MO Efficiency

The splice MO-induced altered splicing can be confirmed using RT-PCR to compare the amplicon size differences, whereas the AUG MO-induced impaired protein synthesis can be confirmed using western blotting with target protein antibodies. Nevertheless, mRNA level reduction does not represent the same level of reduction for protein and the limitation for AUG MO is that the detection of proteins require good quality antibodies (reviewed in Eisen & Smith, 2008).

1.8.4 Determination of MO Specificity

As the MO injection may cause some off-target effects, it is necessary to determine whether the MO-induced phenotypes are actually due to the knockdown of the target gene. A two non-overlapping oligos strategy and rescue experiments are the most commonly used approaches to rule out off-target effects. For the two non-overlapping oligos strategy, if the 2nd MO-induced phenotypes are similar or the same as 1st MO-induced phenotypes, this indicates the specificity of both MOs. A synergistic effect of co-injection of 1st and 2nd MOs in lower amounts also indicates the MO specificities. Moreover, if the co-injection of MO with mRNA rescues the phenotypes induced by MO, the specificity of MO is indicated as well.

1.9 Aims and Objectives

The def6/swap70 family which is a special group of GEFs is poorly characterised. Therefore, in the first part of this thesis, bioinformatics analyses were carried out to identify the family members in vertebrates and invertebrates and to investigate their relationships among others.

The functions of *swap70a* in zebrafish embryogenesis and the signalling pathway which *swap70a* may be involved in during development are unknown. Thus, in the second part of this thesis, gain-of and loss-of function analyses were performed to dissect the functions of *swap70a* in zebrafish embryogenesis and to predict the possible signalling pathways which *swap70a* is involved in.

2. Materials and Methods

2.1 Materials

2.1.1 Technical Equipment

Name	Company
Minispin plus centrifuge	eppendorf
Micromax RF 3593 centrifuge	IEC
Centrikon T-42K centrifuge	Kontron instrument
Dual-intensity UV transilluminator	UVP, LLC
Electrophoresis power supply EPS300	Pharmacia Biotech
Gradient PCR machine	Takara
Heating block DRI-BLOCK DB3	Jencons Techne
Horizon 58 electrophoresis apparatus	GIBCO BRL life technologies
Molecular Imager Gel Doc XR System	Bio-Rad
Nanodrop spectrophotometer ND1000	Nanodrop
Orbital incubator SI50	Stuart Scientific
Pipettes (0.2-2 µl, 2-20 µl, 50-200 µl and 100-1000 µl)	Gibson
Water bath	Jencons

2.1.2 Molecular Biology

Plasmids:

Name	Source
pEGFP-C1 (Figure 2.1A)	Clontech
pβUT3 (Figure 2.1B)	Prof. Roger Patient
pExpress-1+SWAP70 (Figure 2.1D)	Geneservice
pSC-B-amp/kan (Figure 2.1F)	Stratagene

Reaction kits:

Name	Company
StrataClone Blunt PCR Cloning Kit	Stratagene
First-Strand cDNA Synthesis	Invitrogen
mMESSAGE mMACHINE T3 Kit	Ambion
Pfu DNA Polymerase	Stratagene
Plasmid Midi Kit	QIAGEN
QIAprep Spin Minprep Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN
QIAshredder	QIAGEN
Quick T4 Ligase Kit	New England Biolads
REDTaq Ready Mix PCR Reaction Mix	Sigma-Aldrich
RNeasy Mini Kit	QIAGEN

Solutions:

Name	Compositions	Company
Resuspension Solution I	50mM glucose 10mM EDTA 25mM Tris-HCl pH 8.0 Filter sterilised Stored at 4°C	BDH BDH Sigma-Aldrich
Lysis Solution II	1% w/v SDS 0.2M NaOH	Sigma-Aldrich Fisher Scientific
Neutralisation Solution III	60 ml of 5M potassium acetate 11.5 ml of acetic acid 28.5 ml of dH_2O	Sigma-Aldrich Fisher Scientific
1X TBE	10.6g of Tris base 5.5g of boric acid 4 ml of 0.5M EDTA 96 ml of dH_2O	Sigma-Aldrich Fisher Scientific BDH

Reagents and chemicals:

Name	Company
100bp DNA ladder	New England Biolabs
100X BSA	New England Biolabs
10X DNaseI buffer	New England Biolabs
14.3 M β–mercaptoethanol	Sigma-Aldrich
1kb DNA ladder	New England Biolabs
2-log DNA ladder	New England Biolabs
2X RNA loading dye	Fermentas
6X DNA loading dye	New England Biolabs
Agarose powder	Bioline
Ampicillin 50 µg/ml	Sigma-Aldrich
Buffer 1/2/3/4/EcoRI	New England Biolabs
Chloroform	Fisher Scientific
Ethanol	Sigma-Aldrich
Ethidium bromide	
Isopropanol	Fisher Scientific
Phenol:chloroform:isoamylalcohol (25:24:1)	BDH
Primers	Invitrogen/ Sigma-Aldrich
RNA*later* solution	Ambion
RNaseZap	Ambion

Competent cells:

Name	Company
StrataClone SoloPack competent cells	Stratagene
Subcloning Efficiency DH5α competent cells	Invitrogen

Enzymes:

Name	Company
BamHI	New England Biolabs
DNaseI	Roche
EcoRI,	New England Biolabs
FastAP *alkaline phosphatase*	Fermentas
HindIII	New England Biolabs
NheI	New England Biolabs
XbaI	New England Biolabs
XhoI	New England Biolabs

2.1.3 Zebrafish Techniques

Micro-injection

Name	Company
Borosilicate glass capillaries 1mm O.D. x 0.58mm I.D.	Harward apparatus
Incubator	LEEC
Microloader (0.1 µl – 20 µl)	Eppordorf
Morpholino	Gene-Tools
MS-222 4g/L	Invitrogen
Flamming/brown micropipette puller Model P-97	Sutter instrument
Nitrogen (oxygen free)	BOC
Picospritzer micro-injector	Intracel
Fine forceps, jeweler #5	World Precision Instrument

Microscopes for micro-injection and image visualisation

Name	Company
Stereomicroscope Stemi SV 6	Zeiss
Stereomicroscope Stemi 2000-C	Zeiss
Eyepieces W-PI10x/23 Br. Foc	Zeiss
Cold Light Sources KL1500 LCD	Zeiss
Stereomicroscope SMZ1500	Nikon
Eyepieces C-W10X A/22	Nikon
Camera DS-5MC	Nikon
Digital sight DS-U1	Nikon
Mercury lamp for fluorescent	Nikon

2.1.4 Database, Software, and Online Programmes

Database	Website	Reference
Ensembl release 59 (August 2010)	http://www.ensembl.org	Flicek *et al*., 2010
iHOP	http://www.ihop-net.org/UniPub/iHOP/	Hoffmann & Valencia, 2004
NCBI blastp and tblastn searches	http://blast.ncbi.nlm.nih.gov/Blast.cgi	---
Uniprot blast search	http://www.uniprot.org/blast/	The UniProt Consortium, 2010

Software	Usage	Reference
Adobe Illustrator CS4	Figure edition	---
Adobe Photoshop CS4	Photo edition	---
ACT-2U Version 1.61	Fish image capture	---
ClustalX2	Format convertion and alignment display	Larkin *et al*., 2007
FinchTV version 1.4	Retrieving and revising sequencing results	---

Jalview 2.5.1 release	Alignment adjustment and edition	Waterhouse *et al*., 2009
MEGA 4.0	Construction of neighbour-joining tree and tree display and edition	Kumar *et al*., 2008 Tamura *et al*., 2007
Microsoft Excel 2007	Statistic analysis	---
PhyML 3.0	Construction of maximum likelihood tree	Guindon & Gascuel, 2003
ProtTest 2.4	Parameter estimation for tree construction	Abascal *et al*., 2005
Quantity One 4	Gel image capture	---
SimVector 4	Graphic production of plasmid vector	---

Online Programme	**Usage and Website**	**Reference**
ClustalW2	Multiple sequence alignment http://www.ebi.ac.uk/Tools/clustalw2/index.html	Larkin *et al*., 2007
EMBOSS open source	Sequence pairwise alignment http://www.ebi.ac.uk/Tools/emboss/align/	Rice *et al*., 2000
Eurofin PCR primer design tool	Primer design http://www.eurofinsdna.com/shop/eshop-features/design-calculation-tools/oligo-design/pcr-primer-design.html	---
First Glance in Jmol version 1.45	Structure model display http://molvis.sdsc.edu/fgij/	---
Genomicus version 58.01	Synteny Analysis http://www.dyogen.ens.fr/genomicus-58.01/cgi-bin/search.pl	Muffato *et al*., 2010
I-TASSER	3D structure prediction http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Roy *et al*., 2010 Zhang, 2009 Zhang, 2008
Jcoils	Coiled coil and Leucine zipper predictions http://fly.swmed.edu/jcoils/	Lupas *et al*., 1991 Bornberg-Bauerv *et al*., 1998
OligoAnalyzer 3.1	Prediction of melting temperature, home-/hetero-dimer, hairloop formation http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/	---
Pfam	Domain prediction http://pfam.sanger.ac.uk/	Finn *et al*., 2010
PhyML server	Maximum likelihood tree construction http://www.atgc-montpellier.fr/phyml/	---
PSIPRED server version 3.0	Secondary structure prediction http://bioinf.cs.ucl.ac.uk/psipred/	Jones, 1999 Bryson *et al*., 2005
SMART	Domain prediction http://smart.embl-heidelberg.de/	Schultz *et al*., 1998 Letunic *et al*., 2008

2.2 Methods

2.2.1 Molecular Techniques

Preparation of Plasmid

The plasmid DNA was extracted using QIAGEN QIAprep Spin Minprep Kit or Plasmid Midi Kit depending on the amount of high quality plasmid DNA required. Alternatively, phenol/chloroform extraction was used for colony screening and initial analysis, such as presence and orientation of insert DNA.

Minipreparation of Plasmid DNA

Under aseptic conditions, a colony was picked with sterile toothpick and inoculated into 5 ml of LB medium containing 100 µg/ml ampicillin. The culture was incubated at 37°C overnight with agitation at 250 rpm. 3 ml of overnight culture was pelleted at 7,000 rpm for 1 minute. The unused portion of the culture was stored at 4°C for either plasmid DNA preparation or being discarded into 2% Trigene. After centrifugation, medium was discarded into 2% Trigene whereas bacterial pellet was kept for next step (Method 1 or 2).

Method 1: QIAprep Spin Miniprep Kits

The bacterial pellet was re-suspended in 250 µl of Buffer P1 with brief vortexing. The bacterial cell was lysed by addition of 250 µl of Buffer P2, followed by inverting. The mixture was then neutralised by addition of 350 µl of Buffer N3, followed by inverting. Lysed cells were pelleted by centrifugation at 13,000 rpm for 10 minutes. The supernatant was applied to QIAprep spin column directly, followed by centrifuging at 13,000 rpm for 1 minute. The spin column was washed with 750 µl of Buffer PE and centrifuged for 1 minute. The flow-through was discarded and the spin column was further centrifuged for an additional 1 minute to remove residual wash buffer. Plasmid DNA was eluted by applying 30 µl of nuclease-free water to the centre of QIAprep spin column, standing for 1 minute, and centrifuging for 1 minute. The plasmid was then stored at -20°C until further investigation.

Method 2: Phenol/Chloroform Extraction

The bacterial pellet was re-suspended in 100 µl of Resuspension Solution I (see 2.1.2) by vortexing. The bacterial cell was then lysed by adding 150 µl of Lysis Solution II and incubating at room temperature for 5 minutes. Mixture was

neutralised by adding 150 µl of ice cold neutralisation Solution III and incubating on ice for 5 minutes. Lysed cells were pelleted by centrifugation at 13,000 rpm for 5 minutes. The supernatant was transferred to a fresh tube. RNA was removed by adding 3 µl of RNase A (10 mg/ml) and incubating at 37°C for 30 minutes. Equal volume of phenol/chloroform was added to the mixture. The mixture was vortexed for 30 seconds and centrifuged at 13,000 rpm for 1 minute to allow phase separation. The upper aqueous phase containing plasmid DNA was transferred to a fresh tube. Plasmid DNA was precipitated by incubating at -20°C for 10 minutes after the addition of two volumes of ice cold 100% ethanol. The plasmid was then pelleted by centrifuging at 13,000 rpm for 5 minutes and washed by the addition of 0.5 ml 70% ethanol, followed by centrifuging at 13,000 rpm for 2 minutes. The supernatant was discarded completely and the pellet was air-dried at room temperature for 10 minutes. The plasmid DNA pellet was re-dissolved in 30 µl nuclease-free water and then stored at -20°C.

Midipreparation of Plasmid DNA (QIAGEN Plasmid Midi Kit)

To prepare a starter culture, a colony was picked with sterile toothpick and inoculated into 3 ml of LB medium containing 100 µg/ml ampicillin. The starter culture was then incubated at 37°C for 6 hours with agitation at 250 rpm. The starter culture (or unused portion of culture mentioned above) was poured into 100 ml LB medium with 100 µg/ml ampicillin, followed by overnight incubation at 37°C with agitation. Bacterial cells were collected and centrifuged at 6000 rpm at 4°C for 15 minutes. The medium was discarded into 2% Trigene whereas the bacterial pellet was re-suspended in 4 ml Buffer P1. Bacterial cells were lysed by mixing with Buffer P2 thoroughly and incubating at room temperature for 5 minutes. The mixture was neutralised by mixing with Buffer P3 thoroughly and incubating on ice for 5 minutes. The lysed cells were pelleted by centrifuging at 13,000 rpm for 20 minutes at 4°C. At the same time, a QIAGEN-tip 100 was equilibrated by applying 4 ml of Buffer QBT and emptied by gravity flow. The supernatant after centrifugation was applied to the QIAGEN-tip 100 and allowed to enter the resin by gravity flow. The QIAGEN-tip 100 was then washed two times with 10 ml Buffer QC and emptied by gravity flow again. Plasmid DNA was eluted with 5 ml Buffer QC into a sterile 50 ml falcon tube. Plasmid DNA was precipitated by adding 3.5 ml of room temperature isopropanol and was pelleted by centrifuging at 6,000 rpm at 4°C for 1 hour. The supernatant

was carefully discarded. The plasmid DNA pellet was washed with 2 ml room temperature 70% ethanol, followed by centrifuging at 13,000 rpm at 4°C for 10 minutes. The supernatant was carefully discarded. The plasmid DNA pellet was air-dried, re-dissolved in nuclease-free water and further diluted to 1 µg/µl with nuclease-free water. Plasmid was stored at -20°C.

Restriction Enzymes Digestion

In order to linearise the plasmid DNA or to check the orientation and the presence of target insert, plasmid DNA was digested with one or two restriction enzymes, respectively. Typically, a 20 µl reaction was prepared by mixing 1 µg DNA plasmid, 2 µl of 10X digestion buffer, 0.2 µl of 100X BSA, 5 U of each restriction enzyme, and nuclease-free water. The reaction was then incubated at 37°C for 1 hour. If needed, the reaction was simply scaled up to 50uL with the use of 10 U of each restriction enzyme. The reaction was incubated at 37°C for 3 hours prior to gel electrophoresis.

Gel Electrophoresis

1%-1.5% gels were prepared by mixing 30 ml of TBE buffer with 0.3 g-0.45 g agarose powder followed by heating. After cooling down the gel, 5% v/v ethidium bromide was added and mixed thoroughly with gel solution, followed by pouring the mixture into the gel tank. A comb was inserted into the gel to prepare the loading wells. The sample to be loaded was mixed with 6X loading dye and nuclease-free water. After loading the sample into the wells, electrophoresis was performed at 80V for 35-60 minutes. Gel was then examined under UV light.

DNA Purification

<u>Gel Extraction</u> (QIAquick Gel Extraction Kit)

The DNA to be purified was loaded onto 1% agarose gel and run at 80V for 1 hour. The gel with target fragment was cut with sharp scalpel under low intensity UV light. The gel slice was weighed and put into an eppordorf tube, followed by dissolving in 100 µl Buffer QG per 100 mg agarose gel and incubating at 50°C for 10 minutes. The mixture was vortexed every 2 minutes until the gel completely dissolved. All of mixture was added into QIAquick column, followed by centrifuging at 13,000 rpm for 1 minute. The flow-through was discarded. The column was then

washed with 0.75 ml of Buffer PE, followed by centrifuging for 1 minute. The flow-through was discarded. The column was centrifuged for further 1 minute. DNA was eluted with 30 µl of nuclease-free water. The concentration and purity of DNA were examined by gel electrophoresis and/or using nanodrop measurement (see below).

Phenol-chloroform Purification

Half volume of phenol:chloroform:isoamylalcohol (25:24:1) was mixed with the sample to be purified. The mixture was vortexed for 30 seconds, centrifuged at 13,000 rpm for 3 minutes to allow phase separation. The bottom layer was removed. These steps were repeated once. Then, 1 volume of chloroform was added and mixture was vortexed for 30 seconds. Bottom layer was removed again. Addition of chloroform was repeated once more. The aqueous layer was mixed with 1/10 volume of 3M sodium acetate and 2½ volume of 100% ethanol and then stored at -20°C for 30 minutes for precipitation. Nucleic acids were pelleted by centrifuging at 13,000 rpm for 30 minutes at 4°C and were rinsed with addition of 200 µl of ice cold 70% ethanol and centrifugation at 13,000 rpm for 5 minutes at 4°C. Ethanol was completely removed and the pellet was air-dried for approximate 10 minutes. The pellet was re-suspended into 30 µl of nuclease-free water and stored at -20°C.

Measurement of Nucleic Acid Concentration and Purity

A nanodrop spectrophotometer was used to measure the concentrations and purities of DNA, RNA, and Morpholino. The nanodrop platform was cleaned with 70% ethanol, followed by nuclease-free water. The instrument was then initiated and blanked with 1.5 µl of nuclease-free water. 1.5 µl of sample nucleic acid was loaded onto the platform and its concentration measured. The purity was indicated by the A260/280 ratio. High purity DNA shows >1.8 whereas high purity RNA shows >2.0. In case of measuring morpholino concentration, 1.5 µl of 0.1N HCl served as blank instead of nuclease-free water. 0.1N HCl diluted morpholino solution was measured its absorbance at 265nm. A morpholino specific constant was used to calculate the actual concentration.

Primer Design and Primer Handling

Primers were designed by Eurofin PCR primer design tool with all default settings. Hairpin structure, self- and hetero-dimer formation, and melting

temperatures of primers were predicted by OligoAnalyzer v3.1 with the settings of 0.25 μM for [Oligo], 60 mM for [Na^+], 1.5 mM for [Mg^{2+}], and 0.2 mM for [dNTPs]. Primers were synthesised by commercial companies. Different amount of nuclease-free water was added into each desalted and lyophilisated primer to prepare 100 μM stock solution. The working concentration was 10 μM which was diluted from the stocks. Both stock and working solutions were stored at -20°C.

Polymerase Chain Reaction (PCR)

Standard PCR

A typical 20 μl reaction was prepared by mixing 1 μl of cDNA or 20ng DNA, 10 μl of REDTaq Ready Mix, 1 μl of 10 μM forward primer, 1 μl of 10 μM reverse primer and nuclease-free water. Initial denaturation was carried out at 94°C for 3 minutes. Target DNA was amplified by 35 cycles of template denaturation at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds and extension at 72°C for 1 minute. Final extension was performed at 72°C for 4 minutes, followed by incubation at 4°C. The programme is shown below. Finally, 1% or 1.5% agarose gel electrophoresis was carried out with the PCR reactions. Primer sets used are shown in Table 4.1-3.

Standard PCR programme:

94°C 3 minutes

94°C 30 seconds
55°C 30 seconds } 35 Cycles
72°C 1 minute

72°C 4 minutes

4 ^{o}C ∞

High Fidelity PCR

In order to prepare SWAP70 gene for restriction digestion and sub-cloning into pSC-B-amp/kan, high fidelity PCR was carried out using *Pfu* DNA polymerase. A 50 μl reaction was prepared by mixing 20ng plasmid DNA, 5 μl of 10X *Pfu* PCR buffer, 3 μl of 10 mM dNTPs, 2.5 μl of 10 μM forward primer, 2.5 μl of 10μM reverse primer, 1 μl of *PfuTurbo* DNA Polymerase, and nuclease-free water. Initial denaturation was carried out at 95°C for 2 minutes. Target DNA was amplified by 35 cycles of template denaturation at 95°C for 30 seconds, primer annealing at 60°C for

30 seconds and extension at 72°C for 2 minute. A final extension was performed at 72°C for 10 minutes, followed by incubation at 4°C. The programme is shown below. Finally, 1% agarose gel electrophoresis was carried out after the PCR reactions. Primer sequences are shown in Table 2.4.

Pfu PCR programme:

95°C 2 minutes
95°C 30 seconds
60°C 30 seconds } 35 Cycles
72°C 2 minutes
72°C 10 minutes
4 °C ∞

Colony PCR

To determine if colonies contain target inserts, colony PCR was performed. For each reaction, a 15 µl reaction was prepared by mixing 7.5 µl of REDTaq ReadyMix, 0.2 µM forward primer, 0.2 µM reverse primer, and nuclease-free water. Colonies were picked up by pipette tips and streaked on LB agar plate containing 100 µg/ml ampillcilin followed by immersing the tips into PCR reactions. All the reactions were put into the PCR machine. The tips were then discarded into 2% Trigene. After the temperature of the PCR reached to 94°C, bacterial cells were lysed by heating at 94°C for 10 minutes. Target DNA was amplified by 30 cycles of template denaturation at 94°C for 1 minute, primer annealing at 60°C for 40 seconds and extension at 72°C for 1 minute. A final extension was performed at 72°C for 5 minutes, followed by incubation at 4°C. The programme is shown below. Finally, 1% agarose gel electrophoresis was carried out after the PCR reactions.

Colony PCR programme:

94°C 10 minutes
94°C 1 minute
60°C 40 seconds } 30 Cycles
72°C 1 minute
72°C 5 minutes
4 °C ∞

Primer Name	Primer Sequence (5' to 3')	Product Size (bp)	Annealing Temperature
63599 Exon 1 FP	GTTTACCGCTCTGGATCTTG	1711	55
63599 Last Exon RP	TGATCTATTCTTGTTCACCTGC		
LOC Exon 1 FP	AGTGTCCAAATCACAACTCAAG	1538	55
LOC Last Exon RP	TAACACTCGTACTGGCCTG		
15g12.1 Exon 1 FP	GAGTTGCTGAAGTCTGTCTG	1762	55
15g12.1 Last Exon	TGGGACAACCGCTCTTTAC		

Table 2.1 Primer sets used for amplifying transcripts from exon 1 to last exon of *zgc:63599*, *LOC570940*, and *si:dkeyp-15g12.1.*

Primer Name	Primer Sequence (5' to 3')	Product Size (bp)	Annealing Temperature
SWAP70 Exon 6 FP	AGCTTTGAAATCAGCGCC	336	55
SWAP70 Exon 8 RP	CTCTTTCTCCATTTCCAGCC		
63599 Exon 11 FP	CCTGCTTAAACTCGTTCAACC	108	55
63599 Exon 12 RP	TGCCACATCTTCTCCAGTTC		
Def6 Exon 10 FP	GCAAGCACCAATGTTAAACAC	202	55
Def6 Exon 11 RP	CACACCCTCCTCTACTTTCC		
LOC Exon 1 FP	AGTGTCCAAATCACAACTCAAG	187	55
LOC Exon 3 RP	TTTAACAAACGTGCCCTCC		
15g12.1 Exon 4 FP	TCTGCCTGTTCAACCTACTC	192	55
15g12.1 Exon 5 RP	TCAAATGCTCCAAAAACTCCC		
EF1α RP	GCATCAAGGGCATCAAGAAG	555	55
EF1α FP	GAGAAGGAAGCCGCTGAGAT		

Table 2.2 Primer sets used for determining the expression profiles of *def6* paralogues in zebrafish.

Primer Name	Primer Sequence (5' to 3')	Product Size (bp)	Annealing Temperature
SWAP70 Exon 1 FP	TGTGGATAAAAGTGGGAAAGTG	447	55
SWAP70 Intron 1 RP	AGCAAAAGAAGCAGCAAAAC		
SWAP70 Exon 1 FP	TGTGGATAAAAGTGGGAAAGTG	148	55
SWAP70 Exon 2 RP	GTATCCCTGATTGGAAACCG		
SWAP70 Exon 5 FP	GACACAAAAGAAAGAACTGGAC	521	55
SWAP70 Exon 8 RP	CTCTTTCTCCATTTCCAGCC		
SWAP70 Exon 6 FP	AGCTTTGAAATCAGCGCC	336	55
SWAP70 Exon 8 RP	CTCTTTCTCCATTTCCAGCC		

Table 2.3 Primer sets used for determining the efficiency of morpholino-mediated knockdown of SWAP70 experiment.

Primer Name	Primer Sequence (5' to 3')	Product Size (bp)	Annealing Temperature
SWAP70 + *XhoI* FP	*CTCGAG* GTCAGCAGACGGAGAAACTA	1963	60
SWAP70 + *BamHI* RP	*GGATCC* AGAGACAGGAGCGATGAGAC		

Table 2.4 Primers used for introducing restriction sites to SWAP70 PCR product for subcloning the open reading frame into pSC-B-amp/kan plasmid.

Constructions of Vectors

Preparation of pβUT3+EGFP Vector

pEGFP-C1 and pβUT3 are shown in Figure 2.1A and B, respectively. pEGFP-C1 plasmid was linearised by *NheI*. The EGFP fragment was released from pEGFP-C1 by further digestion with *BamHI*. pβUT3 was digested by *XbaI*, which is a compatible enzyme to *NheI*, and *BamHI* to produce two cohesive ends. The pβUT3+EGFP vector (Figure 2.1C) was then prepared by ligation of EGFP and pβUT3.

Preparation of SWAP70 Insert

pExpress1+SWAP70 is shown in Figure 2.1D. A forward primer with an *XhoI* site, which was designed to allow in-frame cloning, and a reverse primer with a *BamHI* site were used to amplify the SWAP70 cDNA sequence from an pExpress1+SWAP70 plasmid using *Pfu* DNA polymerase. The SWAP70 PCR amplicon (Figure 2.1E) was then sub-cloned into a pSC-B-amp/kan plasmid (Figure 2.1F). *XhoI* and *BamHI* were used to digest the pSC-B-amp/kan plasmid to release the SWAP70 fragment. The SWAP70 fragment was gel extracted and purified.

pβUT3+EGFP+SWAP70 Plasmid

The *XhoI* and *BamHI* Double digested pβUT3+EGFP vector was firstly dephosphorylated and then ligated with gel extracted SWAP70. The pβUT3+EGFP+SWAP70 plasmid (Figure 2.1G) was dispatched to Geneservice for sequencing to examine the success of in-frame cloning.

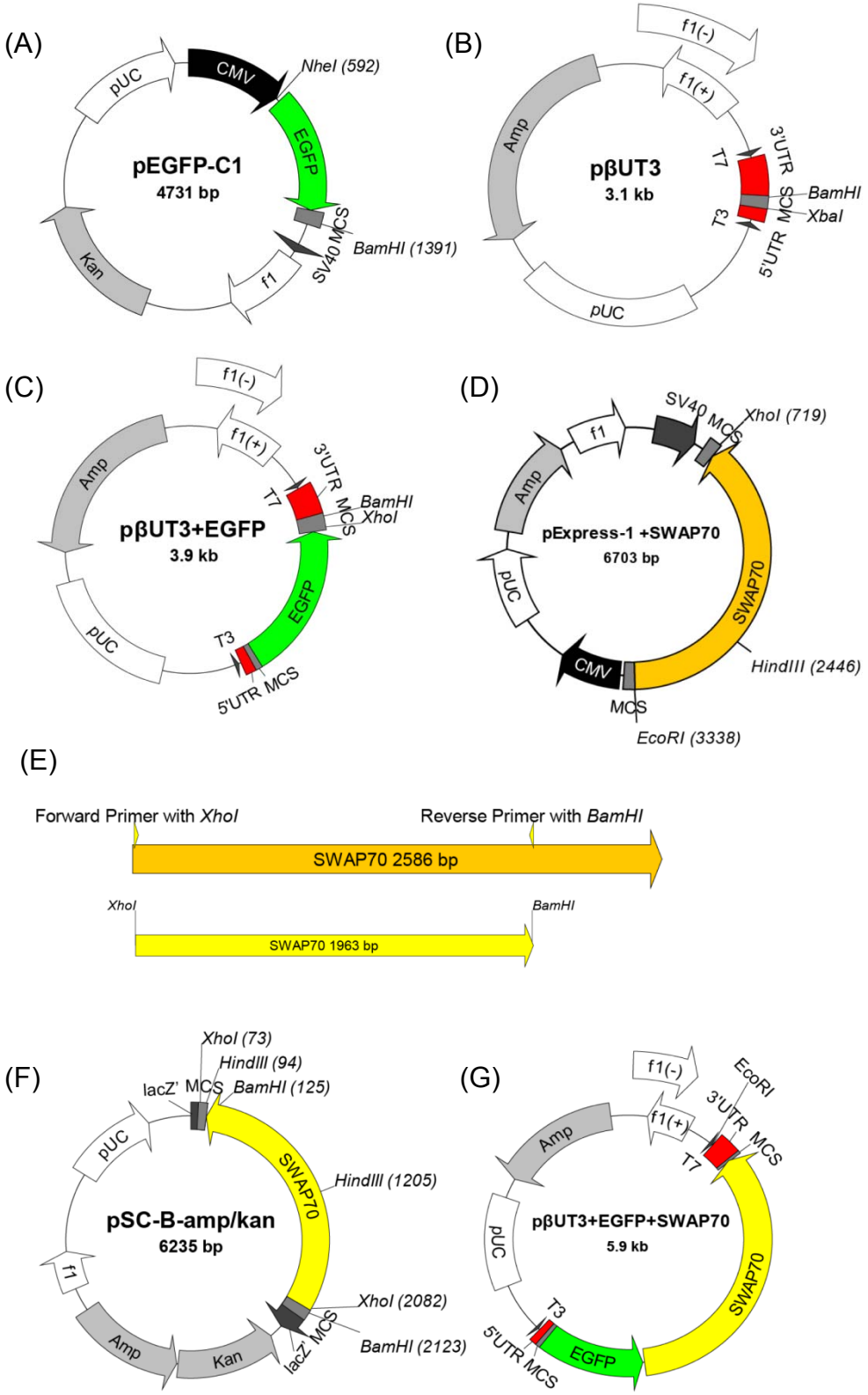


Figure 2-1 Plasmid Constructions. (A) pEGFP-C1 vector. (B) pβUT3 vector. (C) pβUT3+EGFP vector. (D) pExpress1+SWAP70 vector. (E) *Pfu* PCR amplification. (F) pSC-B-amp/kan vector. (G) pβUT3+EGFP+SWAP70 vector.

Molecular Cloning

Preparation of Vector DNA

pβUT3 and pβUT3+EGFP vectors were digested by *XbaI* and *BamHI*, and *XhoI* and *BamHI*, respectively. Cloning vector DNA was dephosphorylated by *alkaline phosphatase* to prevent self-ligation and to improve the ligation efficiency. A 20 µl reaction was prepared by mixing 1 µg linear plasmid DNA with 2 µl of 10X FastAP reaction buffer, 1 µl of FastAP thermosensitive *alkaline phosphatase*, and nuclease-free water. The dephosphorylation reaction was incubated at 37°C for 10 min, followed by inactivation at 75°C for 5 minutes. The concentration of dephosphorylated vector DNA was 50 ng/µl.

Preparation of Insert DNA

XhoI and *BamHI* for pBC-B-amp/kan, and *NheI* and *BamHI* for pEGFP-C1 were used. Digested plasmids were all loaded onto a 1% agarose gel. The fragments of the *SWAP70* gene and *EGFP* gene were cut out from the gel under UV visualization and further purified using a gel extraction kit. Concentrations of the fragments were measured by Nanodrop.

Ligation of Vector DNA and Insert DNA

EGFP was ligated into pβUT3 to produce pβUT3+EGFP plasmid. *SWAP70* was ligated into pβUT3+EGFP to produce pβUT3+EGFP+SWAP70 plasmid. A 20 µl quick ligation reaction was prepared by mixing 100 ng of dephosphorylated vector DNA with 3-fold molar excess of insert DNA, 10 µl of 2X *Quick T4 Ligase* reaction buffer, 1 µl of *Quick T4 Ligase*, and nuclease-free water. The ligation reaction was incubated at room temperature for 5 min followed by chilling on ice.

Transformation

Half of the ligation reaction was used to transform into DH5α competent cell by heat-shock. The ligation reaction was added into a tube of defrosted competent cells and incubated on ice for 30 minutes. Transformation was then carried out by heating competent cells with the ligation mix at 42°C for 20 seconds followed by chilling on ice for 2 minutes. 200 µl of SOC medium was added to the competent cells and the cells were incubated at 37°C with agitation at 250 rpm for 1 hour. Half

of the competent cells were spread on an LB agar containing 100 µg/ml ampicillin and incubated at 37°C overnight.

Colony Screening

The colonies were examined with either colony PCR or plasmid mini-preparation using phenol/chloroform followed by restriction digestion. Plasmid DNA of recombinant clones was sent for sequencing to confirm the success of the in-frame cloning.

StrataClone Blunt PCR Cloning

Pfu PCR reaction products were 1:10 diluted. The ligation reaction mixture was prepared by mixing 3 µl of StrataClone Blunt Cloning Buffer, 2 µl of diluted PCR reaction, and 1 µl StrataClone Blunt Vector Mix amp/kan. The ligation reaction was then incubated at room temperature for 5 minutes followed by putting on ice. StrataClone Solo Pack competent cells were thawed on ice and 1 µl of the cloning reaction mixture was added into competent cells. The mixture was incubated on ice for 20 minutes. Heat-shock was carried out by heating at 42°C for 45 seconds, followed by incubation on ice for 2 minutes. 250 µl of pre-warmed LB medium was added to the competent cells. The mixture was incubated at 37°C for at least 1 hour with agitation at 250 rotations per minute. 16 µl of 50 mg/ml X-gal was spread on LB-amplicllin agar plate before spreading 150 µl of competent cells on the plate. Plate was incubated at 37°C overnight. After incubation, only white colonies were picked for colony screening.

Sequencing

DNA samples to be sequenced with specific primers were sent to geneservice. Sequencing spectra were checked to ensure the correct nucleotide annotations.

Preparation of Capped mRNA for Microinjection

After the in-frame cloning of pBUT3+EGFP+SWAP70 was confirmed, the plasmid DNA was linearised using *EcoRI* and purified using the phenol/chloroform purification method. Capped mRNA was synthesised using Ambion mMESSAGE mMACHINE T3 kit with linearised plasmid as template. A 20 µl reaction was

prepared by mixing 1 µg of linerised DNA, 10 µl of 2X NTP/CAP, 2 µl of 10X reaction buffer, 2 µl of T3 enzyme mix, and nuclease-free water. The reaction was incubated at 37°C for 2 hours, followed by addition of 1 µl of *DNaseI* and incubated at 37°C for a further 15minutes. Gel electrophoresis was performed with 1 µl of the reaction to confirm the success of reaction. Synthesis was then stopped by the addition of 115 µl of nuclease-free water and 15 µl of ammonium acetate stop solution. The capped mRNA was extracted with 1 volume of phenol/chloroform, followed by vortexing for 30 seconds and spinning at 13,000rpm for 3 minutes. The bottom layer was removed. The capped mRNA was further extracted with 1 volume of chloroform, followed by vortexing for 30 seconds and spinning at 13,000 rpm for 3 minutes. The bottom layer was removed. The capped mRNA was precipitated using 1 volume of isopropanol and storing at -20°C for 30 minutes. The Capped mRNA was pelleted by spinning at 13,000 rpm for 15 minutes at 4°C. All the isopropanol was removed. The pellet was then air-dried at room temperature for 15 minutes and re-suspended in 30 µl nuclease-free water. The concentration and purity were checked using a nanodrop spectrophotometer and gel electrophoresis. The Capped mRNA was stored at -80°C.

Reverse Transcription

Total RNA Extraction

The intact embryos (50-100 embryos) were stabilised in RNA*later* RNA stabilization reagent (see 2.1.2) at room temperature for maximum 7 days. Before homogenising of the embryos, RNA*later* reagent was completely removed. 600 µl of buffer RLT Plus was previously mixed with 6 µl of 14.3 M β–mercaptoethanol and then added into the tube containing embryos. An autoclaved pestle was used to disrupt the embryos, followed by using the QIAshredder homogenizer. Lysed embryos were centrifuged at 13,000rpm for 2 min and for a further 3 min. The supernatant was carefully removed to a new eppendorf tube. The supernatant was mixed with 600 µl of 70% ethanol. All the mixture was loaded onto an RNeasy spin column and centrifuged at 13000rpm for 15 seconds. The flowthrough was discarded. The column was washed by addition of 700 µl of Buffer RW1 and centrifugation at 13000 rpm for 15 seconds. The flowthrough was discarded again. The column was then washed with 500 µl of buffer RPE, followed by centrifugation at 13,000 rpm for 15 seconds. The flowthrough was discarded. Another 500 µl of Buffer RPE was used to wash the column and the membrane was dried by centrifuging at 13000rpm for 2

min. 20 µl nuclease-free water was added to elute the total RNA. Total RNA concentration and purity was checked by nanodrop spectrophotometer whereas the integrity was checked by gel electrophoresis. Total RNA was stored at -80°C.

DNaseI Treatment

DNaseI treatment was used to remove any DNA contamination from the isolated RNA. A 10 µl reaction was set up by mixing 1 µg total RNA, 1 µl 10X *DNaseI* reaction buffer, and 10 units of *DNaseI*. The reaction was incubated at 37°C for 15 minutes and stopped by addition of 1 µl of 2 mM EDTA and incubation at 65°C for 10 minutes in PCR machine.

First Strand cDNA Synthesis

DNaseI treated RNA was mixed with 500 ng of oligo$(dT)_{12-18}$ primer and 1 µl of 10mM dNTPs. The mixture was heated at 65°C for 5 min in a PCR machine, followed by quick chilling on ice for 2 minutes and brief centrifuging. The mixture was then mixed thoroughly with 4 µl of 5X First-Strand Buffer and 2 µl 0.1M DTT, followed by incubating at 42°C for 2 min. 1 µl of SuperScript™ II *reverse transcriptase* (200U) was added into the mixture by pipetting up and down. Reverse transcription was carried out at 42°C for 50 min, followed by 15 minutes at 70°C to stop the reaction. 1 µl of cDNA was used for PCR amplification and cDNA was stored at -20°C.

2.2.2 Bioinformatics Analysis

Data Acquisition

Def6 family genes in vertebrates and teleost fishes are shown in protein family ID ENSFM00250000001889 in Ensembl release 59 (August 2010). As the teleost genomes are mainly available in Ensembl database, the sequences of nucleotide and amino acid of teleosts were acquired from Ensembl whereas those of selected vertebrates were obtained from NCBI. Blastp and tblastn searches in NCBI and a blast search in Uniprot using zebrafish full length *def6* and human full length *DEF6* were carried out to identify *def6*-related genes in invertebrates and vertebrates which had not been yet annotated in Ensembl.

Exon-intron Structure

All the exon-intron information was obtained from Ensembl release 59 (August 2010). Exons were aligned and highlighted in grey according to the multiple amino acid sequence alignment.

Structure Prediction

Full length amino acid sequences were submitted to Pfam (Finn *et al*., 2010) and SMART (Schultz *et al*., 1998 and Letunic *et al*., 2008) for domain predictions. Sequences were also submitted to JCoils which is a Java implementation of NCoils (Lupas *et al*., 1991) and 2Zip (Bornberg-Bauerv *et al*., 1998) for coiled coil region predictions.

Multiple Amino Acid Sequence Alignment

Amino acid sequences in FASTA format were submitted into the ClustalW2 programme (Larkin *et al*., 2007) with all default settings. The alignment was viewed and 50% or 60% identity was coloured using Jalview 2.5.1 release (Waterhouse *et al*., 2009). A full length alignment was saved by 'print screen' and the colouring was further modified by Adobe Illustrator CS4. For tree construction, alignment was manually adjusted to remove divergent N-terminus and C-terminus sequences, and most of the gaps. The adjusted alignment was saved in FASTA format. MEGA4.0 was used to convert the alignment format to MEGA whereas ClustalX2 was used to convert the alignment format to PHYLIP.

Pairwise Analysis

Pairwaise analysis of two sequences in fasta format was performed using EMBOSS open source software analysis package (Rice *et al*., 2000). The settings were 10.0 for gap open, 0.5 for gap extend, and blosum50 for matrix. The percentages of identity, similarity and gap between amino acid sequences were recorded.

Secondary Structure Prediction

Secondary structures of full length amino acid sequences were predicted using PSIPRED method (Jones, 1999) with mask low complexity regions as filtering option at University College London server, version 3.0 (Bryson *et al*., 2005). The output PSIPRED results were manually modified and edited using Adobe Illustrator CS4.

Tertiary Structure Prediction

Three dimensional structures of full length amino acid sequences were predicted and generated using I-TASSER (Roy *et al*., 2010, Zhang, 2009 and Zhang, 2008). The predicted molecular contact and rainbow coloured models were visualised using First Glance in Jmol, version 1.45 and 'print screen' saved. The figures were manually modified and edited using Adobe Illustrator CS4.

Tree Construction

Estimation of Parameters for Tree Construction

ProtTest 2.4 (Abascal *et* al., 2005) was run with phylip format alignment to find out the best parameter settings for maximum likelihood tree constructions. For the programme options, BIONJ tree and slow optimization strategy were selected. All the substitute matrices and add-ons shown in the programme were tested with 4 gamma categories. After the programme running, the best model parameters were shown according to Akaike Information Chriterion (AIC) framework suggestion. The top ranked models and the model-averaged estimate of parameters were used for maximum likelihood tree constructions. The model-averaged estimate of gamma distribution (+G) value was also applied to neighbour-joining tree constructions.

Neighbour-joining Tree Construction

The alignment in FASTA format was converted into MEGA format for neighbour-joining tree construction using MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007). MEGA 4.0 settings are JTT+G as the substitution model, pairwise deletion, random seed, and 2000 bootstrapping. Tree generated was visualised and modified using MEGA 4.0.

Maximum Likelihood Tree Construction

Maximum likelihood tree was generated using PhyML 3.0 (Guindon & Gascuel, 2003) in PhyML server, http://www.atgc-montpellier.fr/phyml/ with the alignment in PHYLIP format. Substitution model was set according to ProtTest results. SPR tree improvement, optimisations of topology and branch lengths, and 1000 bootstrapping were also used. Tree generated was visualised and modified using MEGA 4.0. Adobe Illustrator CS4 was used to edit the trees.

Synteny Analysis

Genomicus version 58.01 (Muffato *et al.*, 2010) was used for synteny analysis. The server has integrated with Ensembl release 58 and thus the information is all based on Ensembl database. The DEF6 and SWAP70 syntenies were manually edited using Adobe Illustrator CS4.

2.2.3 Zebrafish Techniques

Zebrafish (*Danio rerio*) was maintained according to Westerfield (2000). For experiments, wild type embryos and morpholino/capped mRNA injected embryos were moved to a new fish Petri-dish with fresh aquarium water. Embryos were raised at 28.5°C.

2.2.4 Anti-sense Technology – Morpholino

Morpholino Design and Action

Two splice blocking morpholinos and one translation blocking morpholino against *swap70a* were designed and synthesised by Gene Tools. Morpholino sequences are shown in Table 2.5. The actions of two splice blocking morpholinos are shown in Figure 2.2. For the *swap70a* +Intron1 morpholino action, the first intron which contains an in-frame stop codon is included during splicing event. Truncated protein is then produced and thus the *swap70a* gene function is knocked-down *in vivo*. For the *swap70a* ∆Exon6 morpholino, the exon 6 is deleted during the splicing event which causes a frame-shift and thus a truncated protein is produced.

Morpholino Name	Constant[1]	Morpholino Type	Morpholino Sequence (5' to 3')
swap70a +Intron1 MO	31.11	Splice blocking (inclusion of intron 1)	AGAGCAAAACGACAAACCTTCAGCT
swap70a ΔExon6 MO	31.89	Splice blocking (exclusion of exon 6)	TGAATCATCTCTAACCTTGAATCCA
swap70a AUG MO	32.13	Translation blocking	TGAGAAGCTCGTCCCTTAGTCCCAT
Standard Control MO	32.39	Negative control morpholino	CCTCTTACCTCAGTTACAATTTATA

[1] Constant is used for calculating morpholino concentration

Table 2.5 Morpholinos used in this thesis

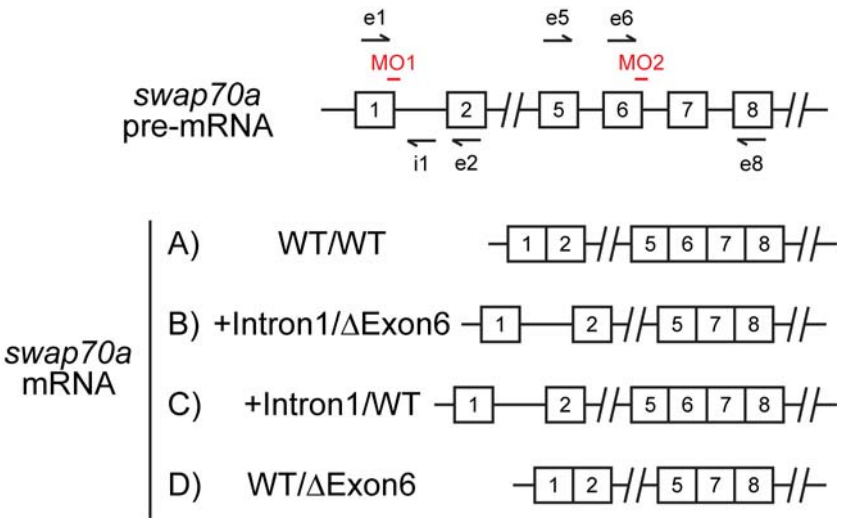


Figure 2-2 Morpholino Actions and Possible Outcomes. MO1 represents *swap70a* +Intron1 morpholino and MO2 represents *swap70a* ΔExon6 morpholino. Primers used for determining morpholino efficiencies are shown as black arrow. The boxes indicate exons and black lines indicate untranslated region and introns. Boxes and lines are not shown to scale. Injection of MO1 may cause outcomes A and C, injection of MO2 may cause outcomes A and D, and co-injection of MO1 and MO2 may cause outcomes A, B, C, and D.

Morpholino Handling

Lyophilised morpholino was dissolved in nuclease-free water to prepare 50ng/nl concentration stock. The morpholino solution was heated at 65°C for 5 minutes, followed by brief vortexing and centrifuging at 13,000 rpm for 1 minute, thrice. The morpholino was heated again at 65°C for less than 1 minute prior to aliquoting. The concentration was checked by nanodrop measurement. The heating steps which are critical to handle morpholino were repeated for diluting the morpholino. The stock and diluted morpholino sealed with parafilm was then stored in -20°C.

Measurement of Morpholino Concentration

The heating steps mentioned above were performed. While the morpholino solution was still hot, 1 µl of solution was added into 99 µl of 0.1N HCl. The 0.1N HCl diluted morpholino solution absorbance was measured at 265nm using the nanodrop spectrophotometer. The concentration was calculated by the equation shown below.

Morpholino concentration = A_{265} X specific constant X dilution factor

2.2.5 Microinjection

The injection needles were prepared using glass capillaries and a needle puller. The needle puller settings were P=500, Heat=295, pull=200, VEL=115, and time=115. For micro-injection, 3 µl of the solution to be injected into embryos was loaded into the injection needle. The injection needle was then fixed in the injection apparatus. The needle was broken using a fine forceps. The process was viewed under a microscope with the highest magnification (X50). The bubble size which equals 500 pl was calibrated by changing the pulse duration. Embryos at the 2-4 cell stage were aligned along the slide edge on a fish Petri-dish lid. The cells of embryos were aligned facing the slide edge to allow the needle to pass through the yolk during injection. Injected embryos were then moved to a fish Petri-dish with fresh aquarium water containing a trace amount of methyl blue. Wild type embryos were also moved to a new fish Petri-dish with fresh aquarium water. Wild type and injected embryos were incubated at 28.5^{o}C. At different developmental stages, embryos were taken out from the incubator for image capture and phenotypic analysis. Embryos were dechorinated at 24 hours post-fertilisation stage.

2.2.6 Phenotypic Analysis

The phenotypes of dechorinated embryos were determined manually and the images were captured. The body length was measured with a micrometer under microscopy with 10X magnification. Adobe photoshop CS4 was used to edit the images captured.

3. Identification and Analysis of *def6* Paralogues in Zebrafish

3.1 Data acquisition

According to the Ensembl database (http://www.ensembl.org/) with the version of Ensembl release 58 (May 2010), there are five Def6 paralogues identified in zebrafish (protein family ID ENSFM00250000001889). The details of the five paralogues along with different gene IDs in different databases are shown in Table 3.1. There are some alternative names for each gene; but in this thesis, the names of *SWAP70*, *zgc:63599*, *Def6*, *LOC570940*, and *si:dkeyp-15g12.1* are used.

Gene Name[1]	Gene Name[2]	Official Symbol in NCBI	Alternative Name[3]	NCBI Gene ID	Ensembl Gene ID	ZFIN Gene ID	UniProt ID
SWAP70	*si:dkey-8l13.4*	si:dkey-8l13.4	fc45f07/swp70/wu:fc45f07	562490	ENSDARG00000057286	ZDB-GENE-030131-3587	Q1LYE0
zgc:63599	*zgc:63599*	zgc:63599	MGC63599/LOC393645	393645	ENSDARG00000051819	ZDB-GENE-040426-1182	Q6PEH7
Def6	*zgc:63721*	zgc:63721	MGC63721/si:dkey-6n6.5	394015	ENSDARG00000012247	ZDB-GENE-040426-1246	Q7SYB5
LOC570940	*-novel-*	LOC570940	---	570940	ENSDARG00000044524	---	---
si:dkeyp-15g12.1	*si:dkeyp-15g12.1*	si:dkeyp-15g12.1	---	563690	ENSDARG00000034717	ZDB-GENE-060503-87	Q1L978

[1] The names which are used in this thesis.

[2] The names which are used in Goudevenou, 2010.

[3] The names which are shown in NCBI and the server of information hyperlinked over proteins (iHOP) (Hoffmann and Valencia, 2004).

Table 3.1 The Gene ID Summary of Def6 Paralogues in Zebrafish.

3.2 Verification of the cDNA sequences of def6 paralogues zgc:63599, LOC570940, and si:dkeyp-15g12.1

Def6 (NM_201040) and *SWAP70* (BC115308 and NM_001044986) cDNA sequences had been previously confirmed (Song *et al*., 2004, Strausberg *et al*., 2002 and Takada & Appel, 2010) and the expression pattern of both genes during zebrafish development had been determined by *in situ* hybridisation in our laboratory (Martin & Goudevenou unpublished data, 2007). The genes *zgc:63599* (NM_200672), *LOC570940* (XM_694476), and *si:dkeyp-15g12.1* (XM_687050), however, were only predicted based on computation analysis. Furthermore, the gene record of *LOC570940* was predicted by automated computational analysis based on a genomic sequence NW_001878313.1 and was annotated by GNOMON gene prediction method. During my analysis, this genomic sequence was discontinued and removed as a result of human review and standard genome annotation processing (NCBI staff, personal communication, July 2010).

Therefore, in order to determine whether the three Def6 paralogues, *zgc:63599*, *LOC570940*, and *si:dkeyp-15g12.1* are potentially functional genes, RT-PCR analysis was carried out. Using gene-specific primers in the first and last exons (see Table 2.1) and cDNA prepared from 72 hpf stage zebrafish embryo mRNA as template, all three genes appeared to be expressed at this time of development (data not shown but see also Figure 6.1). The amplicon fragment sizes of *LOC570940* (around 1500bp) and *si:dkeyp-15g12.1* (around 1700bp) were similar to the predicted sizes (1526bp and 1762bp, respectively). However, the fragment size of *zgc:63599* (around 1700bp) was different from the predicted size of 601bp. Nevertheless, sequence analysis of the purified amplicons showed that the predicted cDNA sequence for *si:dkeyp-15g12.1* was correct matching the predicted mRNA sequence (XM_687050) but the predicted sequence for *zgc:63599* (NM_200672) was incomplete (Figure 3.1). The partial coding sequences of *zgc:63599*, *LOC570940* and *si:dkeyp-15g12.1* determined here were submitted to Genbank and the accession numbers are HM752768, HM752767, and HM775399, respectively.

```
                 10        20        30        40        50        60        70        80        90       100       110       120       130       140       150       160
NM_200672   1 ATGATTTCCAAGGAAGAAATACTCAAACCCATATGGTACGCGTTTACCGCTCTGGATCTTGACCGAAACGGAAAAGTATCCAAATCCCAGTTAAAGGTTCTGTCGCATAACCTGTGCACCATACTGAAGATCCCTCATAACACTTCAGCCCTGGAGGAGCAC 162
HM752768    1 ------------------------------------------------------------------------------------------TTAAAGGTTCTGTCGCATAACCTGTGCACCATACTGAAGATCCCTCATAACACTTCAGCCCTGGAGGAGCAC 72

                170       180       190       200       210       220       230       240       250       260       270       280       290       300       310       320
NM_200672 163 TTTAAAGATGACGACGAAGGACCGGTGTCAACGCAAGGGTACATGCCCTACCTGAACACCTTCATACTTAGTAAGATCCAGCCCAACTTTGACTTTGTTGAGCTCAATAAGATGTGCTGGACTCTGTCTGCACCGAAGCACATCAACACCATGAATCTGCTC 324
HM752768   73 TTTAAAGATGACGACGAAGGACCGGTGTCAACGCAAGGGTACATGCCCTACCTGAACACCTTCATACTTAGTAAGATCCAGCCCAACTTTGACTTTGTTGAGCTCAATAAGATGTGCTGGACTCTGTCTGCACCGAAGCACATCAACACCATGAATCTGCTC 234

                330       340       350       360       370       380       390       400       410       420       430       440       450       460       470       480
NM_200672 325 ATGTCTGAAAAAGACGCCTTCAGGGTCTGGTGCATCTTCAACTTCCTTTCTGAAGATAAATATCCACTCACAATTGTCACAGAAGAGCTTGAGTATTTCCTGCGCAAGCTGTTGACGGCAATGGGTTGCAGTTGGAATGAAGGCCG---------------- 470
HM752768 235 ATGTCTGAAAAAGACGCCTTCAGGGTCTGGTGCATCTTCAACTTCCTTTCTGAAGATAAATATCCACTCACAATTGTCACAGAAGAGCTTGAGTATTTCCTGCGCAAGCTGTTGACGGCAATGGGTTGCAGTTGGAATGAAGGCCGGTTTGAGGATTATAAA 396

                490       500       510       520       530       540       550       560       570       580       590       600       610       620       630       640
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 397 ATGCATCTCAATGCAAAGAAACACCATCTGAATGCATGGGAGCTGATCGAACTTATTGACATGGGTAATTTCACTAAAGGCATCAACCCACAGACTGTCTCCATGGGCATCAATGAAGTCTTTCAGGAACTTGTAATGGATGTTATTAAGCAGGGCTATATG 558

                650       660       670       680       690       700       710       720       730       740       750       760       770       780       790       800
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 559 ATGAAAAAAGGCC-ATAAGAGGAAGAACTGGACAGAGCGGTGGTTCGAGCTCCATCTAAACTACATGTCGTATTATGTGAGCGAAGACCTTGTAGAGCAGAAGGGGTGTATTCCGCTTGATCGCAACTGCTGCGTTGAGTCCATGCCAGACAAAGACGGAAA 719

                820       830       840       850       860       870       880       890       900       910       920       930       940       950       960       970
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 720 GAAAAACCTCTTCATTATAAGATGTGTTGAAAAAAGTTTTGAGATCAGTGCATCGGATAAAAAGGCAAAACAGGAGTGGATTCAAGCCATCCAAGACTGCATTGTCCGCATAAGGCAGGGCCTTTCCTCCGCACACCGAGAATCCAGACAGAAGCGCAGAGA 881

                980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100      1110      1120      1130
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 882 GCTGAGAAACAGGATAAAAGCTGAGCAGGAGATCTTGGAGAACAGAATGAAACAGCTACAGTTGGCCAACGAGAGCAAGCAGGCCCAGCTGGAGGCCCTGAGTAAGGACCTGAATAAAGCCACAGCTAGGGGTGATCTGGAGGAACAGAGAAGACTTCAGGC 1043

               1140      1150      1160      1170      1180      1190      1200      1210      1220      1230      1240      1250      1260      1270      1280      1290
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 1044 CCAAAAAGAACTGCTGGAGCAATACAAACGAGACCTGGAGCAAGAGAAAATGGCTCGAATGCAAATGGAGGAGCAGGTGATTGAGAAGTCGAATGAAGTGGAGCAGTACTGGCAGCGTATGCAGGAACTGGAGGAAATGTACCTTCAGCTCAAACAGGCTCT 1205

               1300      1310      1320      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440      1450
NM_200672   ------------------------------------------------------------------------------------------------------------------------------------------------------------------
HM752768 1206 GCAGGATGAGAAGCAGGCCAAGGAGGATGAAGAAATACTGCGCAAACTGCAGGCCAGACTGCTACAAGAGGAGGAAAACAAGAGACTTGAATTAGAGCAGATTCACCTGCAGCAGCAACGGGTCTTATCCCAGTCCCAGAAGGAGAAGGAGGAGCTGGCGAA 1367

               1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560      1570      1580      1590      1600      1610
NM_200672   471 ------------------------------------------------------------------------------------------------------------------------------TAAGAAGACAAAGACCTGGAAACACAAAGTCTCTAAACA 509
HM752768 1368 GGAGCGGGAAGAGAAGGAGCGAGCCCTGCAAGCGGCACAGGAGCAGCTGGAGAGCCTGAAAAGACAGAGGGAGGGAGCGGAGGAGGAATACATGGCTGTAACAAAGAAACTGGAACGAGCAAGTAAGAAGACAAAGACCTGGAAACACAAAGTCTCTAAACA 1529

               1630      1640      1650      1660      1670      1680      1690      1700      1710      1720      1730      1740      1750      1760
NM_200672 510 TGAAGGCCTGCTTAAACTCGTTCAACCAGGCGGCAAGACACCCAGTAAAATGACCAACTGGGGACAGGCGGCGTTCACTGAATCAGAGCTGAATGAACTGGAGAAGATGTGGCAGGTGAACAAGAATAGATCAGAAGAAGAGTAA 654
HM752768 1530 TGAAGGCCTGCTTAAACTCGTTCAACCAGGCGGCAAGACACCCAGTAAAATGACCAACTGGGGACAGGCGGCGTTCACTGAATCAGAG------------------------------------------------------------ 1617
```

Figure 3-1 The Nucleotide Sequence Alignment of *zgc:63599*. The upper one is the predicted mRNA sequence (NM_200672) and the lower one is the partial coding cDNA sequence (HM752768) determined in this thesis. The 100% match regions are highlighted in yellow and the nucleotide sequences of HM752768 that is missing in NM_200672 are indicated in red.

3.3 Novel exon-intron annotations for zgc:63599 and LOC570940 genes

Comparing the obtained partial coding sequences to the information in Ensembl database, the exon-intron structure of *si:dkeyp-15g12.1* (ENSDARG00000034717) was confirmed. The sequence of *si:dkeyp-15g12.1* (HM775399) showed 100% match to the genomic zebrafish nucleotide sequence from clone DKEYP-15G12 in linkage group 4 (CR382295.10). However, the exon-intron structures of *zgc:63599* and *LOC570940* were found to be incorrect in ENSDARG00000051819 and ENSDARG00000044524, respectively. Nucleotide blast searches indicated that the cDNA sequence of *zgc:63599* (HM752768) and *LOC570940* (HM752767) obtained here exhibited 100% match to the genomic zebrafish nucleotide sequence from clone CH211-168K15 (AL935125.13) and clone CH73-352P18 in linkage group 22 (FP067424.7), respectively. Accordingly the exon-intron structure for both genes could be determined correcting the Ensembl data ENSDARG00000051819 and ENSDARG00000044524. As summarised in Figure 3.2, *zgc:63599* contains 12 exons in total rather than the predicted 6 and *LOC570940* has 11 exons rather than the predicted 17.

The novel exon-intron annotations for *zgc:63599* and *LOC570940* were submitted to the Third Party Annotation (TPA) database in NCBI. The accession numbers are BK007099 and BK007098, respectively.

3.4 Exon-intron structures are highly conserved among Def6 paralogues

All five Def6 paralogues have highly similar exon-intron structures. With the exception of the 5'-untranslated region (UTR), exons 1, 2, 5, 6 and 8/9 have identical lengths in all paralogues (96, 141, 147, 109 and 167 nucleotides, respectively). Exons 3 and 4 exhibit variable lengths and exon 7 in *Def6* and *LOC570940* is represented in 2 exons (exons 7 & 8) in *SWAP70*, *zgc:63599* and *si:dkeyp-15g12.1*. Similarly, exons 9 and 10 of *Def6* and *LOC570940* as well as exons 10 and 11 of *SWAP70* and *zgc:63599* are split in *si:dkeyp-15g12.1* (Figure 3.2).

Importantly, it appears that *SWAP70* and *zgc:63599* have 12 exons of identical length that is different from Def6 and LOC570940, that both have 11 exons also of identical length suggesting that these pairs of paralogues are distinct from each other. Similarly, *si:dkeyp-15g12.1* has an exon–intron structure that is distinct from the other 4 paralogues.

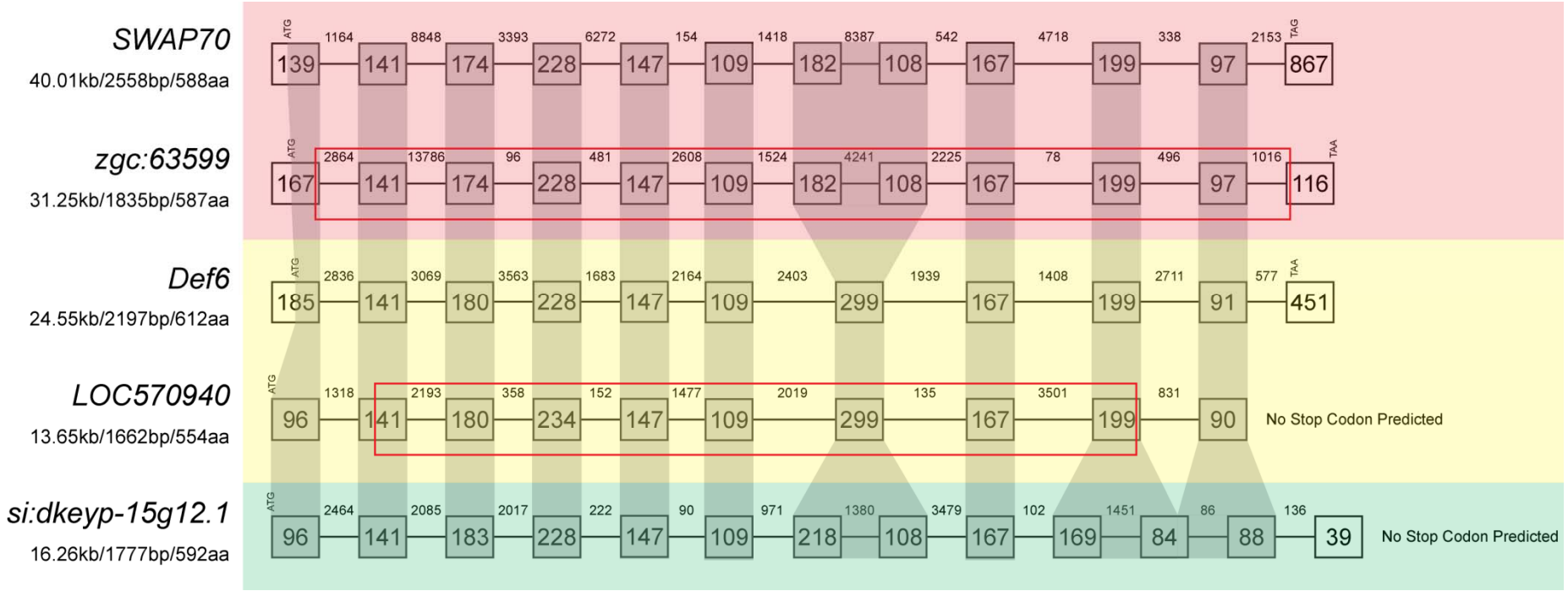


Figure 3-2 Exon-intron Structure Comparisons of Def6 Paralogues in Zebrafish. The exon-intron structures of *SWAP70*, *Def6* and *si:dkeyp-15g12.1* are based on Ensembl transcript ID ENSDART00000100105, ENSDART00000005359, and ENSDART00000101056, respectively. For *zgc:63599* and *LOC570940*, the beginning and the end of the exon-intron structures are based on Ensembl transcript IDs ENSDART00000061302 and ENSDART00000065377 while other regions are based on the novel exon-intron annotations (BK007099 and BK007098) which are indicated with red boxes. The lengths of genomic sequences, transcript sequences, and amino acid sequences are shown under the gene names. The locations of the start and stop codons are shown as well. Small black blocks represent exons and the black lines represent introns. Both black blocks and lines are not shown to scale. Exons with aligned nucleotide lengths are highlighted in grey.

3.5 Def6 Paralogues share highly similar domain structures

Domain structure comparison was carried out using three domain prediction programmes, Pfam 24.0 (Finn *et al*., 2010), SMART (Schultz *et al*., 1998 and Letunic *et al*., 2008), and JCoil which is a Java implementation of NCoils (Lupas *et al*., 1991) and 2Zip (Bornberg-Bauerv *et al*., 1998).

Pfam and SMART predicted a Pleckstrin Homology (PH) domain in the middle of each sequence with high confidence (Figure 3.3). At the N-terminal ends, SWAP70 and zgc:63599 are insignificantly predicted to have EF-hand motif by Pfam and zgc:63599 and LOC570940 are predicted by SMART with low confidences. However, both prediction tools do not predict any EF hand for Def6 and si:dkeyp-15g12.1 which are highly conserved in the N-terminus (see Figure 3.4). JCoil was used to predict the coiled coil in those regions, as there is no identifiable domain predicted in any sequence after the PH domain. There are two or three coiled coil regions identified in the sequences. SWAP70, zgc:63599, and Def6 have three coiled coil regions whereas LOC570940 and si:dkeyp-15g12.1 only have two. But, Def6 and LOC570940 contain two blocks of sequence which are at the similar positions with similar sizes. The si:dkeyp-15g12.1 has only two small coiled coil regions and there is a long gap between PH domain and the first coiled coil region.

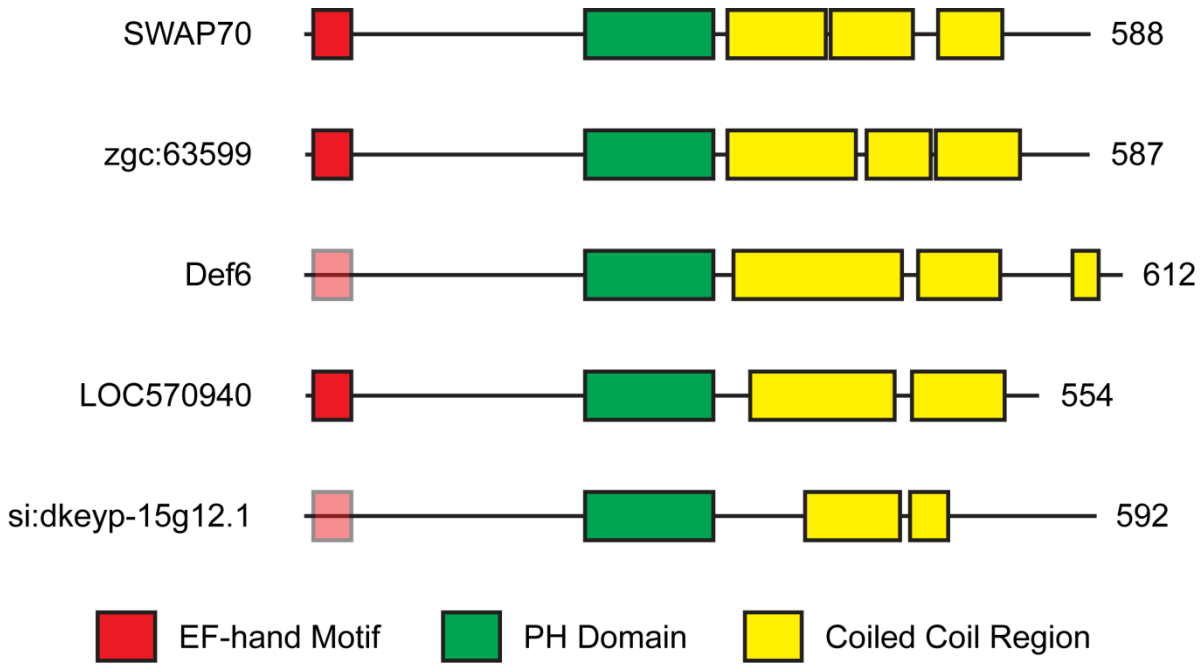


Figure 3-3 Domain Structures of Def6 Paralogues in Zebrafish. EF-hand motif and PH domain were predicted using Pfam (Finn *et al*., 2010) and SMART (Schultz *et al*., 1998 and Letunic *et al*., 2008). Coiled coil regions were predicted using JCoils which is a Java implementation of NCoils (Lupas *et al*., 1991) and 2Zip (Bornberg-Bauerv *et al*., 1998). EF-hand motifs of Def6 and si:dkeyp-15g12.1 were proposed based on the highly conversed N-terminal end among Def6 paralogues (Figure 3.4). The lengths of the amino acid sequences are shown next to the end of C-terminus.

3.6 N-terminus and Pleckstrin Homology (PH) domain region are highly conserved among Def6 paralogues

Multiple amino acid sequence alignment of the Def6 paralogues was generated using ClustalW2 (Larkin *et al*., 2007) and the alignment regions above 50% identity were coloured using Jalview 2.5.1 release (Waterhouse *et al*., 2009; Figure 3.3). There are three different conservation levels that are indicated with dark (100% identical), light (80% identical), and very light (60% identical). Different colours were used to indicate the various domains that had been described earlier for human DEF6 (Mavrakis *et al*., 2004).

The N-terminal region from 1 to 72 amino acids containing a putative EF-hand motif is extremely conserved (60% identical) and the Pleckstrin-Homology domain (position 214-311) are highly conserved (32% identical). Amino acid sequence from position 73 to 213 (DSH) is less conserved (20% identical) and the C-terminal end (position 314-644) encoding a Dbl-homology–like (DHL) domain shows only 10% identity and 30% similarity.

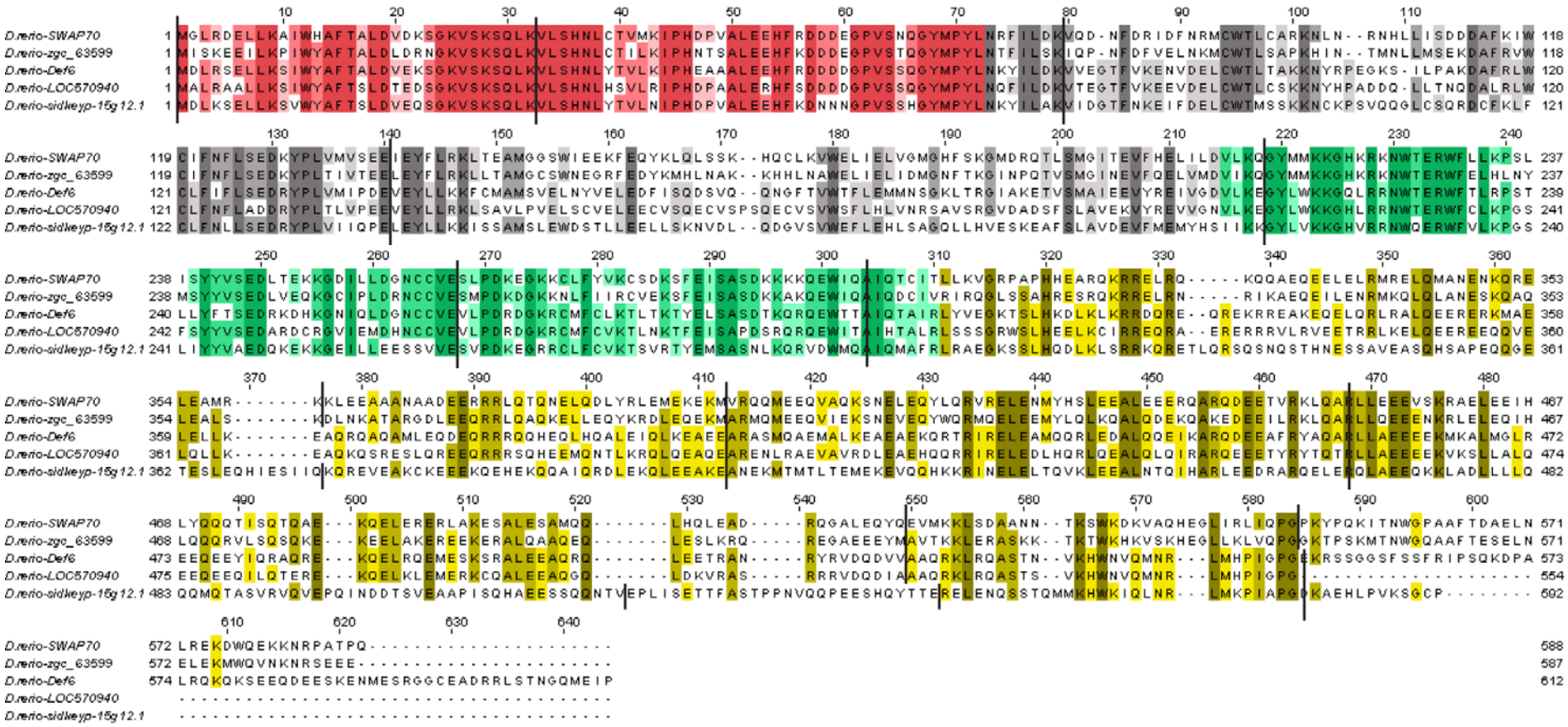


Figure 3-4 Multiple Sequence Alignment of Def6 Paralogues in Zebrafish. Full length amino acid sequences were aligned using ClustalW2 online programme (Larkin *et al*., 2007) with all default settings. The alignment was coloured based on the setting of 'above 50% identity threshold' using Jalview 2.5.1 release (Waterhouse *et al*., 2009). The regions of EF-hand motif, DSH region, PH domain, and DH-like domain were manually highlighted with red, grey, green, and yellow, respectively, according to Mavrakis *et al*., 2004. Vertical lines indicate exon boundaries.

3.7 Def6 paralogues show high relatedness

To visualise and quantify the relatedness of Def6 Paralogues, unrooted tree and pairwise analysis were performed. Both analyses are based on full length amino acid sequences shown in Figure 3.4. Unrooted tree was generated using distance-matrix method (neighbour-joining) with ClustalW2 (Larkin *et al*., 2007) and MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007) while statistical pairwise analysis was generated using EMBOSS open source software package (Rice *et al*., 2000).

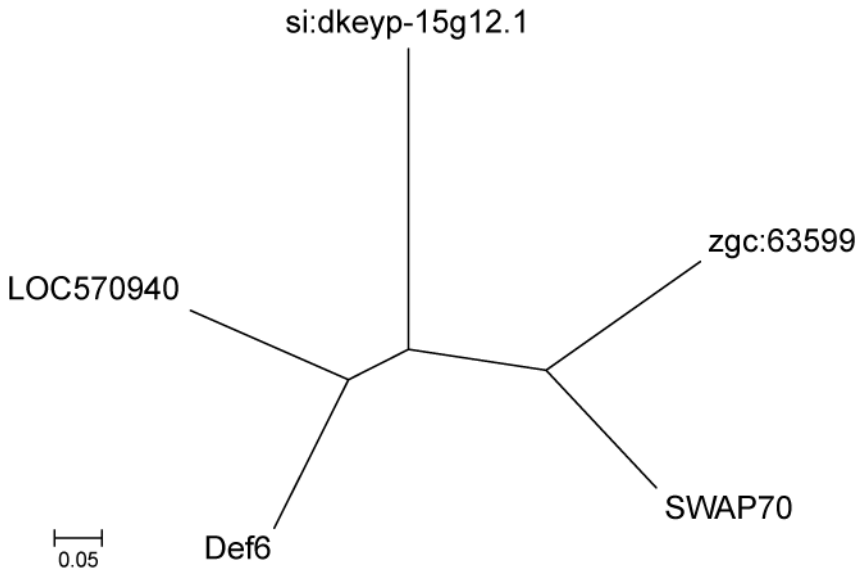
According to the unrooted tree shown in Figure 3.5A, two clusters which are Def6/LOC570940 and SWAP70/zgc:63599 were clearly identified. In other words, SWAP70 and zgc:63599 as well as Def6 and LOC570940 are recently diverged from two ancestor genes. The long branch length of si:dkeyp-15g12.1 indicates that there are more amino acid changes. Also, this unrooted tree put si:dkeyp-15g12.1 as an outgroup to SWAP70/zgc:63599 and Def6/LOC570940. Relatively, there is a shorter branch length between si:dkeyp-15g12.1 and Def6/LOC570940 ancestor gene. This suggested that si:dkeyp-15g12.1 is more closer to Def6 ancestor gene instead of SWAP70 ancestor gene.

Similarly, as shown in Figure 3.5B, zgc:63599 shows high identity (63.1%) and similarity (81%) to SWAP70 whereas LOC570940 shows high identity (58%) and similarity (75.4%) to Def6. Furthermore, the si:dkeyp-15g12.1 protein shows more significant similarity to Def6/LOC570940 (~43% identity and ~63% similarity) than SWAP70/zgc:63599 (34-37% identity and 56-58% similarity).

However, the pairwise analysis shows that SWAP70/zgc:63599 group is close to Def6/LOC570940 group rather than si:dkeyp-15g12.1, although the unrooted tree put si:dkeyp-15g12.1 in the middle between two groups. SWAP70 and zgc:63599 have 38-43% identities and 61-63% similarities to Def6 and LOC570940. But, there are only 34-37% identities and 56-58% similarities among SWAP70, zgc:63599 and si:dkeyp15g12.1.

In summary, based on the unrooted tree and pairwise analysis, SWAP70 is closed to zgc:63599; Def6 is closer to LOC570940; and si:dkeyp-15g12.1 is an outgroup among Def6 paralogues.

(A)



(B)

Identity%/Similarity%/Gap%	SWAP70	zgc:63599	Def6	LOC570940	si:dkeyp-15g12.1
SWAP70	---	63.1/81.0/0.2	42.7/62.5/9.5	41.4/62.5/10.3	37.3/57.8/16.8
zgc:63599		---	40.9/61.7/9.4	38.3/61.1/10.1	34.4/56.4/14.9
Def6			---	58.0/75.4/10.1	42.4/62.2/13.6
LOC570940				---	44.0/64.4/10.3
si:dkeyp-15g12.1					---

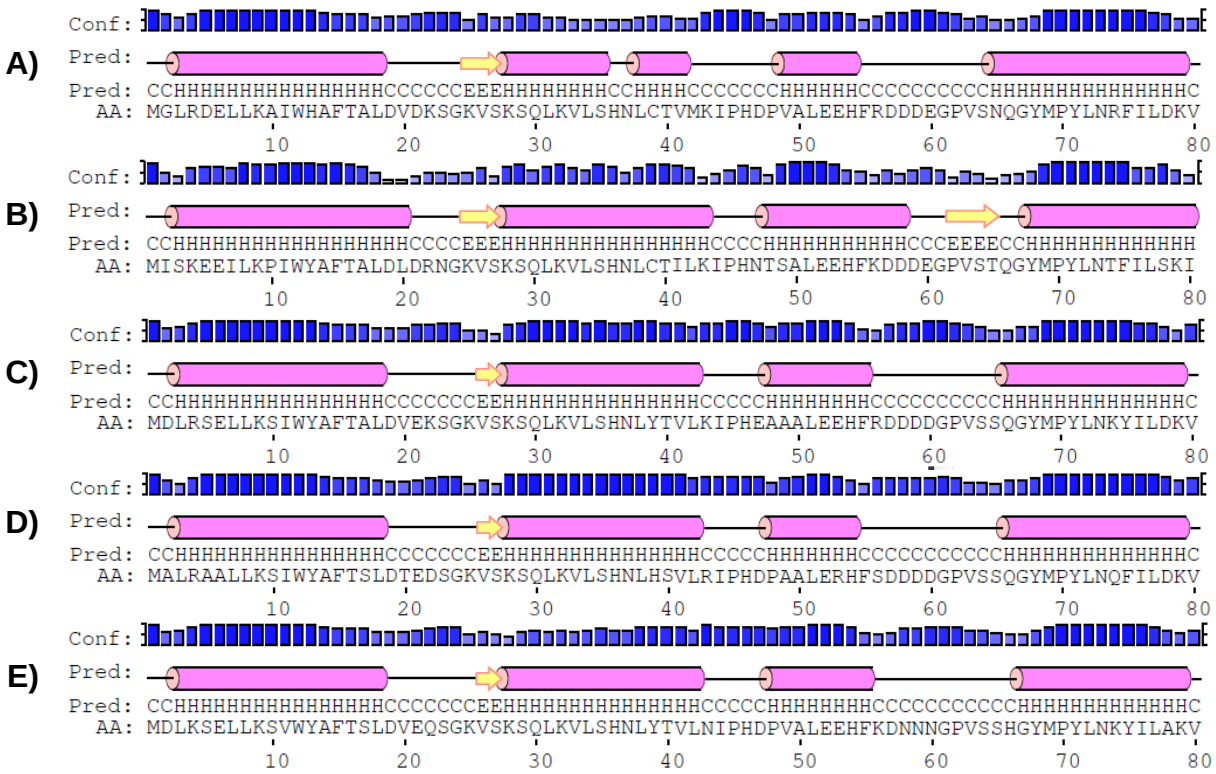
Figure 3-5 Relatedness of Def6 Paralogues in Zebrafish. (A) Unrooted Tree. The amino acid alignment shown in figure 3.3 was used to generate the unrooted tree using ClustalW2 with neighbour-joining distance-matrix method (Larkin *et al*., 2007) and MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007). (B) Pairwise Analysis. EMBOSS open source software analysis package (Rice *et al*., 2000) was used. The settings were 10.0 for gap open, 0.5 for gap extend, and blosum50 for matrix. The numbers indicate the percentages of identity, similarity and gap between amino acid sequences in order.

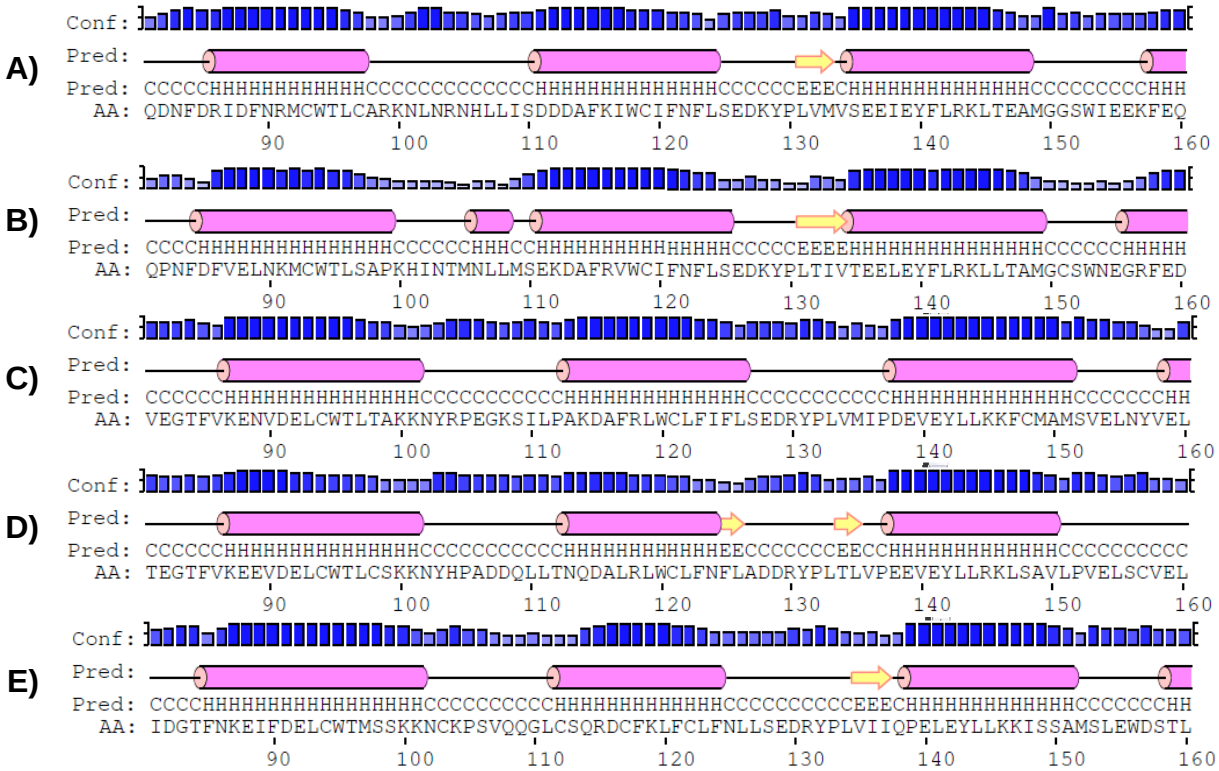
3.8 Def6 Paralogues are extremely conserved in secondary structures

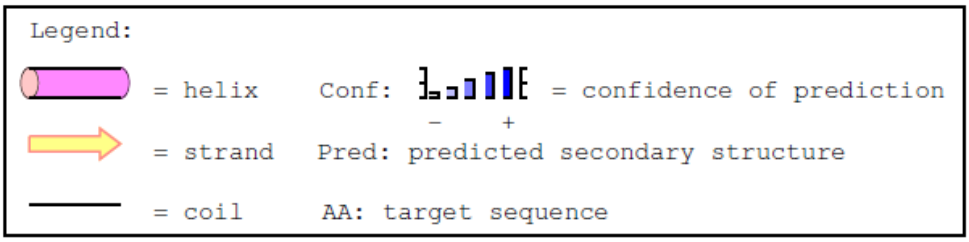
To understand the structural conservation, secondary structure was predicted based on the full length amino acids shown in Figure 3.4. The predictions were accomplished using PSIPRED method (Jones, 1999) at University College London server (Bryson *et al*., 2005). The output results from PSIPRED server were manually modified and are shown in Figure 3.6.

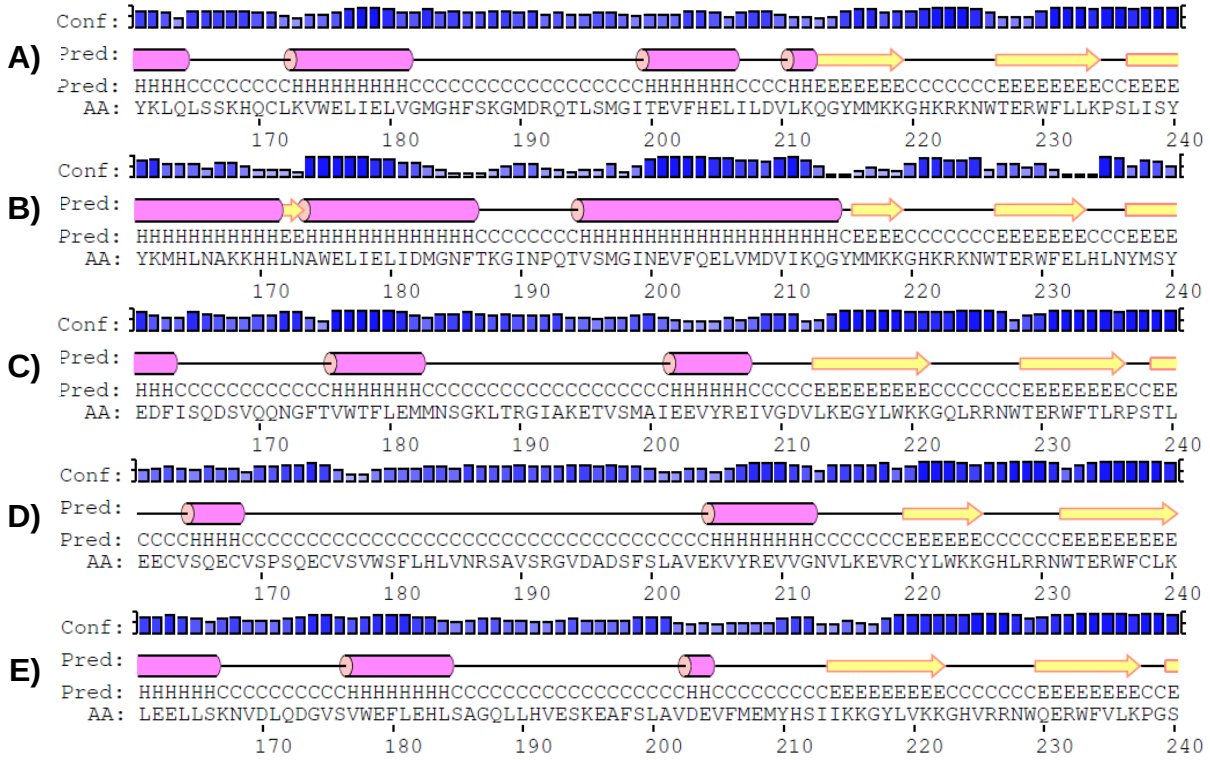
The five Def6 paralogues show a very similar structure pattern. At the N-terminal end, there are nine to ten α-helices. All α-helices are highly supported by PSIPRED method whereas the confidence for the regions in between α-helices is lower. There are seven β-sheets in a series followed by one α-helix. This region is predicted as PH domain according to Pfam and SMART. The arrows indicating β-sheets are poorly supported while the regions in between are significantly supported. After the PH domain, there are long α-helices predicted with high confidence in each sequences, except si:dkeyp-15g12.1. The prediction of si:dkeyp-15g12.1shows that there are some linear coils introduced to the regions after PH domain.

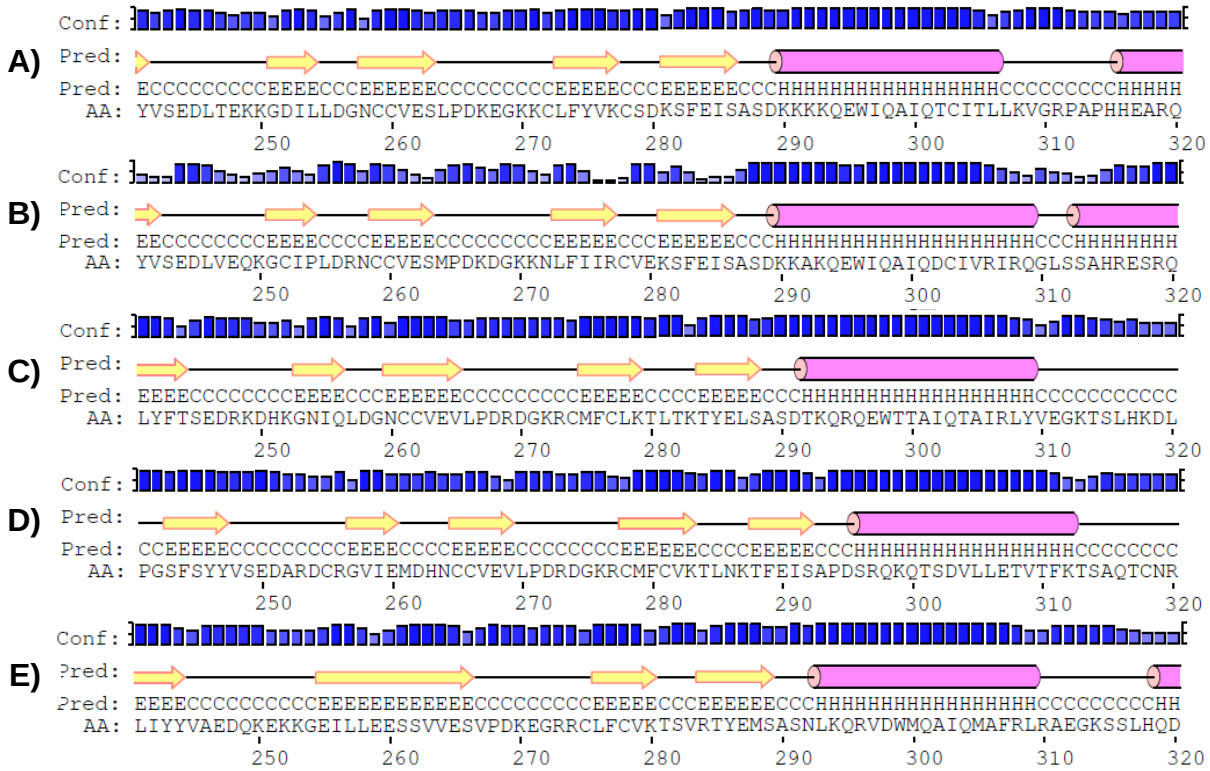
To conclude, all Def6 paralogues show the high structural conservation in secondary structure predictions. There are extensive α helixes before and after the PH domain region, but the C-terminal ends are less conserved.


A)
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHCCCCCCEEEHHHHHHHHCCHHHHCCCCCCCHHHHHHCCCCCCCCCCHHHHHHHHHHHHHHHC
AA: MGLRDELLKAIWHAFTALDVDKSGKVSKSQLKVLSHNLCTVMKIPHDPVALEEHFRDDDEGPVSNQGYMPYLNRFILDKV
10 20 30 40 50 60 70 80
B)
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHHHCCCCEEEHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHCCCEEEECCHHHHHHHHHHHHH
AA: MISKEEILKPIWYAFTALDLDRNGKVSKSQLKVLSHNLCTILKIPHNTSALEEHFKDDDEGPVSTQGYMPYLNTFILSKI
10 20 30 40 50 60 70 80
C)
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHCCCCCCCEEHHHHHHHHHHHHHHHCCCCCHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHC
AA: MDLRSELLKSIWYAFTALDVEKSGKVSKSQLKVLSHNLYTVLKIPHEAAALEEHFRDDDDGPVSSQGYMPYLNKYILDKV
10 20 30 40 50 60 70 80
D)
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHCCCCCCCEEHHHHHHHHHHHHHHHCCCCCHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHC
AA: MALRAALLKSIWYAFTSLDTEDSGKVSKSQLKVLSHNLHSVLRIPHDPAALERHFSDDDDGPVSSQGYMPYLNQFILDKV
10 20 30 40 50 60 70 80
E)
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHCCCCCCCEEHHHHHHHHHHHHHHHCCCCCHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHC
AA: MDLKSELLKSVWYAFTSLDVEQSGKVSKSQLKVLSHNLYTVLNIPHDPVALEEHFKDNNNGPVSSHGYMPYLNKYILAKV
10 20 30 40 50 60 70 80


A)
Conf:
Pred:
Pred: CCCCCHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCEEECHHHHHHHHHHHHHHCCCCCCCCCHHH
AA: QDNFDRIDFNRMCWTLCARKNLNRNHLLISDDDAFKIWCIFNFLSEDKYPLVMVSEEIEYFLRKLTEAMGGSWIEEKFEQ
90 100 110 120 130 140 150 160
B)
Conf:
Pred:
Pred: CCCCHHHHHHHHHHHHHHCCCCCCHHHCCHHHHHHHHHHHHHHHCCCCCEEEEHHHHHHHHHHHHHHHCCCCCCHHHHH
AA: QPNFDFVELNKMCWTLSAPKHINTMNLLMSEKDAFRVWCIFNFLSEDKYPLTIVTEELEYFLRKLLTAMGCSWNEGRFED
90 100 110 120 130 140 150 160
C)
Conf:
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHCCCCCCCHH
AA: VEGTFVKENVDELCWTLTAKKNYRPEGKSILPAKDAFRLWCLFIFLSEDRYPLVMIPDEVEYLLKKFCMAMSVELNYVEL
90 100 110 120 130 140 150 160
D)
Conf:
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHEECCCCCCCEECCHHHHHHHHHHHHCCCCCCCCCC
AA: TEGTFVKEEVDELCWTLCSKKNYHPADDQLLTNQDALRLWCLFNFLADDRYPLTLVPEEVEYLLRKLSAVLPVELSCVEL
90 100 110 120 130 140 150 160
E)
Conf:
Pred:
Pred: CCCCHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHCCCCCCCCCCEEECHHHHHHHHHHHHCCCCCCCHH
AA: IDGTFNKEIFDELCWTMSSKKNCKPSVQQGLCSQRDCFKLFCLFNLLSEDRYPLVIIQPELEYLLKKISSAMSLEWDSTL
90 100 110 120 130 140 150 160


Legend:
= helix
Conf:
= confidence of prediction
–
+
= strand
Pred: predicted secondary structure
= coil
AA: target sequence





Legend:

= helix

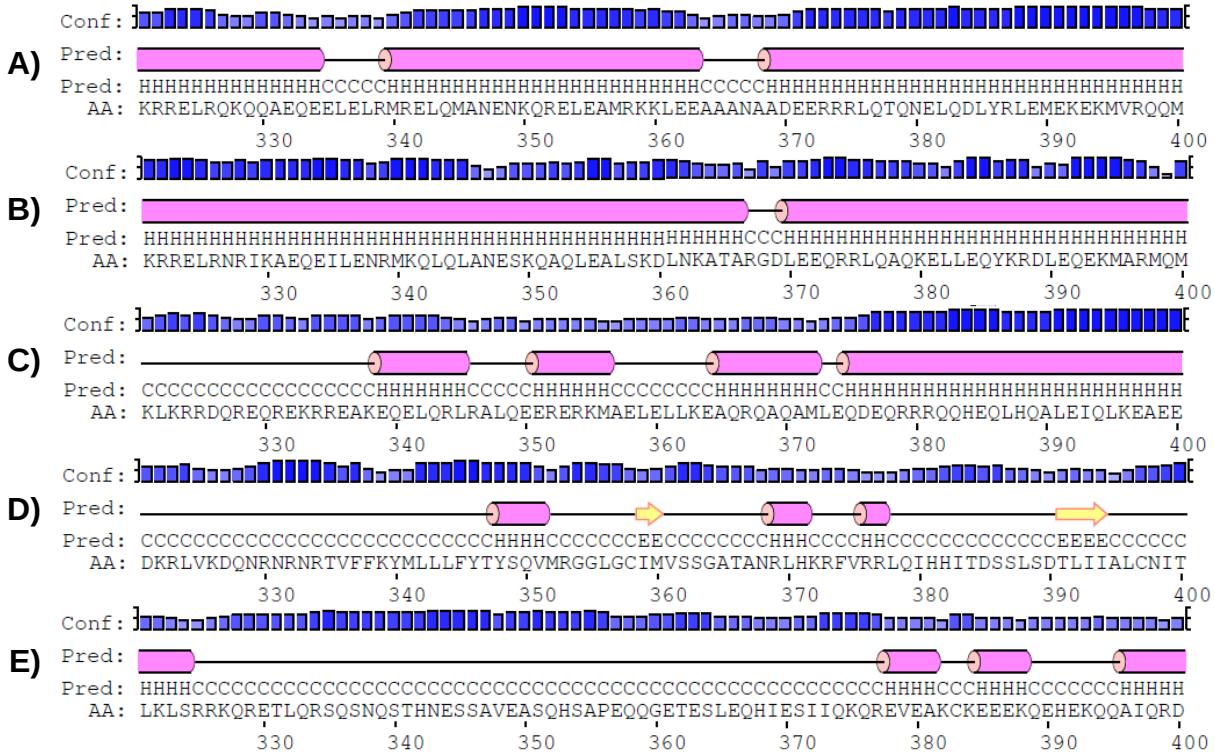
= strand

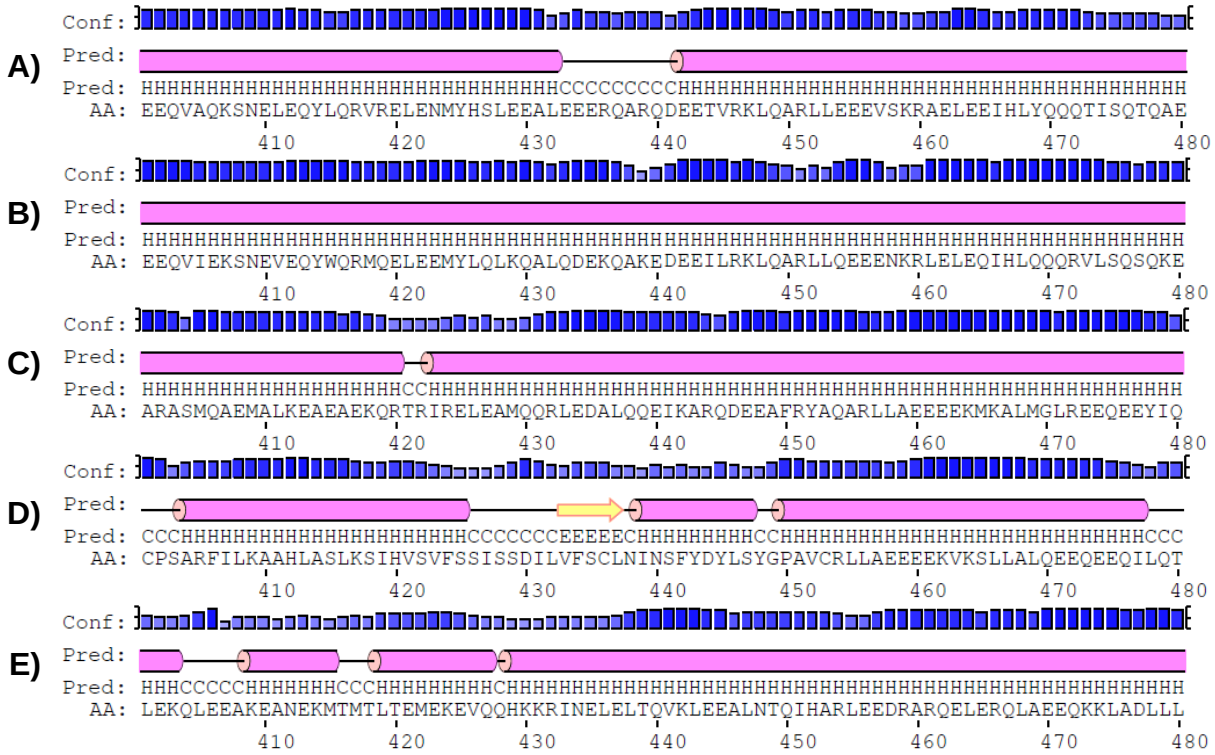
= coil

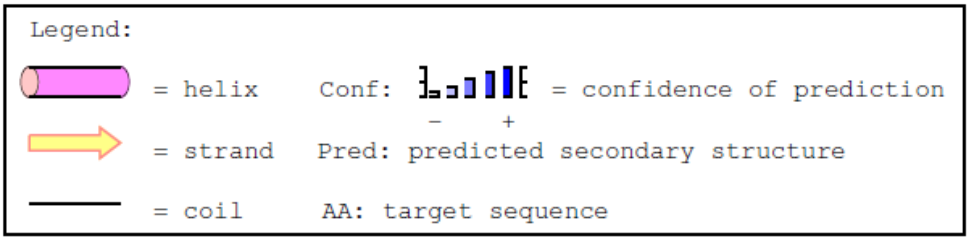
Conf: = confidence of prediction

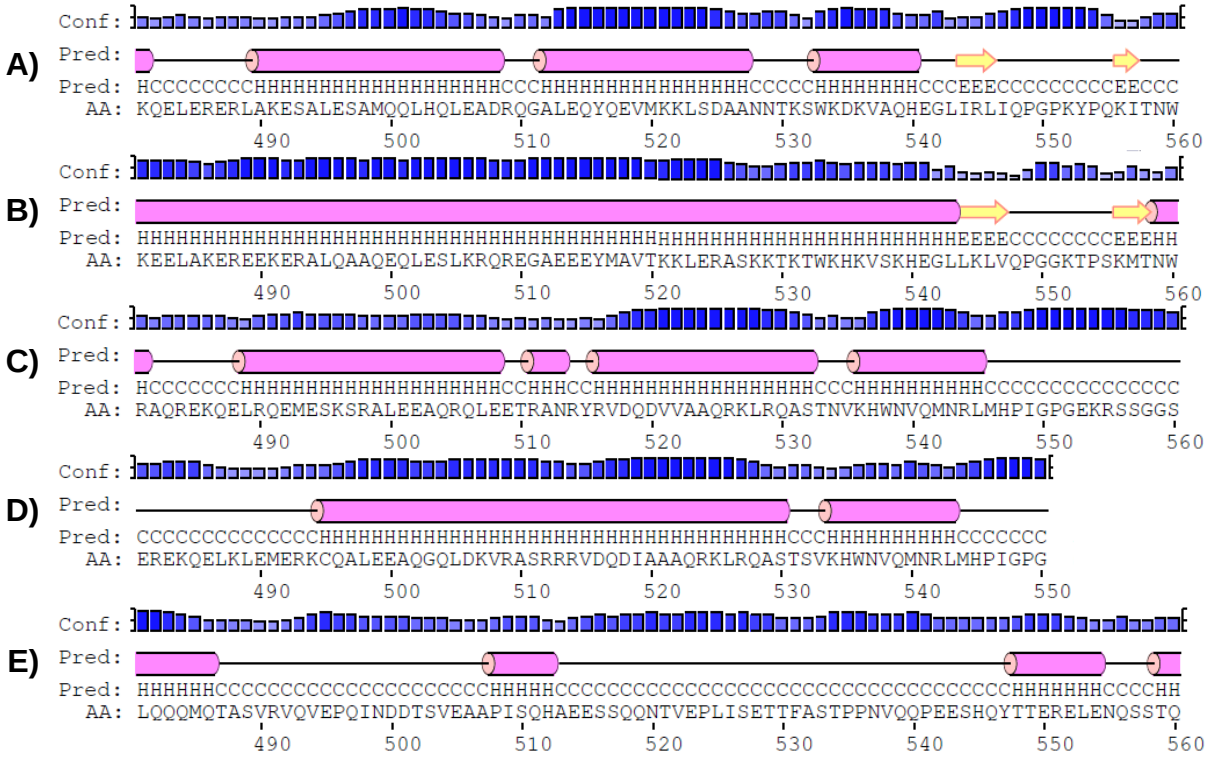
Pred: predicted secondary structure

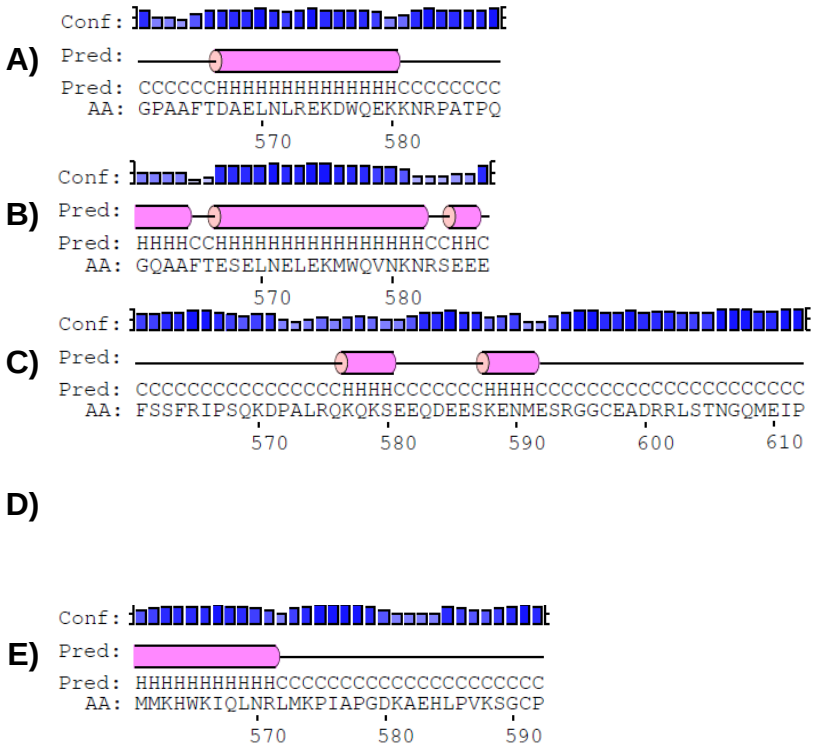
AA: target sequence


A)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: KRRELRQKQQAEQEELELRMRELQMANENKQRELEAMRKKLEEAAANAADEERRRLQTQNELQDLYRLEMEKEKMVRQQM
330 340 350 360 370 380 390 400
B)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: KRRELRNRIKAEQEILENRMKQLQLANESKQAQLEALSKDLNKATARGDLEEQRRLQAQKELLEQYKRDLEQEKMARMQM
330 340 350 360 370 380 390 400
C)
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCHHHHHHCCCCCHHHHHHCCCCCCCCHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHH
AA: KLKRRDQREQREKRREAKEQELQRLRALQEERERKMAELELLKEAQRQAQAMLEQDEQRRRQQHEQLHQALEIQLKEAEE
330 340 350 360 370 380 390 400
D)
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCEECCCCCCCCHHHCCCCHHCCCCCCCCCCCCCEEEECCCCCC
AA: DKRLVKDQNRNRNRTVFFKYMLLLFYTYSQVMRGGLGCIMVSSGATANRLHKRFVRRLQIHHITDSSLSDTLIIALCNIT
330 340 350 360 370 380 390 400
E)
Conf:
Pred:
Pred: HHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCHHHHCCCCCCHHHHH
AA: LKLSRRKQRETLQRSQSNQSTHNESSAVEASQHSAPEQQGETESLEQHIESIIQKQREVEAKCKEEEKQEHEKQQAIQRD
330 340 350 360 370 380 390 400


A)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: EEQVAQKSNELEQYLQRVRELENMYHSLEEALEEERQARQDEETVRKLQARLLEEEVSKRAELEEIHLYQQQTISQTQAE
410 420 430 440 450 460 470 480
B)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: EEQVIEKSNEVEQYWQRMQELEEMYLQLKQALQDEKQAKEDEEILRKLQARLLQEEENKRLELEQIHLQQQRVLSQSQKE
410 420 430 440 450 460 470 480
C)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: ARASMQAEMALKEAEAEKQRTRIRELEAMQQRLEDALQQEIKARQDEEAFRYAQARLLAEEEEKMKALMGLREEQEEYIQ
410 420 430 440 450 460 470 480
D)
Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHHHHHHHHCCCCCCCEEEEECHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
AA: CPSARFILKAAHLASLKSIHVSVFSSISSDILVFSCLNINSFYDYLSYGPAVCRLLAEEEEKVKSLLALQEEQEEQILQT
410 420 430 440 450 460 470 480
E)
Conf:
Pred:
Pred: HHHCCCCCHHHHHHHCCCHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: LEKQLEEAKEANEKMTMTLTEMEKEVQQHKKRINELELTQVKLEEALNTQIHARLEEDRARQELERQLAEEQKKLADLLL
410 420 430 440 450 460 470 480


Legend:
= helix
Conf:
= confidence of prediction
–
+
= strand
Pred: predicted secondary structure
= coil
AA: target sequence


A)
Conf:
Pred:
Pred: HCCCCCCCCHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHCCCCCHHHHHHHHCCCEEECCCCCCCCCEECCC
AA: KQELERERLAKESALESAMQQLHQLEADRQGALEQYQEVMKKLSDAANNTKSWKDKVAQHEGLIRLIQPGPKYPQKITNW
490 500 510 520 530 540 550 560
B)
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHEEECCCCCCCCEEEHH
AA: KEELAKEREEKERALQAAQEQLESLKRQREGAEEEYMAVTKKLERASKKTKTWKHKVSKHEGLLKLVQPGGKTPSKMTNW
490 500 510 520 530 540 550 560
C)
Conf:
Pred:
Pred: HCCCCCCCHHHHHHHHHHHHHHHHHHHCCHHCCHHHHHHHHHHHHHHHHCCCHHHHHHHHHCCCCCCCCCCCCCCCC
AA: RAQREKQELRQEMESKSRALEEAQRQLEETRANRYRVDQDVVAAQRKLRQASTNVKHWNVQMNRLMHPIGPGEKRSSGGS
490 500 510 520 530 540 550 560
D)
Conf:
Pred:
Pred: CCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHCCCCCCC
AA: EREKQELKLEMERKCQALEEAQGQLDKVRASRRRVDQDIAAAQRKLRQASTSVKHWNVQMNRLMHPIGPG
490 500 510 520 530 540 550
E)
Conf:
Pred:
Pred: HHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCHH
AA: LQQQMQTASVRVQVEPQINDDTSVEAAPISQHAEESSQQNTVEPLISETTFASTPPNVQQPEESHQYTTERELENQSSTQ
490 500 510 520 530 540 550 560


A)
Conf:
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHCCCCCCCC
AA: GPAAFTDAELNLREKDWQEKKNRPATPQ
570 580
B)
Conf:
Pred:
Pred: HHHHCCHHHHHHHHHHHHHHHCCHHC
AA: GQAAFTESELNELEKMWQVNKNRSEEE
570 580
C)
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCHHHHCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCC
AA: FSSFRIPSQKDPALRQKQKSEEQDEESKENMESRGGCEADRRLSTNGQMEIP
570 580 590 600 610
D)
E)
Conf:
Pred:
Pred: HHHHHHHHHHCCCCCCCCCCCCCCCCCCCCC
AA: MMKHWKIQLNRLMKPIAPGDKAEHLPVKSGCP
570 580 590

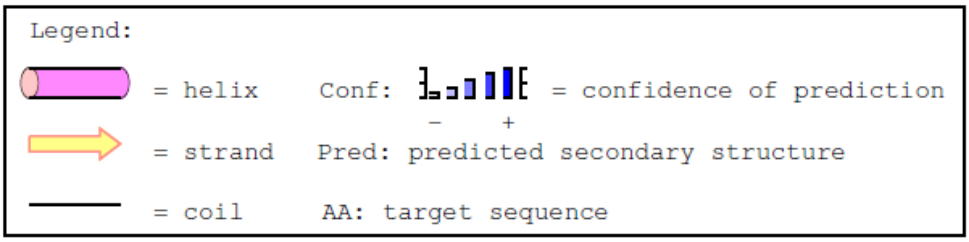

Legend:
= helix
Conf:
= confidence of prediction
–
+
= strand
Pred: predicted secondary structure
= coil
AA: target sequence

Figure 3-6 Secondary Structure Predictions of Def6 Paralogues in Zebrafish. (A) SWAP70. (B) zgc:63599. (C) Def6. (D) LOC570940. (E) si:dkeyp-15g12.1. Prediction was performed using PSIPRED method (Jones, 1999) with mask low complexity regions as filtering option at University College London server, version 3.0 (Bryson *et al*., 2005). The predictions are lined up manually. Conf stands for the confidence of the prediction, Pred stands for prediction result, and AA stands for amino acid sequence. For the Pred sections, letter H and pink cylinder indicate the regions which form alpha helices, letter E and yellow arrow indicate the regions which form beta sheets, and letter C and black straight line indicate the regions which form linear coil.

3.9 SWAP70, zgc:63599, Def6, and LOC570940 are folded into highly similar structures

As shown in Figure 3.6, secondary structures of *Def6* paralogues determined by PSIPRED prediction method were highly conserved. Next, to understand whether the protein folds are conserved among Def6 paralogues, the I-TASSER programme (Roy *et al*., 2010, Zhang, 2009 and Zhang, 2008) at the University of Michigan was employed to predict the 3D structure for each Def6 paralogue. The best models suggested for each paralogue by I-TASSER programme were obtained and viewed using First Glance in Jmol, version 1.45 (http://firstglance.jmol.org) with rainbow coloured and contacts options. The models are rainbow coloured from N-terminus to C-terminus in blue to red (Figure 3.11A). The contacts between amino acids in each model are shown in Figure 3.11B and the contact details are shown in Table 3.2B. The quality of each predicted model is shown in Figure 3.11A. According to Roy *et al*. (2010), the quality of the predicted model is indicated with three values, confidence score (C-score), template modelling score (TM-score), and root mean square deviation (RMSD). The C-score is ranged from -5.0 to +2.0 and the TM-score is ranged from 0 to 1. If C-score is larger than -1.50 toward positive and TM-score is larger than 0.5 toward 1, the predicted model has the correct topology. Besides, the estimated RMSD which compares the resolution of the predicted model with native structure is in the range of 0–30Å. Lower estimated RMSD means better resolution of the predicted model.

As shown in Figure 3.11, si:dkeyp-15g12.1 protein folds into a hammer-like shape structure. In contrast, other Def6 paralogue proteins show a very similar donut shape structure from the top view. Comparatively, Def6 and LOC570940 show the highest C-score and TM-score with better resolution in term of estimated RMSD (Table 3.2). The quality of SWAP70 predicted model is also above the cutoff values of -1.50 for C-score and 0.5 for TM-score. Although zgc:63599 and si:dkeyp-15g12.1 have lower C-score and TM-score with poor resolution, both scores are still higher than the cutoff value to support the predicted foldings.

Table 3.2B shows that the predicted model of SWAP70 has two contact points where the amino acids physically contact to each other in the molecular contact model view. The first contact is between hydrophobic Valine at position 80 and positively charged Arginine at position 486 with 5.05Å in distance. The second one is formed by positively charged Lysine and polar Asparagine *with* 8.09Å in distance.

According to the predicted zgc:65399 folding, there are three amino acid contacts. Hydrophobic Isoleucine at position 80 and negatively charged Glutamate at position 489 form the first contact point with 6.38Å in distance. Both positively charged Lysine at positions 168 and 169 make contact with positively charged Lysine at position 581 and hydrophobic Valine at position 579, respectively, to form the second contact with 11.9Å in distance and the third contact with 8.12Å in distance.

Def6 model shows one contact point and three regions that the distances are close between amino acids. Hydrophobic valine at position 81 and hydrophobic Phenylalanine at position 561 contribute to the only contact point with 7.45Å in distance. Hydrophobic phenylalanine at position 163 is close to both polar amino acids, serine and asparagine at position 604 with 9.18Å in distance and position 606 with 11.81Å in distance, respectively. Also, hydrophobic isoleucine at position 164 and positively charged arginine at position 601 are closed to each other in 11.78Å distance.

The molecular contact models mentioned above share the similar contact regions. The positions for the first amino acids in each contact point are close to others. For example, the first amino acids involved in the contact are around positions 80 and 81 of each model. However, the model of LOC570940 shows varied contact positions. The first contact point with 9.54Å in distance is formed by the side chain interaction between negatively charged aspartate at position 47 and positively charged lysine at position 492. The negatively charged arginine at position 130 makes contact to polar glutamine at position 489 in 5.49Å. The hydrophobic valine at position 194 and positively charged arginine at position 275 contribute to the last contact point in LOC570940 model.

Taken together, SWAP70, zgc:63599, Def6, and LOC570940 share the similar secondary structure. They seem to have a conversed protein folding despite the fact that the amino acid sequences of C-terminal ends are less conserved.

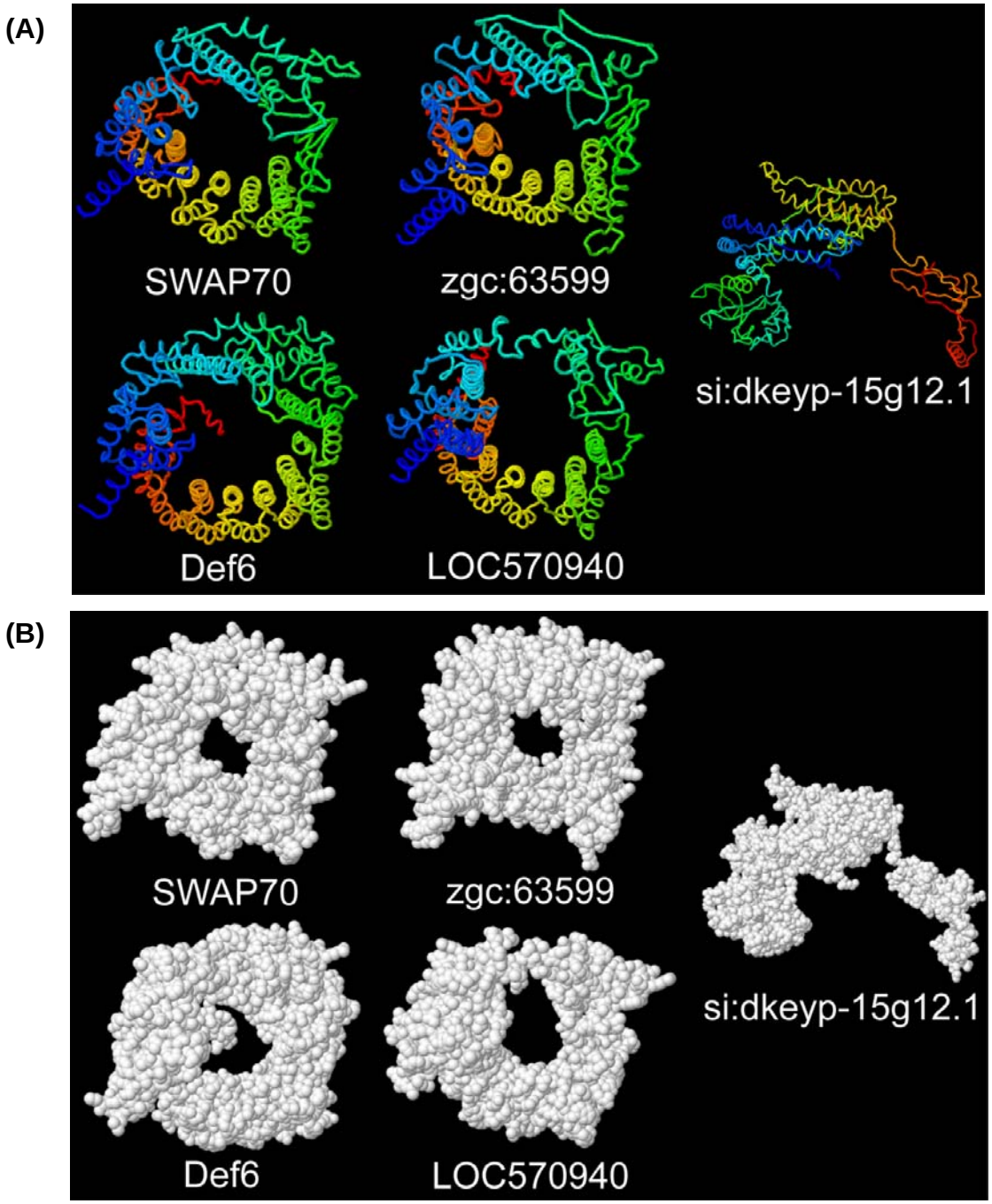


Figure 3-7 Three Dimensional Structure Predictions of Def6 Paralogues in Zebrafish. Full length amino acid sequences were submitted to I-TASSER (Roy *et al*., 2010, Zhang, 2009 and Zhang, 2008) to generate predicted 3D structure models. The model was visualised using First Glance in Jmol, version 1.45. The accuracies of the models are shown in Table 3.2A. (A) Rainbow Coloured Model. N-terminus is in blue and C-terminus is in red. (B) Molecular Contact Model. The orientations are the same as models shown in A. The contacted amino acids are shown in Table 3.2B.

(A)

Gene Name	C-Score	Estimated TM-Score	Estimated RMSD (Å)
SWAP70	-0.95	0.59±0.14	9.9±4.6
zgc:63599	-1.23	0.56±0.15	10.6±4.6
Def6	-0.28	0.68±0.12	8.4±4.5
LOC570940	-0.45	0.68±0.13	8.4±4.6
si:deyp-15g12.1	-1.31	0.55±0.15	10.6±4.6

(B)

Gene Name	Contact Point 1			Contact Point 2			Contact Point 3		
	1st Amino Acid	2nd Amino Acid	Estimated Distance between two amino acids (Å)	1st Amino Acid	2nd Amino Acid	Estimated Distance between two amino acids (Å)	1st Amino Acid	2nd Amino Acid	Estimated Distance between two amino acids (Å)
SWAP70	80 VAL	486 ARG	5.05	162 LYS	582 ASN	8.09	---	---	---
zgc:63599	80 ILE	489 GLU	6.38	168 LYS	581 LYS	11.9	169 LYS	579 VAL	8.12
Def6	81 VAL	561 PHE	7.45	*163 PHE	*604 SER/ *606 ASN	*9.18/ *11.81	*164 ILE	*601 ARG	*11.78
LOC570940	47 ASP	492 LYS	9.54	130 ARG	489 GLN	5.49	194 VAL	275 ARG	8.22

* Contact points which are not physically contacted but the distances are very very closed.

Table 3.2 Three Dimensional Structure Predictions of Def6 Paralogues in Zebrafish. (A) The quality of each model. C-score, estimated TM-score and estimated RMSD indicate the quality level. (B) Amino acids that make contact in the different models are listed. The estimated distances were measured between two amino acid carbon centres using First Glance in Jmol, version 1.45.

4. Identification and Analysis of *def6*-related Genes in Vertebrates

4.1 Data acquisition

In this chapter the bioinformatics analyses of Def6-related genes in vertebrate species are presented. In addition to zebrafish (*Danio rerio*), four teleost fish species including tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*) and medaka (*Oryzias latipes*) were selected. Human (*Homo sapiens*) and mouse (*Mus musculus*) were chosen to be representatives of primates and rodents, while frog (*Xenopus tropicalis*), chicken (Gallus gallus) and anole lizard (*Anolis carolinensis*) were selected to represent amphibians, birds and reptiles, respectively. The priority of the information source is NCBI Genbank but because the teleost genomes are mainly available in Ensembl database, all the gene details for teleosts were acquired from Ensembl (Table 4.1 and 4.2). As shown in Table 4.1, with the exception of zebrafish, only four def6-related genes are found in each teleost species. Those genes are orthologues of *Def6*, *SWAP70*, *zgc:63599*, and *si:dkeyp-15g12.1* according the Ensembl database. However, *LOC570940* orthologues cannot be identified in other teleost species even thought the blastp and tblastn searches were carried out in NCBI, Uniprot, and Ensembl databases. According to the DEF6 protein family shown in Ensembl (ENSFM00250000001889), there are only two genes of *Def6* and *SWAP70* orthologues identified in all available vertebrate genomes. Moreover, both blastp and tblastn searches with *zgc:63599*, *si:dkeyp-15g12.1*, *LOC570940* amino acid sequences also point to *Def6* and *SWAP70* in NCBI, UniProt, and Ensembl databases. The details of orthologues of *Def6* and *SWAP70* in selected vertebrate species were obtained mainly from NCBI Genbank and are shown in Table 4.2.

Common Name	Species	Gene Name[1]	Gene Name[2]	Ensembl Gene ID	No. of Exons	Genomic DNA Length (kb)	Transcript Length (bp)	Peptide Length (aa)	Genome Location	Strand
Tetraodon	*Tetraodon nigroviridis*	*DEF6*	*def6*	ENSTNIG00000015111	11	4.99	1851	616	Chromosome 9:3,242,470-3,247,460	Reverse
		ENSTNIG00000011420	*def6-like*	ENSTNIG00000011420	11	3.51	1644	548	Chromosome 19:2,050,309-2,053,821	Forward
		SWAP70 (1 of 2)	*swap70a*	ENSTNIG00000013737	12	8.19	2528	598	Chromosome 13:12,305,792-12,313,985	Reverse
		SWAP70 (2 of 2)	*swap70b*	ENSTNIG00000009368	11	4.18	1575	525	Chromosome 5:6,426,601-6,430,777	Reverse
Fugu	*Takifugu rubripes*	*DEF6*	*def6*	ENSTRUG00000016125	11	5.41	1833	610	Scaffold_66:1,140,126-1,145,533	Forward
		ENSTRUG00000012423	*def6-like*	ENSTRUG00000012423	12	4.70	1863	559	Scaffold_8:199,053-203,753	Forward
		SWAP70 (1 of 2)	*swap70a*	ENSTRUG00000009866	12	8.70	1797	598	Scaffold_191:206,187-214,881	Forward
		SWAP70 (2 of 2)	*swap70b*	ENSTRUG00000017420	13	5.29	1797	598	Scaffold_14:317,082-322,367	Forward
Stickleback	*Gasterosteus aculeatus*	*DEF6*	*def6*	ENSGACG00000008311	11	8.48	2455	619	Group XII:10,464,460-10,472,941	Reverse
		ENSGACG00000019860	*def6-like*	ENSGACG00000019860	13	4.73	2173	555	Group IV:30,935,886-30,940,616	Reverse
		SWAP70 (1 of 2)	*swap70a*	ENSGACG00000009748	14	12.92	2343	590	Group XIX:11,097,999-11,110,921	Reverse
		SWAP70 (2 of 2)	*swap70b*	ENSGACG00000015680	12	6.86	2152	588	Group II:10,115,093-10,121,947	Forward
Medaka	*Oryzias latipes*	*DEF6*	*def6*	ENSORLG00000006652	11	8.67	1860	619	Chromosome 7:11367,897-11,376,569	Reverse
		ENSORLG00000019580	*def6-like*	ENSORLG00000019580	13	11.17	2288	626	Chromosome 23:17,370,714-17,381,882	Reverse
		SWAP70 (1 of 2)	*swap70a*	ENSORLG00000001090	12	33.21	1829	589	Chromosome 6:2,037,056-2,070,262	Reverse
		SWAP70 (2 of 2)	*swap70b*	ENSORLG00000007129	12	10.22	1761	587	Chromosome 3:18,356,969-18,367,192	Forward
Zebrafish	Danio rerio	*zgc:63721*	*def6a*	ENSDARG00000012247	11	24.55	2197	612	Chromosome 8: 14,348,627-14,373,176	Reverse
		*LOC570940**	*def6b*	HM752767 #	10	13.65	1662	554	Chromosome 22:935,708-1,021,603^	Unknown^
		si:dkey-8l13.4	*swap70a*	ENSDARG00000057286	12	40.01	2558	588	Chromosome 18: 17,073,094-17,113,101	Reverse
		*zgc:63599**	*swap70b*	HM752768 #	12	31.25	1835	587	Chromosome 7: 70,983,022-71,014,271	Reverse
		si:dkeyp-15g12.1	*def6-like*	ENSDARG00000034717	13	16.26	1777	592	Chromosome 4: 18,573,101-18,589,360	Reverse

[1] The gene names which are shown in Ensembl database, version release 59 (August 2010).

[2] The gene names which are proposed according to the phylogenetics analyses in this thesis.

* The details of the genes are shown and discussed in Chapter 3 in this thesis.

NCBI accession numbers of paritial coding sequences are shown instead of Ensembl gene IDs which are not yet assigned.

^ The chromosome location is the location of contig FP067424.7 which contains *LOC570940* in Zebrafish Zv9 (July, 2010) in Pre-Ensembl.

Table 4.1 The Summary of Def6-related Genes in Teleost Species.

Common Name	Species	Gene Name[1]	Gene Name[2]	NCBI Gene ID	Ensembl Gene ID	UniProt ID	No. of Exons	Genomic DNA Length (kb)	Transcript Length (bp)	Peptide Length (aa)	Genome Location	Strand
Anole Lizard	*Anolis carolinensis*	*DEF6**	*def6*	---	ENSACAG00000002129	---	11*	66.94*	1743*	581*	scaffold_238: 518,006-584,941*	Forward*
		*SWAP70**	*swap70*	---	ENSACAG00000009213	---	13*	38.84*	2006*	588*	scaffold_249: 1,339,770-1,378,605*	Reverse*
Chicken	*Gallus gallus*	*DEF6*	*DEF6*	419893	ENSGALG00000002615	---	11	11.41	2526	625	Chromosome 26: 3,939,957-3,951,361	Reverse
		SWAP70	*SWAP70*	423044	ENSGALG00000005750	Q5F4B2	12	27.80	3849	586	Chormosome 5:10,238,478-10,266,272	Reverse
Human	*Homo sapiens*	*DEF6*	*DEF6*	50619	ENSG00000023892	Q9H4E7	11	23.96	2320	631	Chromosome 6:35,265,595-35,289,548	Forward
		SWAP70	*SWAP70*	23075	ENSG00000133789	Q9UH65	12	88.88	4848	585	Chromosome 11:9,685,628-9,774,508	Forward
Mouse	*Mus musculus*	*Def6*	*Def6*	23853	ENSMUSG00000002257	Q8C2K1	11	20.83	2294	630	Chromosome 17:28,344,775-28,365,544	Forward
		Swap70	*Swap70*	20947	ENSMUSG00000031015	Q6A028	12	61.75	4056	585	Chromosome 7:117,365,275-117,427,020	Forward
Frog	*Xenopus tropicalis*	*DEF6**	*def6*	---	ENSXETG00000015117	---	11*	60.11*	1791*	596*	scaffold_771: 208,816-268,929*	Forward*
		swap70	*swap70*	100135363	ENSXETG00000024196	A9UMI4	10*	31.81	2238	586	scaffold_235: 211,507-240,630	Forward*

[1] The gene names which are shown in NCBI or Ensembl.

[2] The gene names which are proposed in this thesis.

* The information is from Ensembl database, version release 59 (August 2010).

Table 4.2 The Summary of Def6-related Genes in Selected Vertebrate Species.

4.2 Exon-intron structures are extremely conserved among all Def6-related orthologues

Exon-intron structures of *SWAP70* orthologues, *zgc:63599* orthologues, *Def6* orthologues, and *si:dkeyp-15g12.1* orthologues are shown in Figure 4.1-4.4, respectively. As *LOC570940* orthologues have not yet found in other teleost species, the exon-intron structure of *LOC570940* was compared to *Def6* orthologues. The primary source of exon-intron structures is Ensembl release 59 (August 2010) while that of *zgc:63599* and *LOC570940* is mentioned in chapter 3.

With few exceptions, SWAP70 orthologues (Figure 4.1) and zgc:63599 orthologues (Figure 4.2) in different species show extremely similar exon-intron structures. There are mainly 12 exons. Excluding the species with uncompleted exon annotations, there are 8 exons having identical nucleotide numbers. In addition, if the species with exon splitting events are ignored as well, all the *SWAP70* and *zgc:63599* orthologues have 10 identical exons containing the same nucleotide numbers.

Similarly, Def6 orthologues and LOC570940 shows extremely conserved exon-intron structures (Figure 4.3). Exon 1 excluding 5'UTR as well as exons 2, 5, 6, 7, 8, and 9 have the identical lengths in all *Def6* orthologues and *LOC570940* (96, 141, 147, 109, 299, 167, and 199 nucleotides, respectively). Only exons 3 and 4 are varied among species.

Nevertheless, the *si:dkeyp-15g12.1* orthologues (Figure 4.4) shows relatively variable exon-intron structures. With the exception of the 5'-UTR, exons 1, 2, 5, 6 and 9/10 contain identical numbers of nucleotides. Exons 3, 4, and 11/12 exhibit variable lengths. The exon 7 in *D.rerio* and *T.rubripes* are split in two exons in other teleost species whereas the exons 10 and 11 in *D.rerio* are joined into one exon in other teleost species.

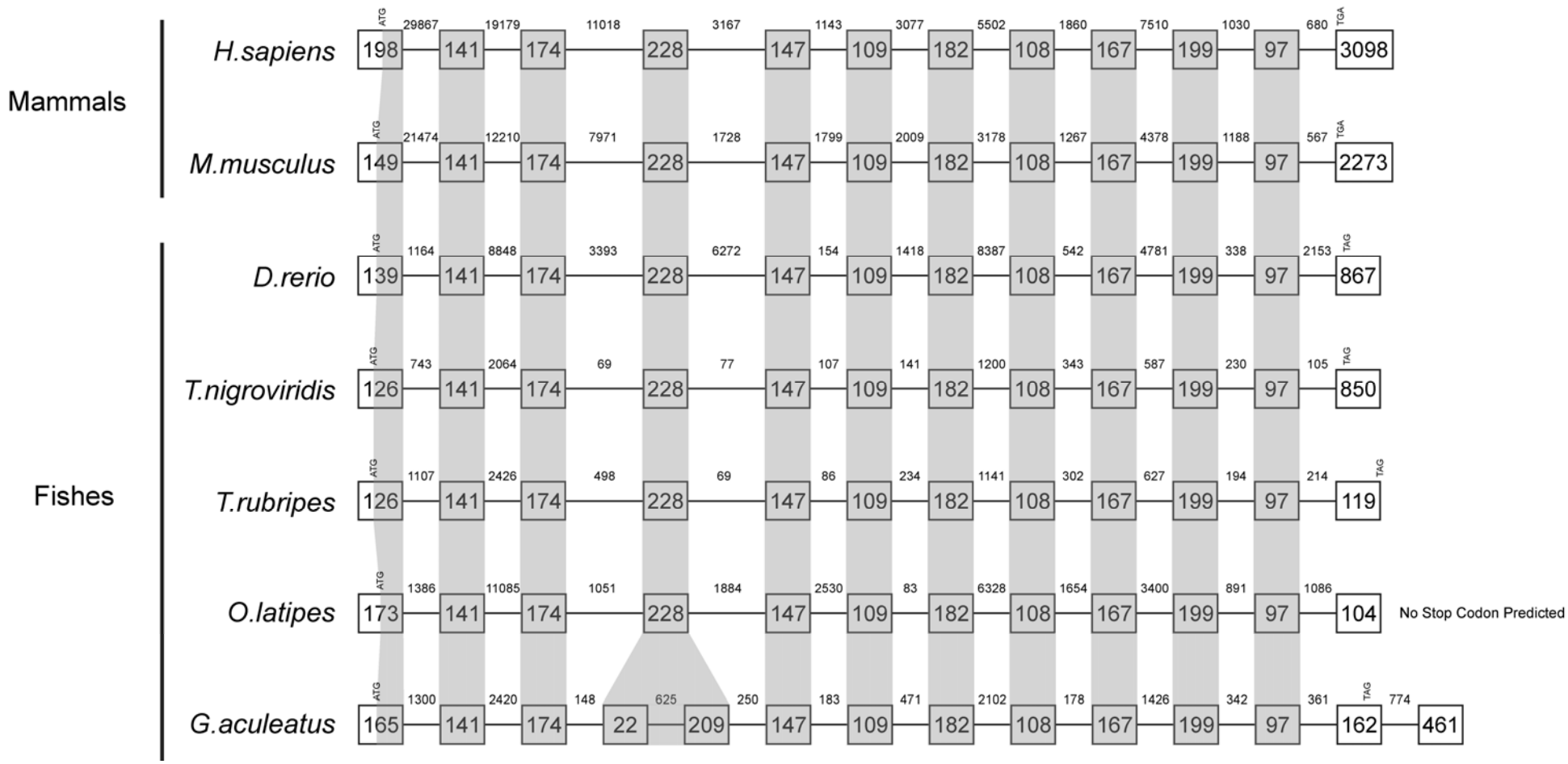


Figure 4-1 Exon-intron Structures of *SWAP70* Orthologues. The exon-intron information is based on Ensembl release 59 (August 2010). The Ensembl gene IDs for each gene are shown in Table 4.1 and 4.2. The locations of the start and stop codons are shown. Small black blocks represent exons and the black lines represent introns. Both black blocks and lines are not shown to scale. Exons with aligned nucleotide lengths are highlighted in grey.

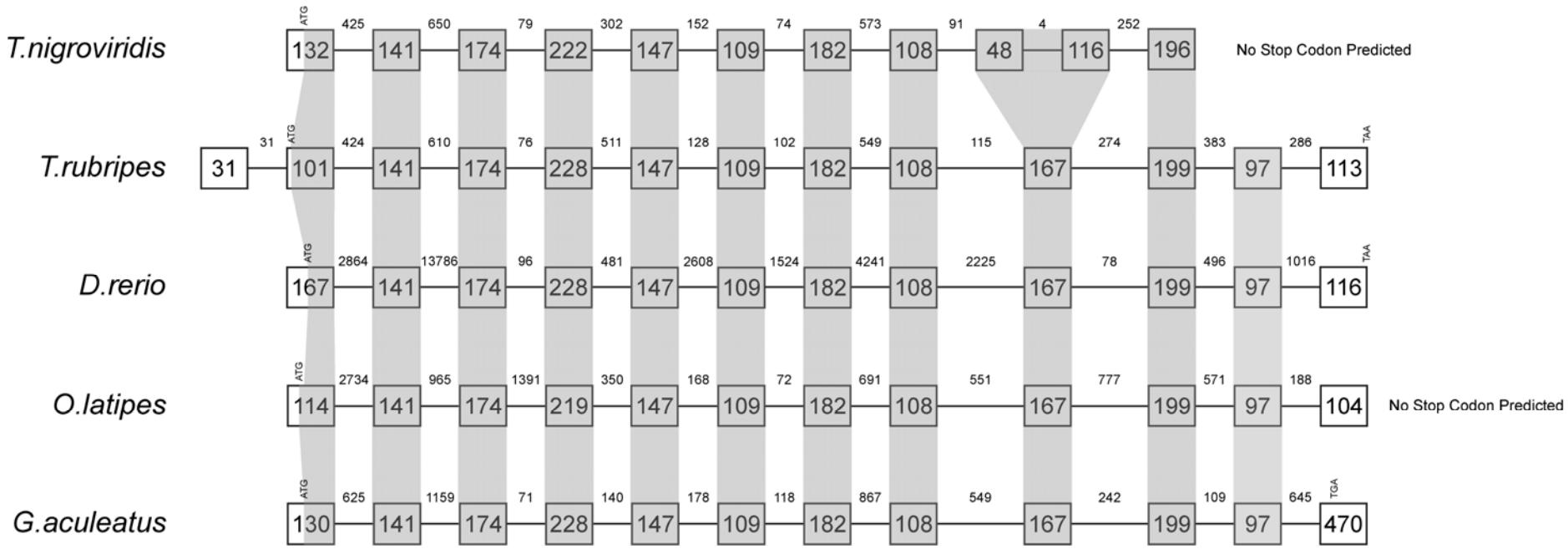


Figure 4-2 Exon-intron Structures of *zgc:63599* Orthologues in Teleost Species. The exon-intron information is based on Ensembl release 59 (August 2010). The Ensembl gene IDs are shown in Table 4.1 and 4.2. The locations of the start and stop codons are shown. Small black blocks represent exons and the black lines represent introns. Both black blocks and lines are not shown to scale. Exons with aligned nucleotide lengths are highlighted in grey.

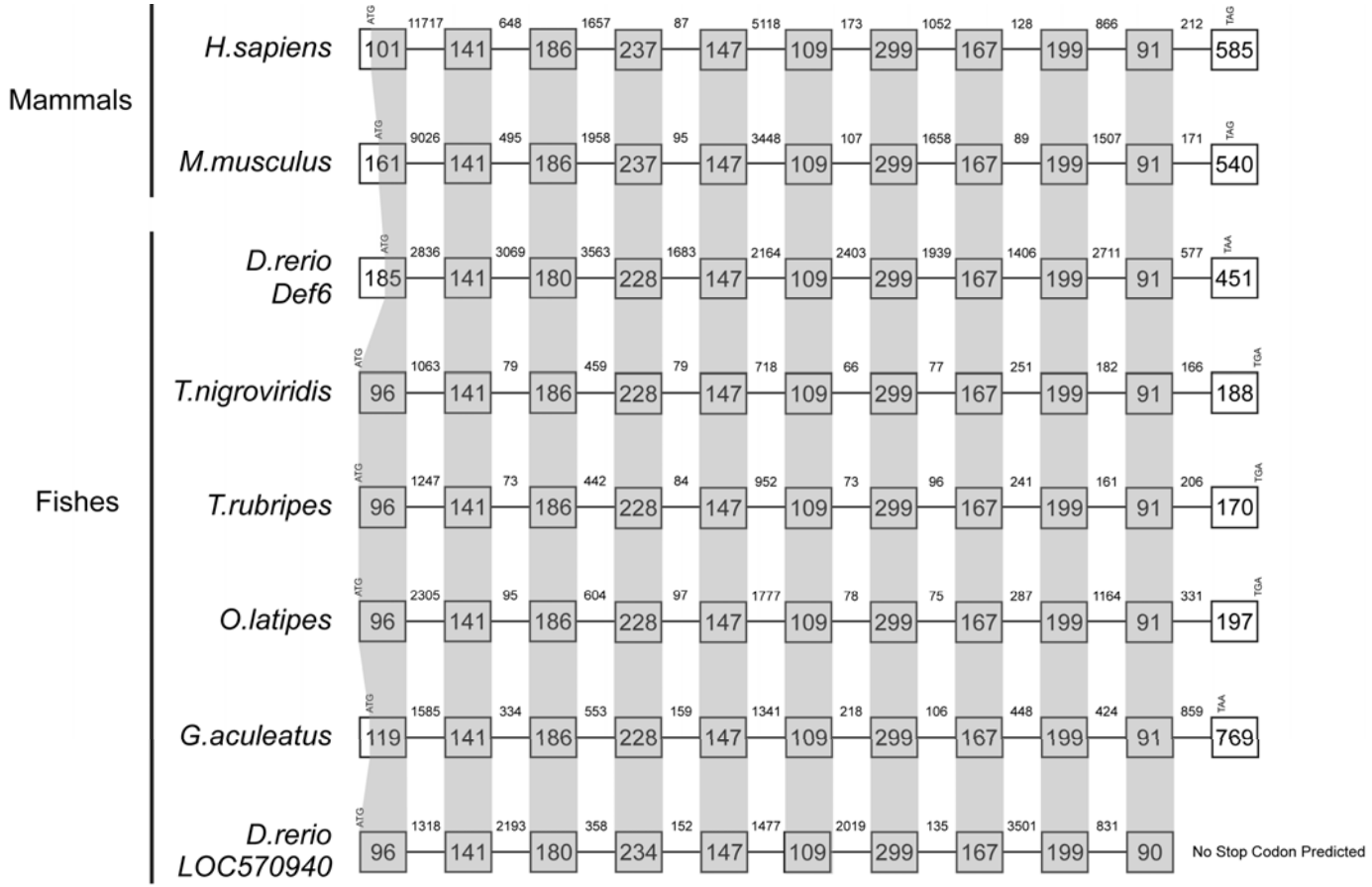


Figure 4-3 Exon-intron Structures of *Def6* Orthologues and *LOC570940*. The exon-intron information is based on Ensembl release 59 (August 2010). The Ensembl gene IDs are shown in Table 4.1 and 4.2. The locations of the start and stop codons are shown. Small black blocks represent exons and the black lines represent introns. Both black blocks and lines are not shown to scale. Exons with aligned nucleotide lengths are highlighted in grey.

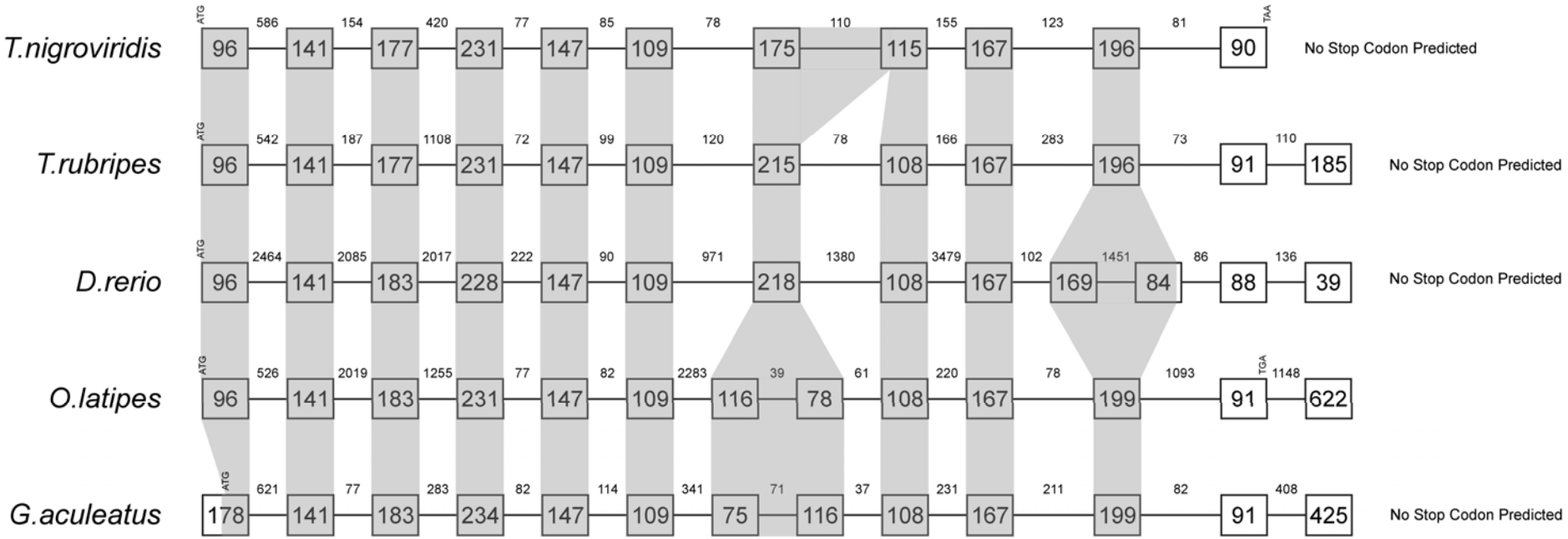


Figure 4-4 Exon-intron Structures of *si:dkeyp-15g12.1* Orthologues in Teleost Species. The exon-intron information is based on Ensembl release 59 (August 2010). The Ensembl gene IDs are shown in Table 4.1 and 4.2. The locations of the start and stop codons are shown. Small black blocks represent exons and the black lines represent introns. Both black blocks and lines are not shown to scale. Exons with aligned nucleotide lengths are highlighted in grey.

4.3 Amino acid sequences of SWAP70 and Def6 are highly conserved among vertebrate species

Multiple amino acid sequence alignments of the Def6 orthologues and SWAP70 orthologues were generated using ClustalW2 (Larkin *et al*., 2007) and the alignment regions above 50% identity were coloured using Jalview 2.5.1 release (Waterhouse *et al*., 2009; see also Figure 3.5 and 3.6). There are three different conservation levels that are indicated with dark (>80% identical), light (>70% identical), and very light (>57% identical).

As shown in Figure 4.5, the amino acid sequences of SWAP70 orthologues are extremely conserved and only three gaps had to be introduced for optimal alignment (positions 2, 34 and 35). Moreover, 48% amino acid positions are 100% identical and 30% positions show >70% identity. In comparison, DEF6 orthologues are slightly less well conserved (Figure 4.6). Gaps were introduced to Def6 orthologue alignment from the positions 107 to 180 and after position 557. Also, only 40% amino acid positions were 100% identical and 28% positions were >70% identical.

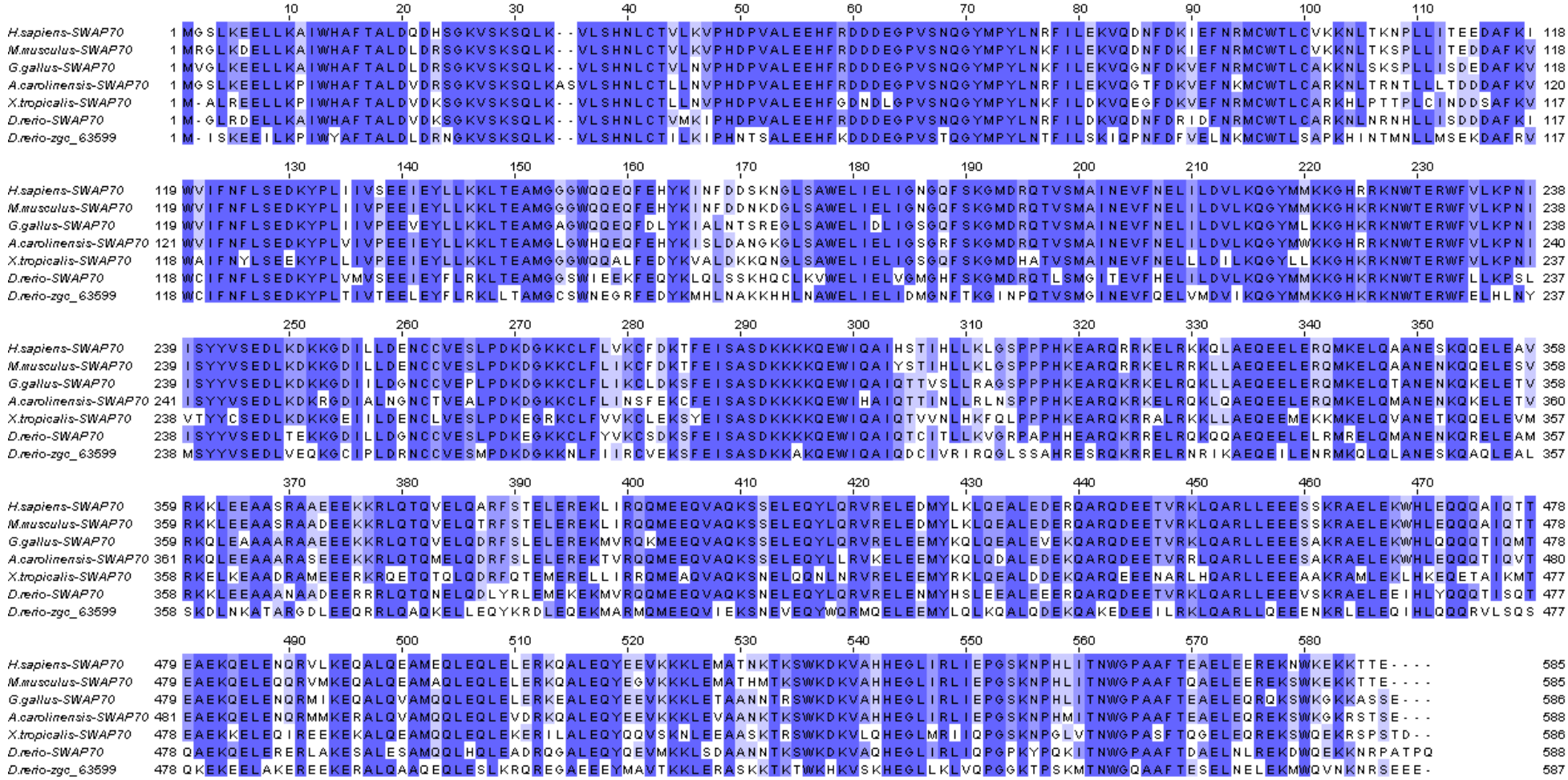


Figure 4-5 Multiple Amino Acid Sequence Alignment of SWAP70 Orthologues. Full length amino acid sequences were aligned using ClustalW2 online programme (Larkin *et al*., 2007) with all default settings. The alignment was coloured based on the setting of 'above 50% identity threshold' using Jalview 2.5.1 release (Waterhouse *et al*., 2009).

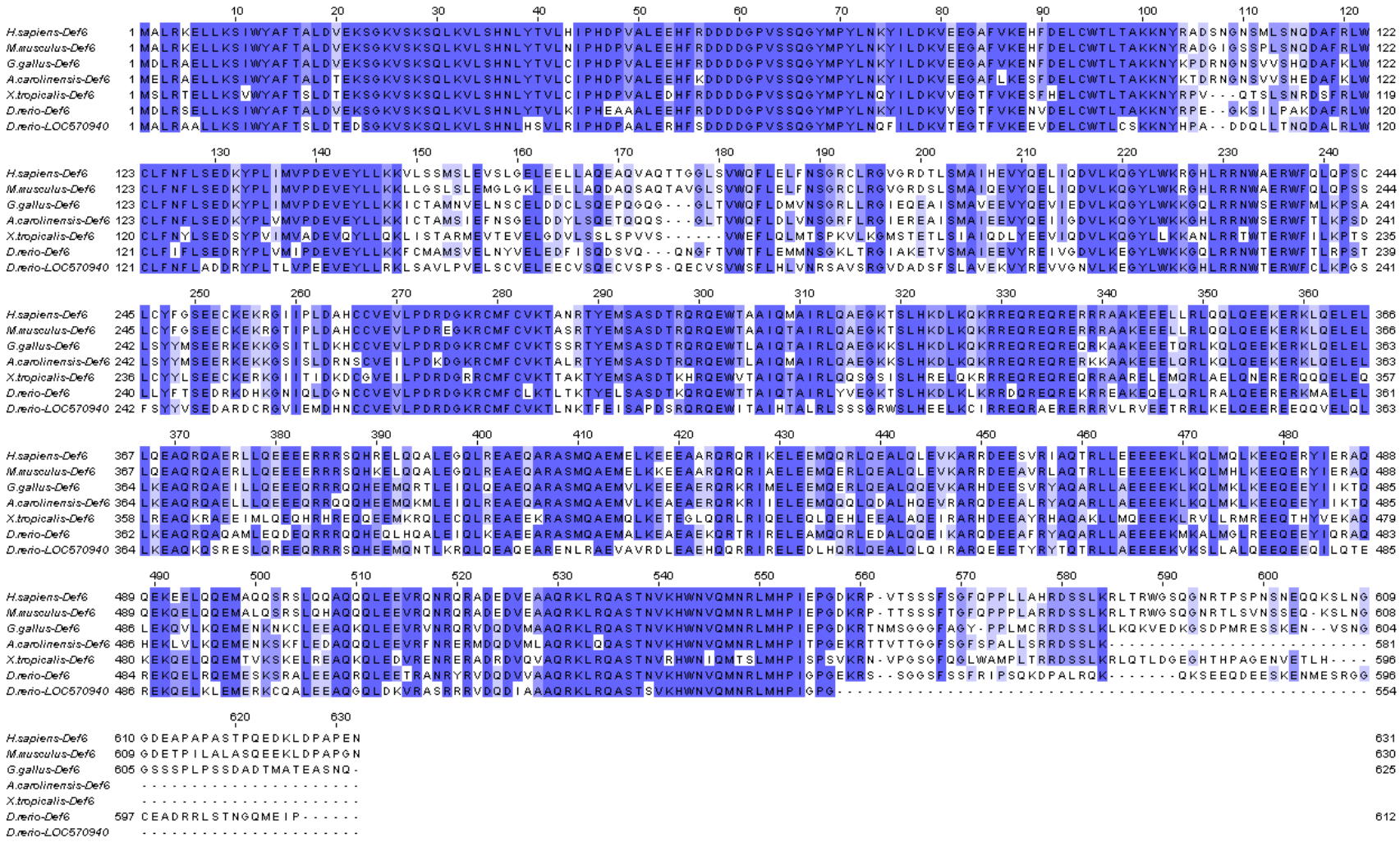


Figure 4-6 Multiple Amino Acid Sequence Alignment of Def6 Orthologues. Full length amino acid sequences were aligned using ClustalW2 online programme (Larkin *et al*., 2007) with all default settings. The alignment was coloured based on the setting of ‘above 50% identity threshold’ using Jalview 2.5.1 release (Waterhouse *et al*., 2009).

4.4 *Def6* and *LOC570940* as well as *SWAP70* and *zgc:63599* are phylogenetically identified as duplicated genes

In order to understand the phylogenetic relationships of Def6-related genes among vertebrate species, the unrooted gene trees of Def6 protein family based on amino acid sequences were constructed using neighbour-joining (NJ) method with 2000 bootstrapping and maximum likelihood (ML) method with 1000 bootstrapping.

According to the phylogenetic tree shown in Figure 4.7, both NJ and ML methods well support the formations of three clades, *Def6* clade, *si:dkeyp-15g12.1* clade and *SWAP70* clade. The orthologues of *si:dkeyp-15g12.1* in teleost are clearly separated from *Def6* and *SWAP70* with high bootstrap value. Also, *Def6* and *LOC570940* are grouped into the same teleost lineage in *Def6* clade while *SWAP70* and *zgc:63599* are put together into the same teleost lineage in *SWAP70* clade. This suggests that *Def6* and *LOC570940* are duplicated genes and SWAP70 and *zgc:63599* are duplicated genes in zebrafish. However, the duplication event of *SWAP70* and *zgc:63599* are poorly supported by the bootstrap values of NJ (53) and ML (61) methods. Therefore, the other two phylogenetic trees (Figure 4.8 and 4.9) were constructed with the orthologues of *Def6* and *SWAP70* from other selected vertebrates namely anole lizard, chicken, and frog.

As shown in Figure 4.8, the teleost lineages are separated from selected vertebrates with higher bootstrap values. The topology of the *Def6* orthologue tree (Figure 4.8) is the same as the topology of *Def6* clade shown in Figure 4.7. In addition, the separation of teleost lineage and selected vertebrate lineage is highly supported by both NJ bootstrap value (87) and ML bootstrap value (92) while most of the bootstrap values for both trees are more and less the same in Figure 4.8. For the *SWAP70* gene tree (Figure 4.9), the duplication event is highly supported with 100 NJ bootstrap value and 100 ML bootstrap value. Also, the bifurcation of *G.aculeatus* and *O.latipes* in SWAP70 orthologues (Figure 4.9) improves much compared to the previous tree (Figure 4.7).

Consequently, according to these phylogenetic tree analyses, *Def6* and *LOC570940* in zebrafish are co-orthologues of *DEF6* in human whereas *SWAP70* and *zgc:63599* are co-orthologues of *SWAP70* in human. And thus, *Def6* and *LOC570940* are named as *def6a* and *def6b* while *SWAP70* and zg*c:63599* are named as *swap70a* and *swap70b*, respectively in zebrafish. Besides, the names of *SWAP70-1* and *SWAP70-2* in other 4 teleost species are properly changed to

swap70a and *swap70b*. As there is no *LOC570940* orthologue identified, *def6* is used instead of *def6a* in the other 4 teleost species. In addition, the clade of *si:dkeyp-15g12.1* is phylogenetically closer to that of Def6, so they are referred to as *def6-like* genes.

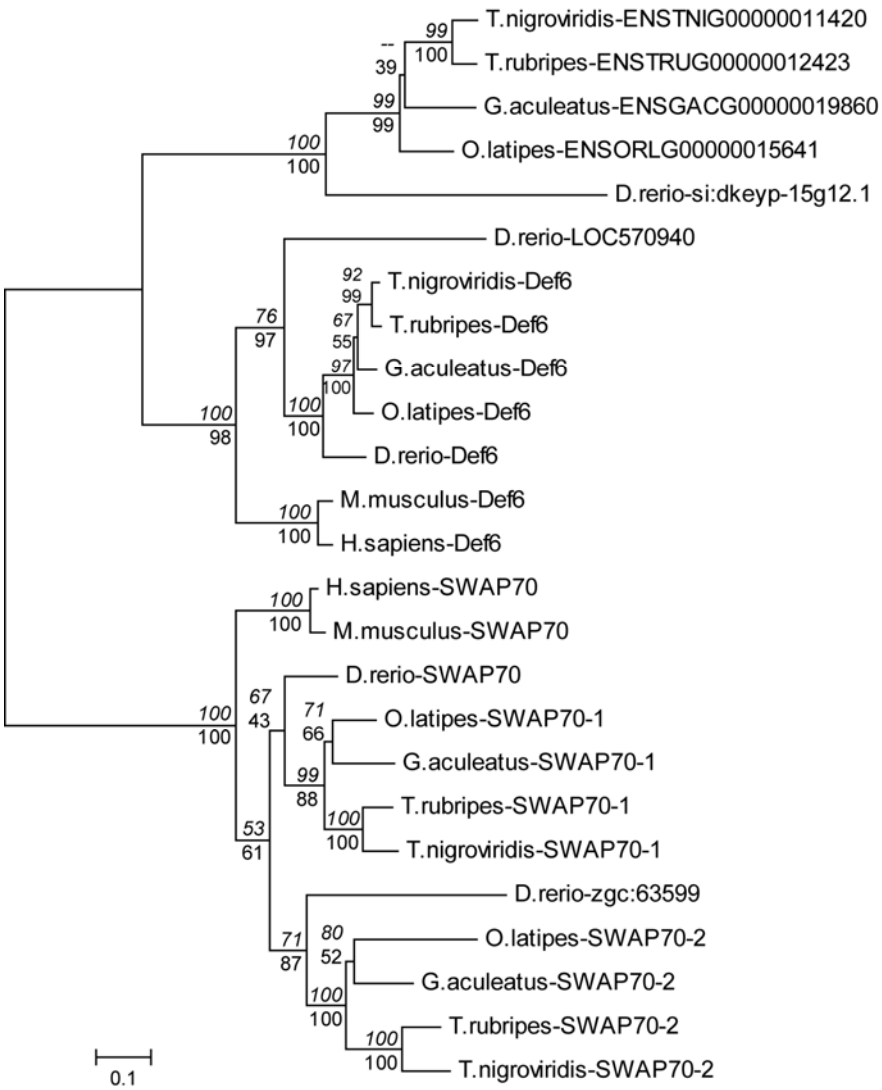


Figure 4-7 Phylogenetic Tree Analysis of DEF6/SWAP70 Family in Vertebrates. The bootstrap percentages from neighbour-joining tree (in *italic* font) and maximum likelihood tree (in regular font) are shown. The tree topology and branch length are based on ML tree. The NJ tree was constructed using MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007) with the settings of JTT+G as the substitution model, pairwise deletion, random seed, and 2000 bootstrapping. The ML tree was constructed using PhyML 3.0 (Guindon & Gascuel, 2003) in the PhyML server, http://www.atgc-montpellier.fr/phyml/ with the settings of JTT+G as the substitution model, SPR tree improvement, topology and branch lengths optimising for tree searching, and 1000 bootstrapping for branch support. The NJ and ML trees were constructed with the same manually adjusted multiple sequence alignment which was performed using ClustalW2 with default settings (Larkin *et al*., 2007) and Jalview 2.5.1 release (Waterhouse *et al*., 2009). The sequences involved in tree constructions are shown in Table 4.1 and 4.2.

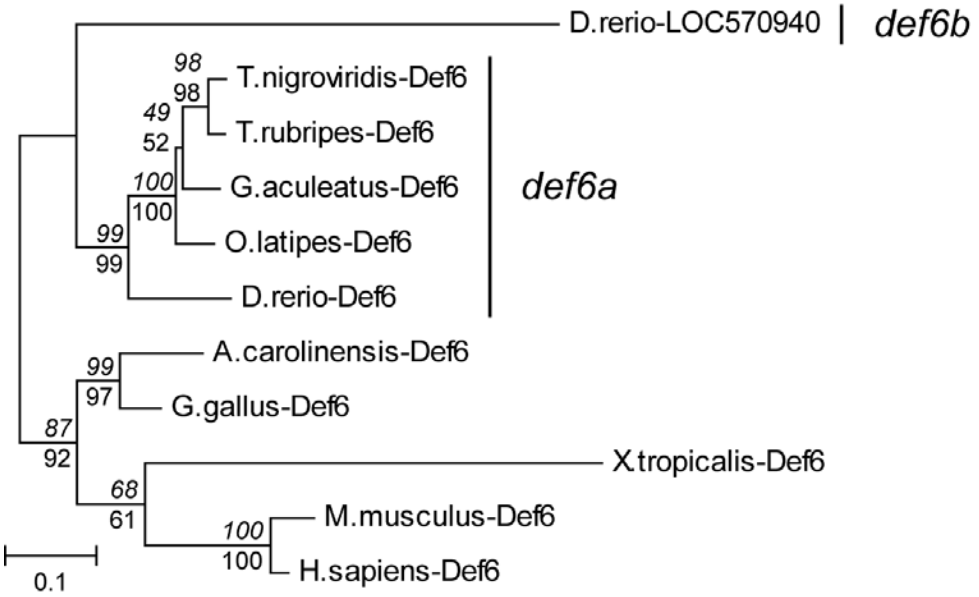


Figure 4-8 Phylogenetic Tree of Def6 Orthologues. The bootstrap percentages from neighbour-joining tree (in *italic* font) and maximum likelihood tree (in regular font) are shown. The tree topology and branch length are based on ML tree. The NJ tree was constructed using MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007) with the settings of JTT+G as the substitution model, pairwise deletion, random seed, and 2000 bootstrapping. The ML tree was constructed using PhyML 3.0 (Guindon & Gascuel, 2003) in the PhyML server, http://www.atgc-montpellier.fr/phyml/ with the settings of JTT+G+F as the substitution model, SPR tree improvement, topology and branch lengths optimising for tree searching, and 1000 bootstrapping for branch support. The NJ and ML trees were constructed with the same manually adjusted multiple sequence alignment which was performed using ClustalW2 with default settings (Larkin *et al*., 2007) and Jalview 2.5.1 release (Waterhouse *et al*., 2009). The sequences involved in tree constructions are shown in Table 4.1 and 4.2.

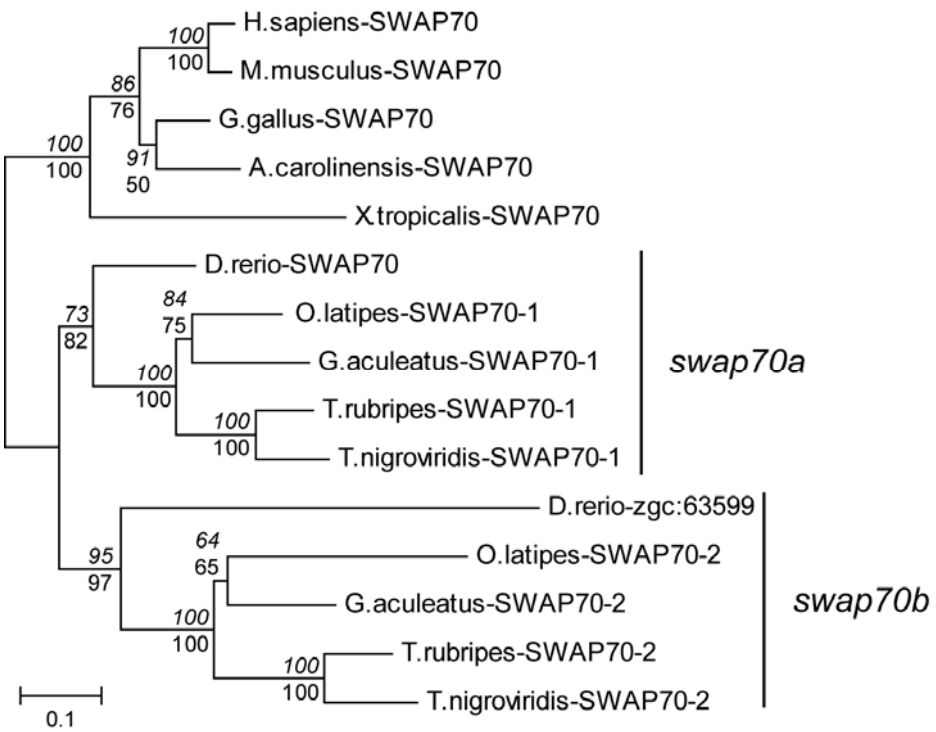


Figure 4-9 Phylogenetic Tree of SWAP70 Orthologues. The bootstrap percentages from neighbour-joining tree (in *italic* font) and maximum likelihood tree (in regular font) are shown. The tree topology and branch length are based on ML tree. The NJ tree was constructed using MEGA 4.0 (Kumar *et al*., 2008 and Tamura *et al*., 2007) with the settings of JTT+G as the substitution model, pairwise deletion, random seed, and 2000 bootstrapping. The ML tree was constructed using PhyML 3.0 (Guindon & Gascuel, 2003) in the PhyML server, http://www.atgc-montpellier.fr/phyml/ with the settings of JTT+I+G+F as the substitution model, SPR tree improvement, topology and branch lengths optimising for tree searching, and 1000 bootstrapping for branch support. The NJ and ML trees were constructed with the same manually adjusted multiple sequence alignment which was performed using ClustalW2 with default settings (Larkin *et al*., 2007) and Jalview 2.5.1 release (Waterhouse *et al*., 2009). The sequences involved in tree constructions are shown in Table 4.1 and 4.2.

4.5 All DEF6 and SWAP70 in selected vertebrates have similar folding

All DEF6 proteins in selected vertebrates were predicted to fold into ‘donut shaped’ with the C-scores which are higher than the accuracy cut-off -1.50 value (Figure 4.10). Also there are 2 to 3 contact points formed between N-terminus and C- terminus in each DEF6 model. In comparison, most of the SWAP70 proteins are also folded into ‘donut shapes’ with varied C-scores. The number of contact points formed between N-terminus and C-terminus range from 0 to 2. However, human SWAP70 is predicted to fold like a ‘C-shape’, despite the fact that human and gorilla SWAP70 differ by only one amino acid and the gorilla SWAP70 protein is predicted to fold like a ‘donut-shape’.

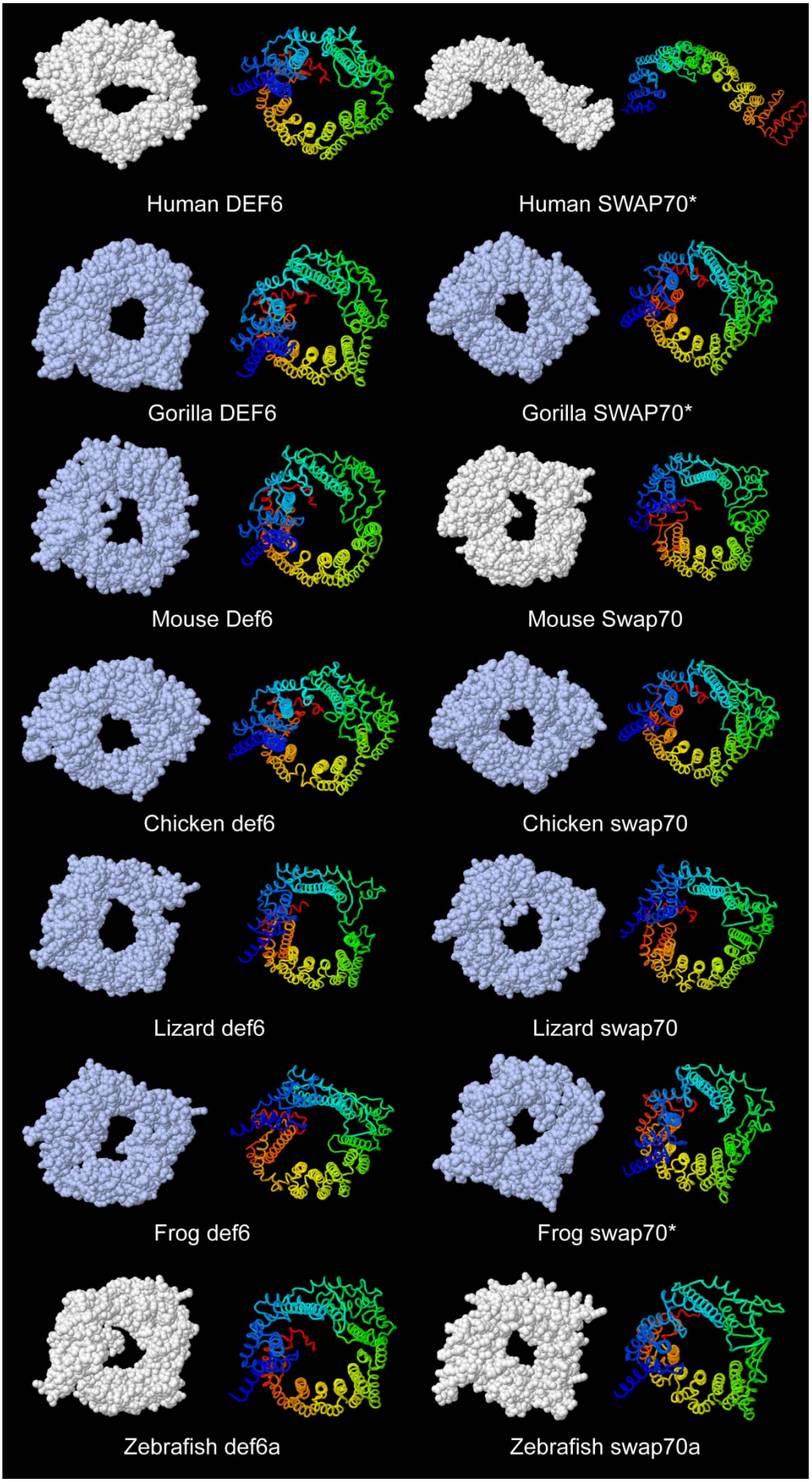

Human DEF6
Human SWAP70*
Gorilla DEF6
Gorilla SWAP70*
Mouse Def6
Mouse Swap70
Chicken def6
Chicken swap70
Lizard def6
Lizard swap70
Frog def6
Frog swap70*
Zebrafish def6a
Zebrafish swap70a

Figure 4-10 Three Dimensional Structure Predictions of DEF6 and SWAP70 Orthologues in Selected Vertebrates. Full length amino acid sequences were submitted to I-TASSER (Roy *et al*., 2010, Zhang, 2009 and Zhang, 2008) to generate predicted 3D structure models. The molecular contact and rainbow coloured models were visualised using First Glance in Jmol, version 1.45. N-terminus is in blue and C-terminus is in red. The C-scores for all models shown here are higher than -1.5 towards positive side.

5. Identification and Analysis of *def6*-related Genes in Invertebrates

5.1 Data acquisition

As of August 2010, there is no literature that reported the presence of def6-related gene in any invertebrates. However, in this chapter, the identification and bioinformatics analyses of def6-related gene in invertebrate species are presented. Blastp and tblastn with the protozoa, plant, and fungi genomes were carried out in NCBI based on the zebrafish def6 and swap70 amino acid sequences. But, only PH domain region was hit as a result. Similarly, there is no gene identified in fruit fly, honey bee, nematode, and cnidaria species. Nevertheless, blastp and tblastn with other invertebrate genomes in NCBI and blast search in Uniprot showed some putative def6-related genes. Only one gene was identified in each species. Furthermore, in Ensembl release 59 (5 August 2010) and EnsemblMetazoa databases, there are five genes identified which are related to def6 in five species. The details of the putative def6-related genes identified are shown in Table 5.1.

Common Name	Species	Offical Symbol in NCBI	Gene Name[1]	NCBI Gene ID	Ensembl Gene ID	UniProt ID	No. of Exons	Genomic DNA Length (kb)	Transcript Length (bp)	Peptide Length (aa)	Genome Location	Strand
Southern house mosquito	*C.quinquefasciatus*	CpipJ_CPIJ010369	*def6/swap70*	6042925	CPIJ010369	B0WTE5	3	3.89	1644	547	supercont3.273: 149,534-153,424	Reverse
Red flour beetles	*T.castaneum*	LOC663300	*def6/swap70*	663300	---	---	3	2.01	1800	531	Chromosome LG8: 16,315,657-16,317,664	Forward
Yellow fever mosquito	*A.aegypti*	AaeL_AAEL011412	*def6/swap70*	5574780	---	Q0IEB4	4	8.09	1875	547	supercont1.579: 453,273-461,360	Reverse
African malaria mosquito	*A.gambiae*	AgaP_AGAP012332	*def6/swap70*	4577474	AGAP012332-PA	A0NGI8	4	2.47	1850	555	Chromosome 3L: 40,710,928-40,713,401	Reverse
Florida lancelet	*B.floridae*	BRAFLDRAFT_239050	*def6/swap70*	7253503	---	C3Z9A4	5	4.55	786	261	422,063-426,707	Reverse
Jewel wasp	*N.vitripennis*	LOC100123666	*def6/swap70*	100123666	---	---	7	2.51	1608	535	Chromosome not placed: 3,588,750-35,91,256	Reverse
Human louse	*P.humanus*	Phum_PHUM446210	*def6/swap70*	8231186	PHUM446210	---	8	2.30	1617	538	DS235780: 105,002-107,304	Forward
Hemichordata	*S.kowalevskii*	LOC100369873	*def6/swap70*	100369873	---	---	8	19.43	1520	421	Chromosome not placed: 111,853-131,277	Reverse
Sea urchin	*S.purpuratus*	LOC585280	*def6/swap70*	585280	---	---	8	16.21	1474	454	Chromosome not placed: 57-16,263	Reverse
Placozoa	*T.adhaerens*	TRIADDRAFT_61339	*def6/swap70*	6758568	---	B3SAQ2	8	7.44	1761	586	338,510-345,949	Reverse
Black-legged tick	*I.scapularis*	IscW_ISCW006031	*def6/swap70*	8053036	ISCW006031	B7PKR8	9	33.06	1479	493	DS735295: 63,935-96,998	Reverse
Pea aphid	*A.pisum*	LOC100169581	*def6/swap70*	100169581	---	---	10	6.60	1601	444	Chromosome not placed: 169,386-175,986	Reverse
Tunicate	*C.intestinalis*	LOC100176638	*def6/swap70*	100176638	ENSCING00000009911	---	11	4.18	2752	602	Chromosome 4q: 1,399,201-1,403,380	Forward

[1] The gene names which are proposed in this thesis.

Table 5.1 The Summary of def6/swap70 Gene Family in Invertebrate.

5.2 Exon-intron structures of invertebrates are simpler than those of vertebrates

The exon-intron structures of invertebrates which are available in Ensembl database were selected and compared to others. The aligned nucleotide regions are indicated with black lines based on the multiple amino acid sequence alignment.

As shown in Figure 5.1, the exon-intron structures of mosquito species are much simpler than others. They have only 3-4 exons in total. In comparison, human louse (*P.humanus*), black-legged tick (*I.scapularis*), and tunicate (*C.intestinalis*) have 8, 9, and 11 exons, respectively. Both human SWAP70 and zebrafish swap70a have 12 exons. Relatively, southern house mosquito (*C.quinquefasciatus*) has the simplest structure. Exon 2 is split in two separate exons in yellow fever mosquito (*A.aegypti*) and african malaria mosquito (*A.gambiae*). Those two exons are further split into 7 exons in human louse. Black-legged tick show two more exon splits in the first 5 exons of human louse. Tunicate exhibit additional exon splitting and joining events and thus the overall exon-intron structure is much closer to that of vertebrate species.

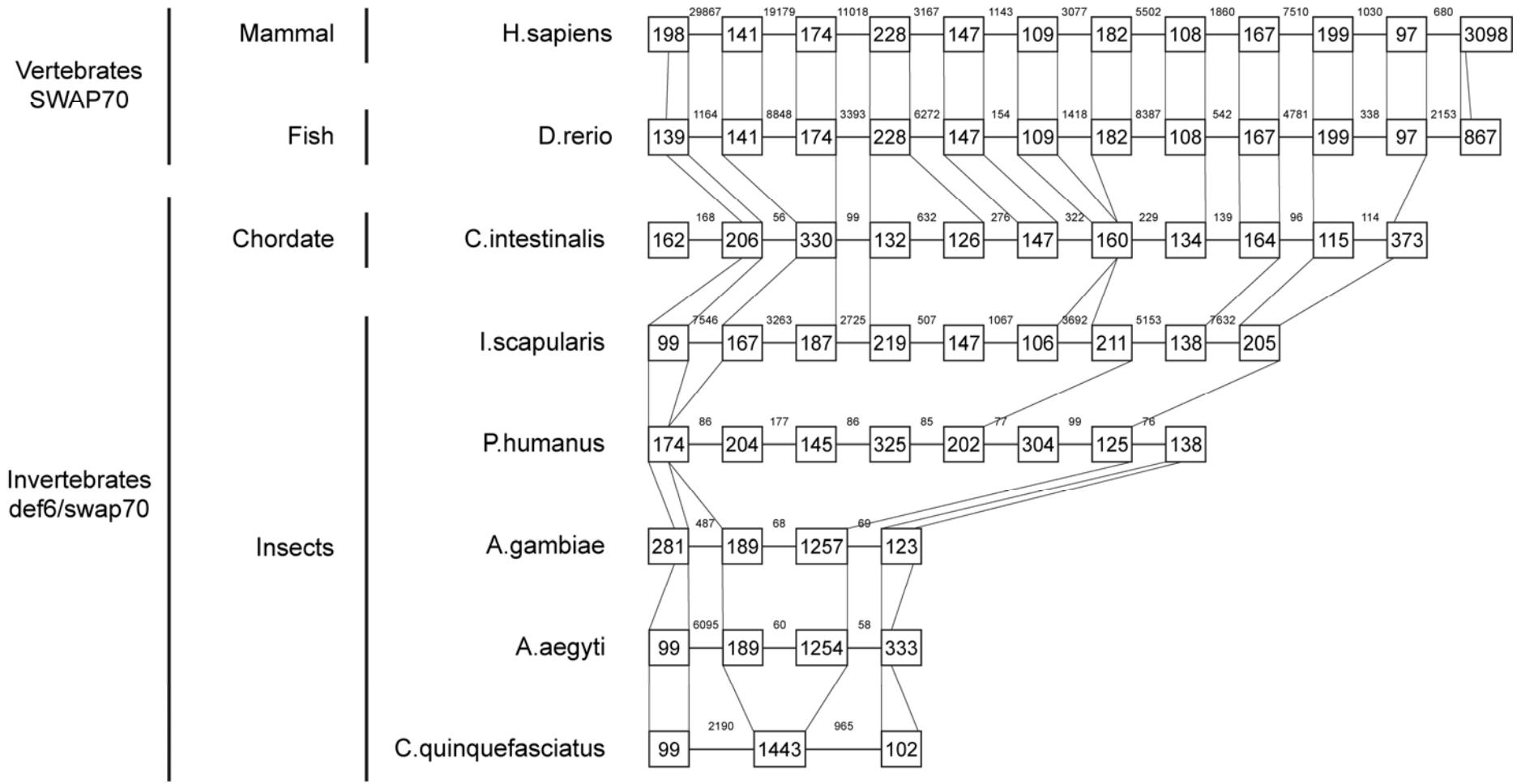


Figure 5-1 Comparison of Exon-intron Structures among Invertebrate def6/swap70 and Vertebrate SWAP70. The exon-intron information is based on Ensembl release 59 (August 2010). The Ensembl gene IDs for each invertebrate gene are shown in Table 5.1. Small black blocks represent exons and the black horizontal lines represent introns. Both black blocks and horizontal lines are not shown to scale. The lines between exon-intron structures indicate the aligned nucleotide regions.

5.3 Amino acid sequences of def6/swap70 in invertebrates are statistically closer to SWAP70 in vertebrates

Multiple amino acid sequence alignments of SWAP70 and DEF6 from six selected vertebrates and def6/swap70 from six selected invertebrates were generated using ClustalW2 (Larkin *et al*., 2007) and the alignment regions above 60% identity were coloured using Jalview 2.5.1 release (Waterhouse *et al*., 2009). Pairwise analysis was performed using EMBOSS open source software analysis package (Rice *et al*., 2000) with settings of 10.0 for gap open, 0.5 for gap extend, and blosum50 for matrix. Full length amino acid sequence was used in both analyses. The alignments of SWAP70 against def6/swap70 and DEF6 against def6/swap70 are shown in Figure 5.2 and 5.3, respectively. There are two different conservation levels that are indicated in dark (>80% identical) and light (>67% identical). Pairwise analysis results are shown in Figure 5.4.

In Figure 5.2, 11.4% of the amino acid positions show >80% identity and 14.6% positions show >67% identity along the multiple sequence alignment. However, there are only 9.4% positions having >80% identical and 11.6% position are >67% identical in Figure 5.3. Therefore, the def6/swap70 in invertebrates and SWAP70 in vertebrates are more similar to each other than to DEF6 in vertebrates.

According to the pairwise analysis (Figure 5.4), with exception of *B.floridae* and *C.intenstinalis*, percentage identities between full length amino acid sequences of invertebrates and vertebrates vary from 22.0% to 28.4%. Similarly, the similarities of amino acid sequences between two groups are ranged from 37.4% to 47.7%. Excluding two species, *N.vitripennis* and *T.castaneum*, the percentage identities between invertebrate def6/swap70 and human SWAP70 are higher than those between def6/swap70 and human DEF6. In addition, with few exceptions, the similarities of def6/swap70 compared to human SWAP70 are also higher.

Importantly, both results of multiple amino acid sequence alignments and pairwise analysis suggest that def6/swap70 genes in invertebrates are statistically closer to SWAP70 orthologues in vertebrates rather than DEF6 orthologues.

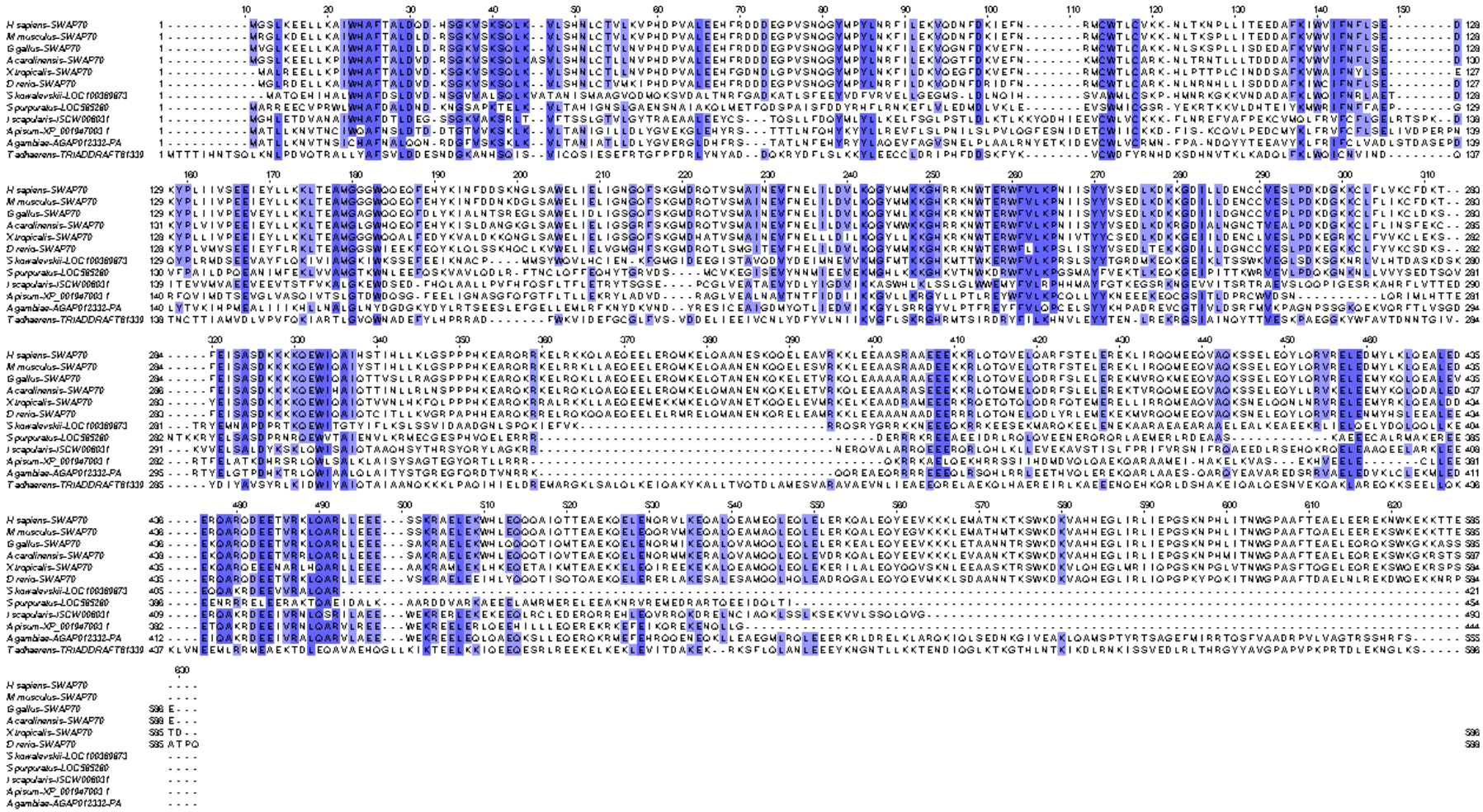


Figure 5-2 Multiple Amino Acid Sequence Alignment of def6/swap70 in Selected Invertebrates and SWAP70 in Selected Vertebrates. Full length amino acid sequences were aligned using ClustalW2 online programme (Larkin *et al*., 2007) with all default settings. The alignment was coloured based on the setting of 'above 60% identity threshold' using Jalview 2.5.1 release (Waterhouse *et al*., 2009).

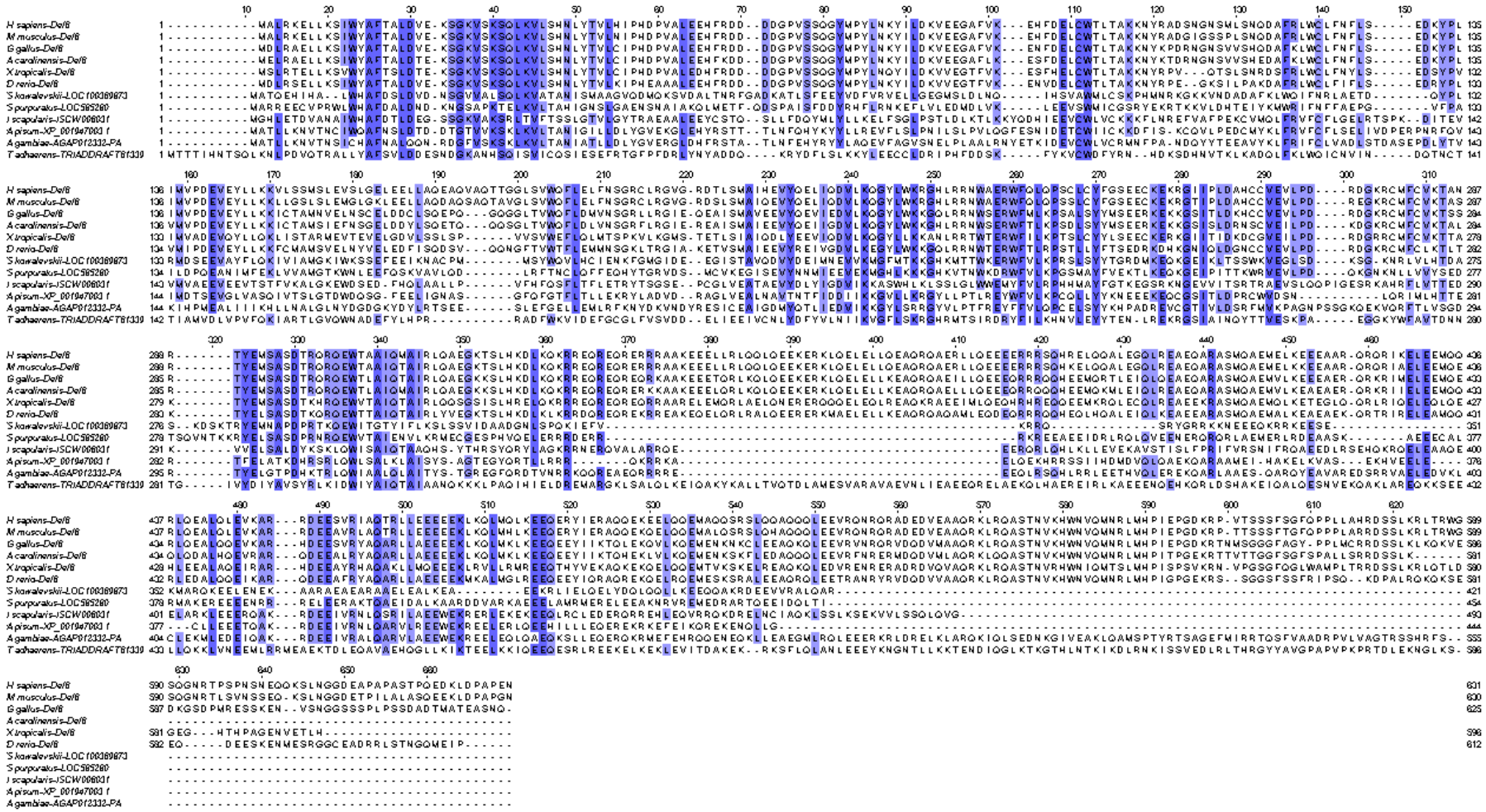


Figure 5-3 Multiple Amino Acid Sequence Alignment of def6/swap70 in Selected Invertebrates and DEF6 in Selected Vertebrates. Full length amino acid sequences were aligned using ClustalW2 online programme (Larkin *et al*., 2007) with all default settings. The alignment was coloured based on the setting of ‘above 60% identity threshold’ using Jalview 2.5.1 release (Waterhouse *et al*., 2009).

Identity%/Similarity%/Gap%	*D.rerio*		*H.sapiens*	
	def6a	**swap70a**	**DEF6**	**SWAP70**
A.aegypti	25.6/45.8/25.5	27.6/45.6/25.7	25.9/46.3/24.4	28.3/46.2/22.3
A.gambiae	28.4/44.8/27.4	27.2/44.5/23.3	27.9/47.3/24.8	28.1/47.7/22.2
A.pisum	24.6/40.8/34.5	22.3/39.9/35.9	24.5/40.3/36.4	24.8/39.7/37.4
B.floridae	16.6/26.0/60.8	18.5/28.2/58.3	15.6/25.8/62.1	19.6/29.2/57.3
C.intestinalis	18.8/33.2/39.6	18.2/33.8/39.2	19.3/33.8/37.1	20.6/36.2/36.7
C.quinquefasciatus	25.5/46.1/24.9	27.2/47.5/20.4	26.8/47.2/23.7	28.3/48.8/22.8
I.scapularis	26.6/41.2/35.1	26.1/38.4/32.1	24.1/37.4/37.6	25.7/41.0/31.3
N.vitripennis	28.0/45.5/23.8	25.8/43.2/27.2	27.9/46.4/26.7	26.5/45.3/27.4
P.humanus	23.5/43.3/27.1	25.9/46.0/25.7	24.7/43.7/27.8	27.2/46.5/23.4
S.kowalevskii	25.4/38.9/37.8	26.3/39.2/38.0	23.7/36.8/39.4	27.0/41.9/34.5
S.purouratus	27.2/42.8/32.4	24.3/42.0/32.2	24.9/40.6/38.5	25.0/41.7/31.6
T.adhaerens	22.0/39.4/33.4	24.9/42.0/26.8	22.5/44.1/28.6	24.8/45.2/21.8
T.castaneum	27.5/43.9/28.9	27.8/45.6/24.6	28.2/48.4/24.7	27.9/47.3/22.9

Figure 5-4 Pairwise Analysis among def6/swap70 Orthologues in Invertebrate and DEF6 and SWAP70 orthologues in Selected Vertebrates. The amino acid sequences of def6/swap70 orthologues in invertebrates were compared with those of zebrafish def6a and swap70 as well as human DEF6 and SWAP70. EMBOSS open source software analysis package (Rice *et al*., 2000) was used. The settings were 10.0 for gap open, 0.5 for gap extend, and blosum50 for matrix. The numbers indicate the percentages of identity, similarity and gap between amino acid sequences in order.

5.4 DEF6 and SWAP70 are co-orthologues of the gene identified in Invertebrates

In order to determine the relationships among def6/swap70 in invertebrate and DEF6 and SWAP70 in vertebrates, the gene trees manually rooted with *T.adhaerens* were constructed based on protein sequences using maximum likelihood (ML) method with 1000 bootstrapping.

The ML tree (Figure 5.5) shows that vertebrate DEF6 and SWAP70 were grouped in the same clade with high bootstrap value (97) supported. Also, arthropod species were separated from others with an absolute bootstrap value (100). On the contrary, other invertebrate species have not got a high bootstrap value to support the tree topology. Given that only one gene was identified in some invertebrate species, DEF6 and SWAP70 in vertebrates are clear orthologues of def6/swap70 as the result.

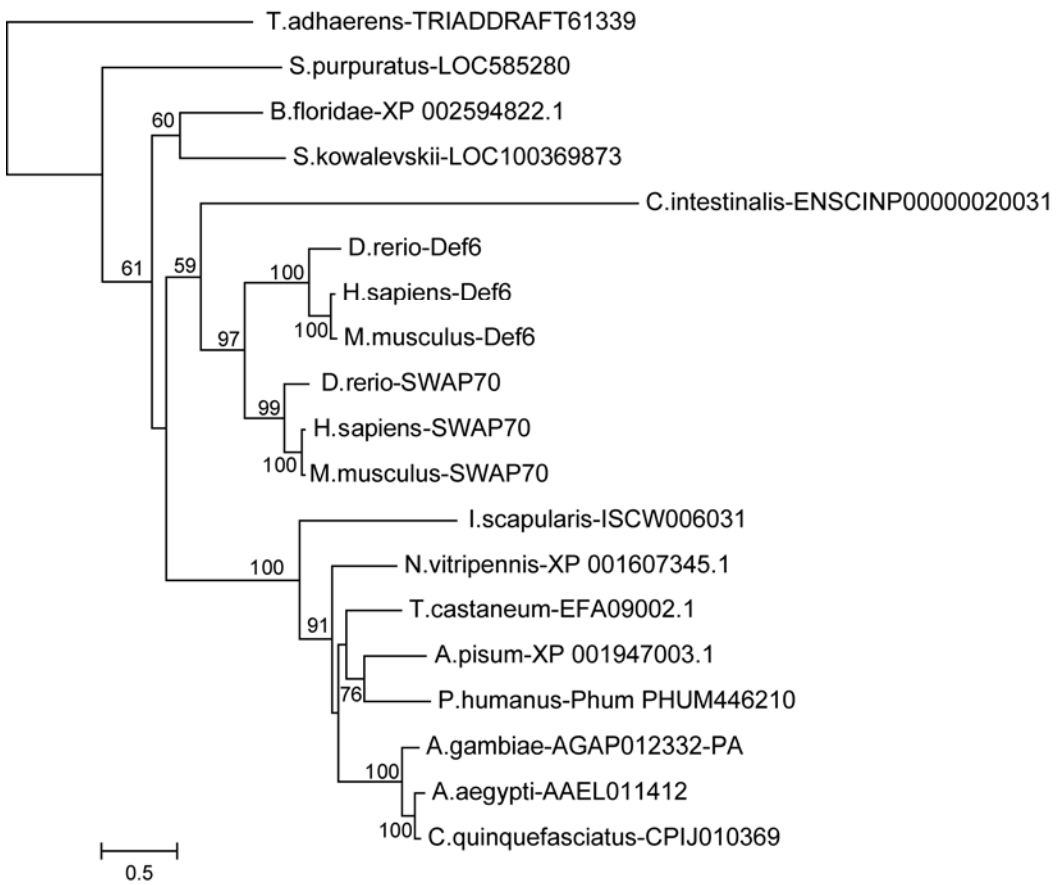


Figure 5-5 Maximum Likelihood Tree of def6/swap70 in Invertebrates. Amino acid sequences of invertebrate shown in Table 5.1 with vertebrate representatives were aligned using ClustalW2 (Larkin *et al*., 2007) with all default settings. Multiple sequence alignment were manually adjusted using Jalview 2.5.1 release (Waterhouse *et al*., 2009) and the tree was constructed using PhyML 3.0 (Guindon & Gascuel, 2003) in the PhyML server, http://www.atgc-montpellier.fr/phyml/ with the settings of JTT+I+G+F as the substitution model, SPR tree improvement, topology and branch lengths optimising for tree searching, and 1000 bootstrapping for branch support. The bootstrap values that are higher than 50 are shown in the tree.

5.5 'Donut shape' structures are predicted in lower invertebrate species

The proteins of def6/swap70 in tricoplax (*T.adhaerens*) and mosquitoes (*A.aegypti, C.quinquefasciatus*, and *A.gambiae*) were predicted to fold into 'donut shapes' with zero to two contact points. In comparison, with the exception of *P.humanus* and *S.purpuratus*, def6/swap70 proteins in other invertebrates were predicted to fold into a curved shaped without any contact point (Figure 5.6). Nevertheless, there is remarkable structural conservation between vertebrates and invertebrates. Some def6/swap70 in invertebrates were predicted to fold like a 'donut shape' which are similar to the folding of DEF6 and SWAP70 in vertebrates. In addition, although the predicted folding of human SWAP70 is not highly supported by I-TASSER, its folding is also similar to the predicted 'C-shaped' folding in invertebrates. Therefore, def6/swap70 in vertebrates is predicted to have similar folding of DEF6 and SWAP70 in vertebrates.

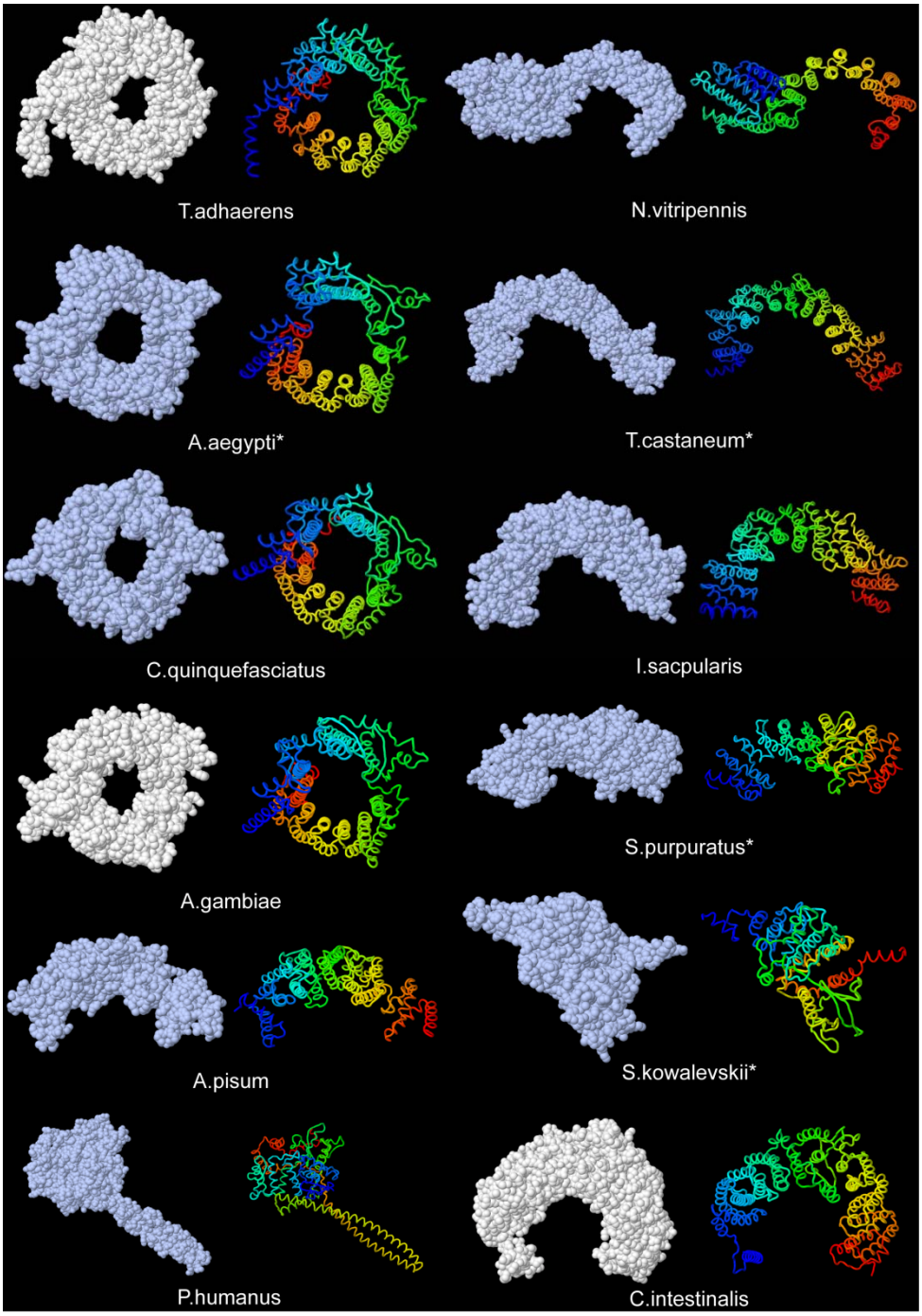


Figure 5-6 Three Dimensional Structure Predictions of def6/swap70 Orthologues in Invertebrates. Full length amino acid sequences were submitted to I-TASSER (Roy *et al*., 2010, Zhang, 2009 and Zhang, 2008) to generate predicted 3D structure models. The molecular contact and rainbow coloured models were visualised using First Glance in Jmol, version 1.45. N-terminus is in blue and C-terminus is in red. The star indicates the models which the C-scores are lower than the accuracy cut-off -1.5 value.

6. Functional Analysis of *swap70a* in Zebrafish Embryogenesis

6.1 Differential gene expression of the def6 paralogues during zebrafish development

To determine the expression profile of the five *def6* paralogues during zebrafish development, RT-PCR analysis was carried out using total RNA isolated from different developmental stages (1 to 72 hpf as indicated in Figure 6.1) and gene specific primer sets amplifying cDNA fragments of 100 bp to 300 bp length. The *DNaseI* treated total RNA without reverse transcription (RT-) was used as negative control in PCR reaction and the elongation factor *ef1a*, a housekeeping gene, was amplified as a positive control.

As shown in Figure 6.1, *swap70a* and *def6a* were expressed in all developmental stages tested. In contrast, *swap70b* was expressed from 24 hpf onwards and *def6b* expression was detectable only at 48 and 72 hpf. *def6-like* was transiently expressed at 3 hpf and then from 24 hpf onwards. In addition, the expression level of *swap70a* seemed to decrease from 1 hpf to 10 hpf whereas the expression level of *def6a* seemed to increase from 1 hpf to 6 hpf. Given that zygotic gene expression starts after midblastula transition (around 3 hpf), amplification of *swap70a* and *def6a* cDNA at 1 hpf indicated that both genes are also maternally expressed. Maternal *swap70a* expression seemed to be much higher than maternal *def6a* expression suggesting a more important role for *swap70a* than *def6a* at this time of development.

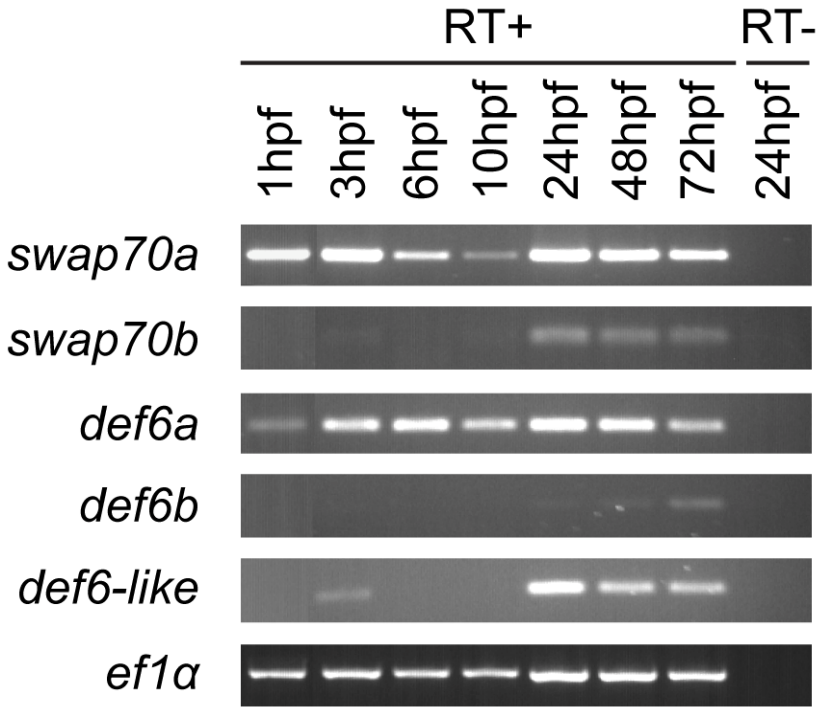


Figure 6-1 Expression Profiles of Def6 Paralogues during Zebrafish Development. RT-PCR with 35 cycles was performed to indicate the expression of five Def6 paralogues during zebrafish development at 1, 3, 6, 10, 24, 48 and 72 hours post-fertilisation (hpf). PCR reactions without reverse transcription (RT-) are shown for 24 hpf only. The *ef1α* amplification with 30 cycles and under less UV exposure served as internal control.

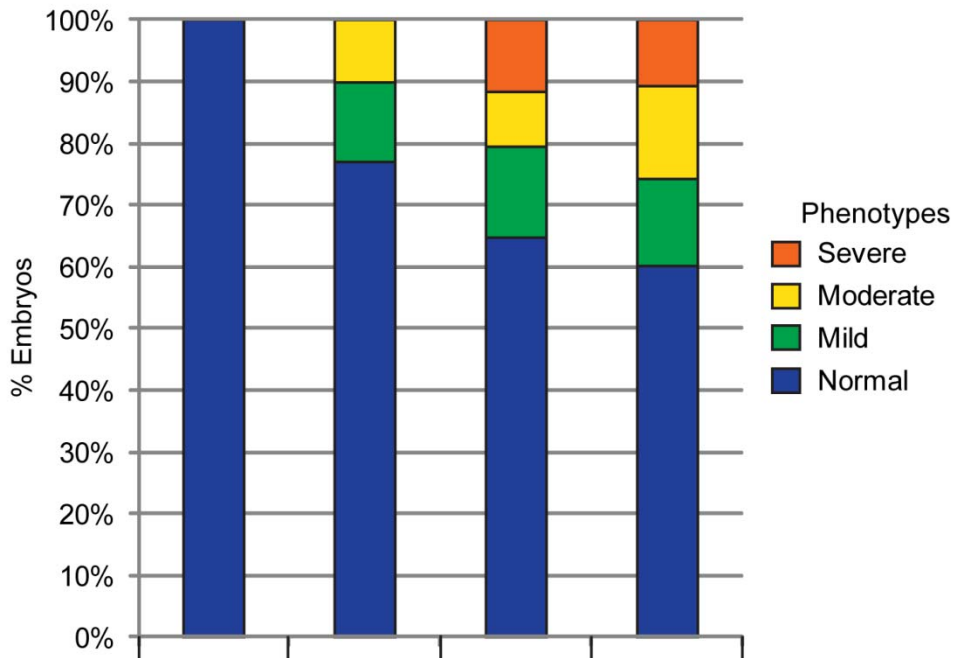
6.2 Gain-of function analysis of *swap70a*

6.2.1 Over-expression of swap70a reveals developmental defects in brain, eyes and tail formation

250pg, 500pg, and 750pg of *in vitro* transcribed GFP-tagged *swap70a* mRNA were injected into 2-4 cells stage embryos. Embryos were examined at 24 hpf under a fluorescent microscope to ensure that the GFP-*swap70a* protein was expressed and the phenotypes were examined using a stereomicroscope.

As shown in Figure 6.2, the *swap70a* overexpression defects were dose-dependent. At low amounts (250pg) of mRNA injection (see Figure 6.3), the body axis of embryos had not been significantly affected. Embryos showed curved tail formation in the mild phenotype group and no mid-hindbrain boundary formed and only one otolith formed in the severe group. Moreover, as the mRNA amount increased, embryos became shorter in body axis with multiple phenotypes. The forebrain region turned dark. Abnormal eyes and tails and an underdeveloped head were found.

As can be seen in Figure 6.3, the amount of GFP fluorescent was correlated with the severity of the phenotypes. More GFP fluorescence is due to higher *swap70a* transcript expression, resulting in the more severe phenotypes found in the head, the back, and the tail.



	WT	250pg mRNA	500pg mRNA	750pg mRNA
No. of Embryos	54	40	42	113
Survival %	100%	98%	81%	100%

Figure 6-2 Statistics of *in vitro* transcribed GFP-tagged *swap70a* RNA Injections. 250pg, 500pg, and 750pg of *swap70a* RNA were injected into embryos. The number of surviving embryos was determined and the phenotypes of embryos were examined at the 24 hpf stage. Based on the seriousness of the phenotypes, embryos were grouped into normal (blue), mild (green), moderate (yellow), severe (orange). Images of embryos injected with *swap70a* RNA are shown in Figure 6.3.

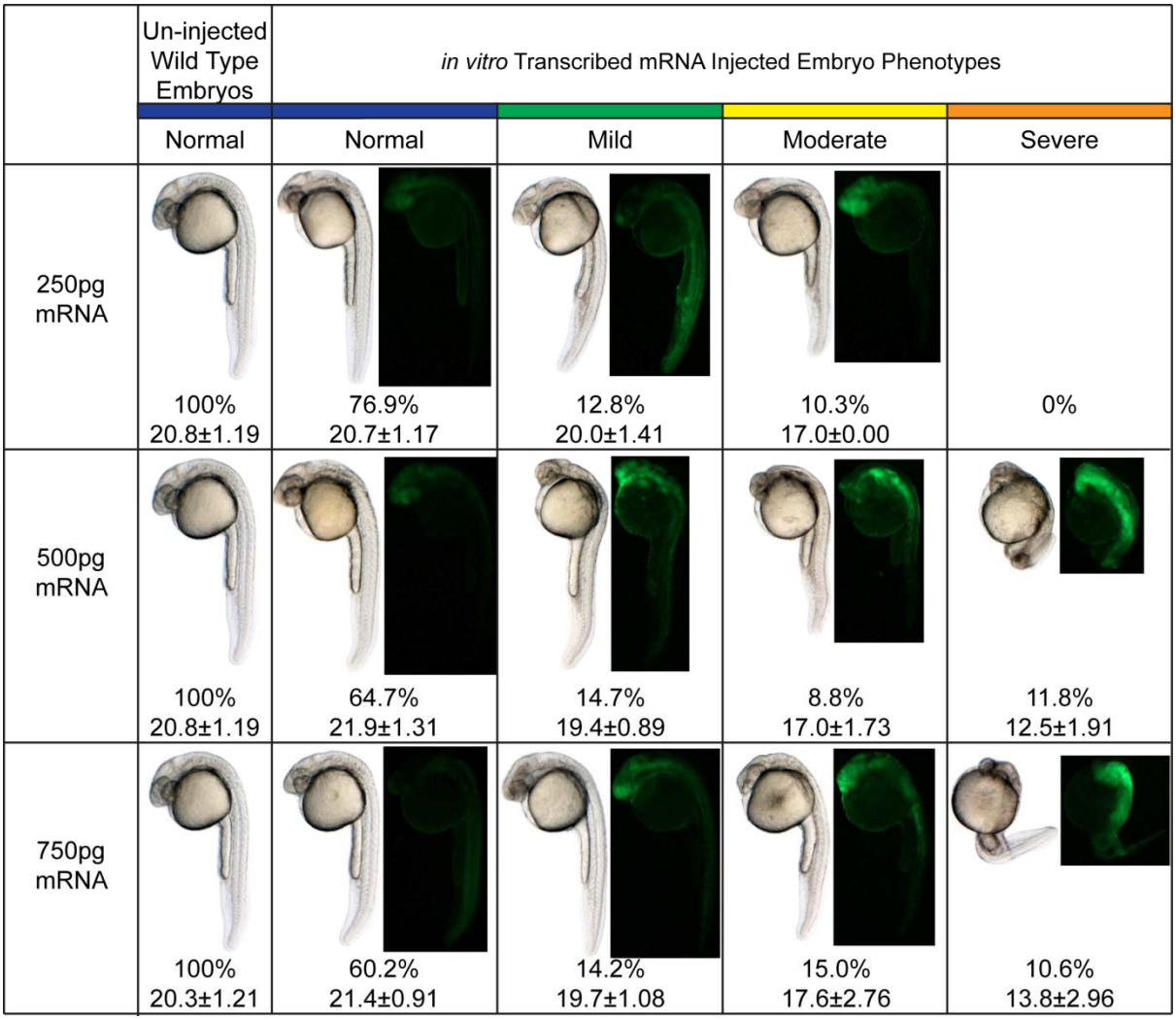


Figure 6-3 Dose-dependent Phenotypes of *swap70a* Overexpression. 250pg, 500pg, and 750pg of *in vitro* transcribed GFP-tagged mRNA were injected into embryos separately in three experiments. Based on the seriousness of the phenotypes, embryos were grouped into normal (blue), mild (green), moderate (yellow), and severe (orange). Un-injected wild type embryos served as the experimental control. The percentages indicate the number of embryos (see Figure 6.2) found in each phenotype group. The axes of embryos were measured from head to tail under microscopy with 10X magnification and were presented as mean ± standard derivation with unit, mm. Note that the intensity of GFP fluorescent is correlated with the severity of the phenotypes underlining that the observed alterations are due to *swap70a* overexpression. The enlarged images of phenotypes are shown in Appendices I to III.

6.3 Loss-of function analysis of *swap70a* by morpholino-mediated knockdown

As *swap70a* and *def6a* were expressed at all stages tested (Figure 6.1), they may have important roles in zebrafish embryogenesis. Therefore, in this thesis, the effects of *swap70a* knockdown on embryogenesis are presented. Antisense morpholino (MO) was employed to knockdown *swap70a in vivo*. The splice MO1 (+Intron1 MO) targeting to the junction between exon 1 and intron 1 was designed by Gene Tools. As a result, the first intron which contains an in-frame stop codon was included into the mature mRNA after splicing and thus a truncated protein was produced. Splice MO2 (ΔExon6 MO) was designed to bind to exon 6 and intron 6 boundary resulting in deletion of exon 6 in mature mRNA. Absence of exon 6 sequence in mRNA causes a frame-shift and thus a putative truncated protein might be produced as well. In order to determine the specificities of MOs used, co-injection of splice MO1 and splice MO2 together at a lower amount and co-injection of splice MO1 and *swap70a* mRNA (also called as rescue experiment) were carried out. In addition, the efficiencies of two splice MOs were determined by RT-PCR, followed by sequencing. Injection of splice MO1 was further used to determine the phenotypes of embryos during development. As splice MOs target to zygotic transcripts only, AUG MO targeting both maternal and zygotic transcripts was also analysed.

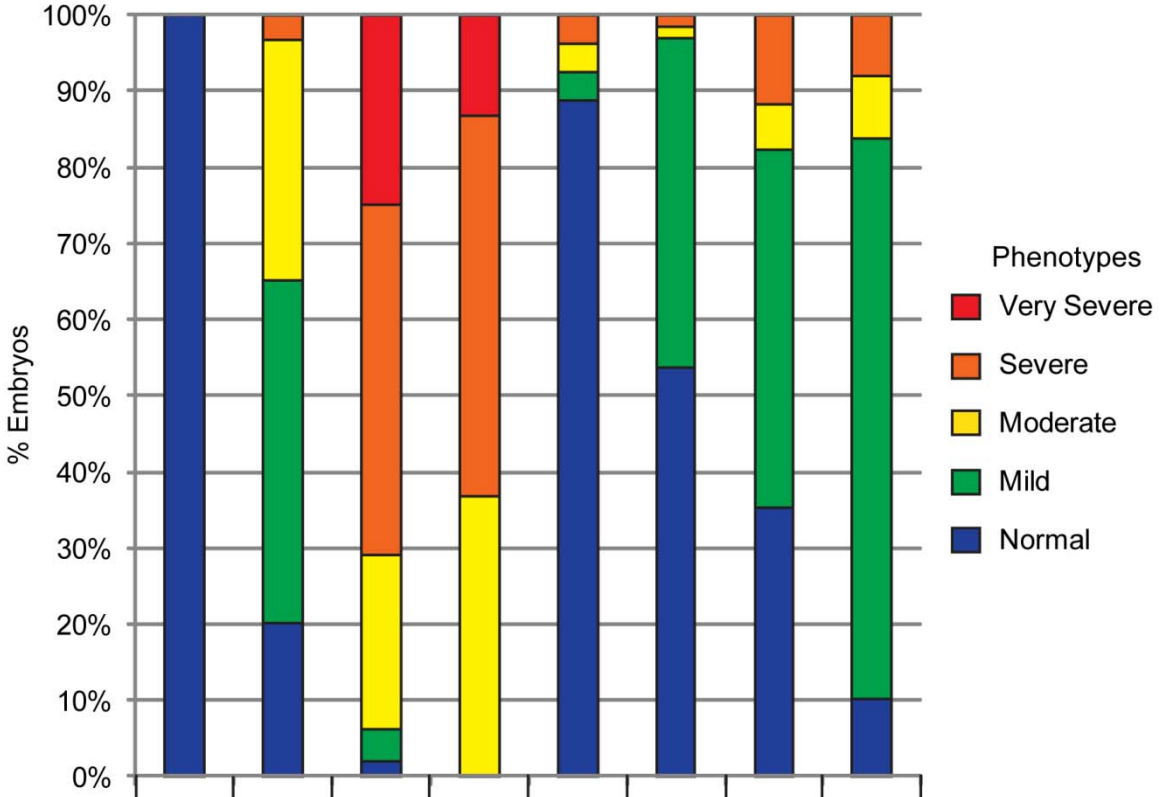
6.3.1 Knockdown of *swap70a* results in multiple defects at 24 hpf

To determine the lowest amount that is sufficient to show specific phenotypes and to minimise off-target effect, splice MO1 was titrated with three different amounts, 2.5ng, 5.0ng, and 7.5ng per embryo; whereas splice MO2 was titrated with four different amounts, 2.5ng, 5.0ng, 7.5ng, and 10ng per embryo. MOs were injected into one of the cells in embryos at 2-4 cells stages. Un-injected wild type embryos served as control. Injected and un-injected embryos were dechorinated and the phenotypes were examined at 24 hours post fertilisation (hpf) stage using a stereomicroscope.

According to the injection statistics (Figure 6.4), embryos were more sensitive to splice MO1 than splice MO2. For splice MO2 injected embryos, the mortalities were 5%-15%. But, more than half embryos died after high dosage splice MO1 injected. In addition, the remaining embryos showed more serious malformation than those injected with splice MO2.

As shown in Figure 6.5, splice MO1 injected embryos at 24 hpf showed varied defects. The main phenotypes were found in brain, eyes, yolk extension, and tail formation. Also, the body axes of injected embryos were statistically shorter. As the amount of splice MO1 increased, the phenotypes became more severe. Surviving embryos were underdeveloped and highly curved tails or even no tails were formed. Occasionally, one or no otolith formation was found in severe and very severe embryos. At the highest dosage, 7.5ng of splice MO1 per embryo, all surviving embryos had not undergone proper development.

Relatively, splice MO2 induced milder phenotypes (Figure 6.6). Also, the body axes were not significantly affected. The truncated protein containing the N-terminal may contribute to the mild phenotypes. Splice MO2 mainly affected the formation of the yolk extension which was thinner and messy, brain and tail formation. Importantly, the severe phenotypes of highly curved tails are similar to those induced by splice MO1, suggesting that both splice MO possibly target to the same *swap70a* transcript.



	WT	2.5ng MO1	5.0ng MO1	7.5ng MO1	2.5ng MO2	5.0ng MO2	7.5ng MO2	10ng MO2
No. of Embryos	55	100	112	95	56	71	60	54
Survival %	100%	89%	44%	32%	95%	94%	85%	91%

Figure 6-4 Statistics of *swap70a* Splice MO1 and Splice MO2 Injections. 2.5ng, 5.0ng, and 7.5ng of splice MO1 were injected into embryos; whereas 2.5ng, 5.0ng, 7.5ng, and 10ng of splice MO2 were injected. The number of surviving embryos was measured and the phenotypes of embryos were examined at 24 hpf stage. Based on the seriousness of the phenotypes, embryos were grouped into normal (blue), mild (green), moderate (yellow), severe (orange), and very severe (red). Images of embryos injected with splice MO1 are shown in Figure 6.5 and those injected with splice MO2 are shown in Figure 6.6.

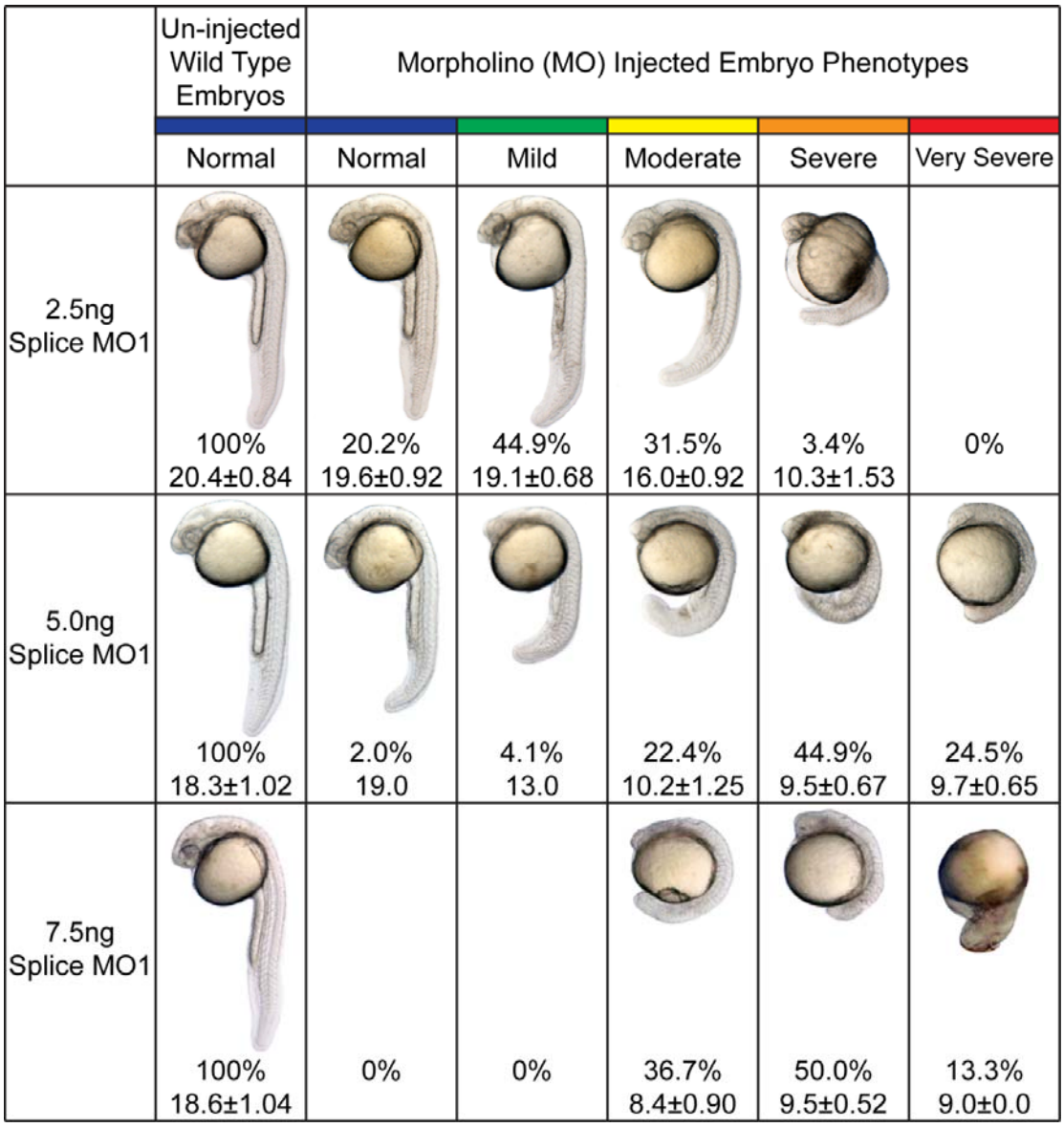


Figure 6-5 Dose-dependent Phenotypes of Splice MO1-mediated *swap70a* Knockdown. 2.5ng, 5.0ng, and 7.5ng of splice MO1 were injected into embryos separately in three experiments. Injected embryos at 24hpf were grouped into normal (blue), mild (green), moderate (yellow), severe (orange), and very severe (red) according to the phenotypes. Un-injected wild type embryos served as the experimental control. The percentages indicate the number of embryos (see figure 6.4) found in each phenotype group. The axes of embryos were measured from head to tail under microscopy with 10X magnification and were presented as mean ± standard derivation with unit, mm. Mean was presented in case of less than 3 embryos found in the phenotype group. The enlarged images of phenotypes are shown in Appendices IV to VI.

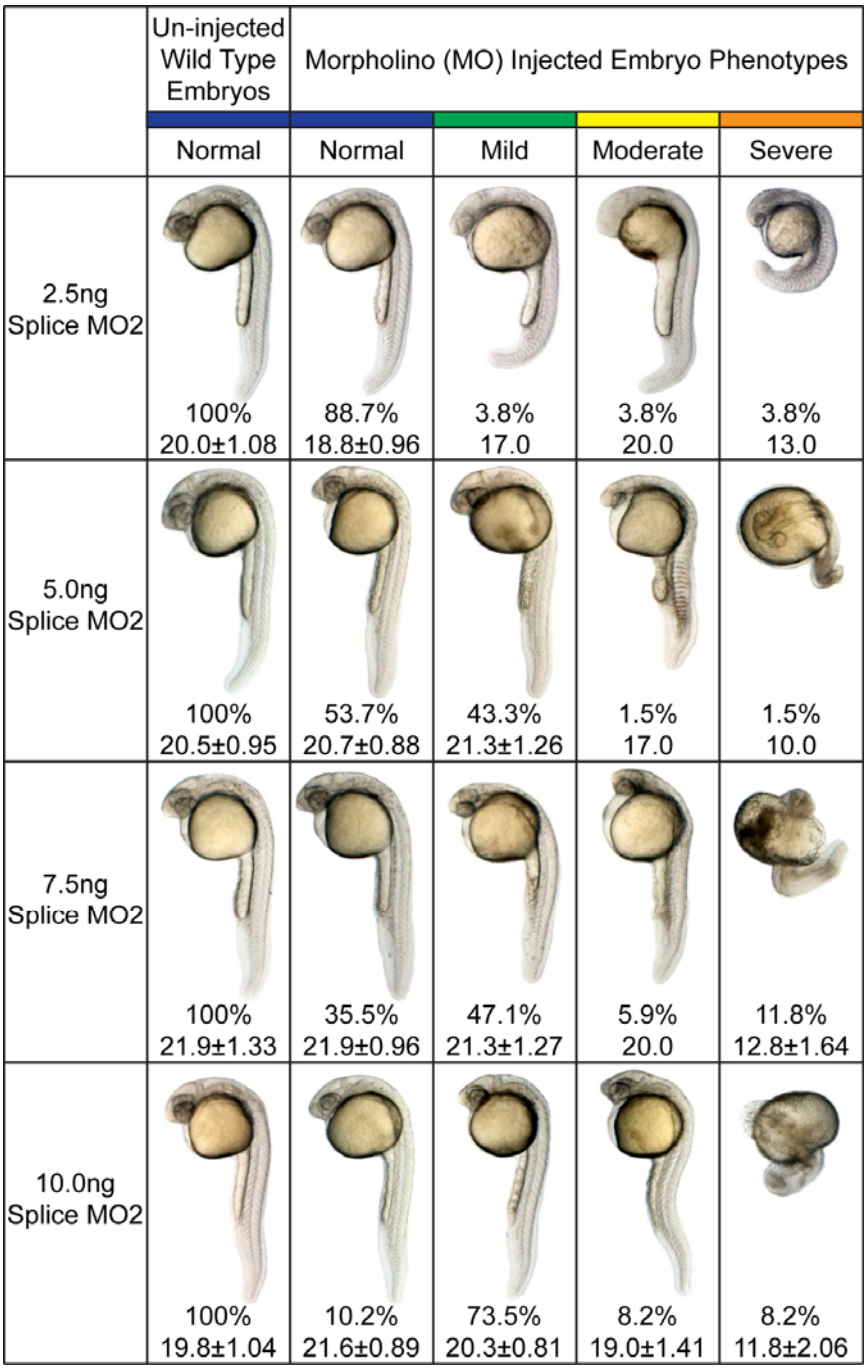


Figure 6-6 Dose-dependent Phenotypes of Splice MO2-mediated *swap70a* Knockdown. 2.5ng, 5.0ng, 7.5ng, and 10ng of splice MO2 were injected into embryos separately in four experiments. Injected embryos at 24hpf were grouped into normal (blue), mild (green), moderate (yellow), severe (orange), and very severe (red) according to the phenotypes. Un-injected wild type embryos served as the experimental control. The percentages indicate the number of embryos (see Figure 6.4) found in each phenotype group. The axes of embryos were measured from head to tail under microscopy with 10X magnification and were presented as mean ± standard derivation with unit, mm. Mean was presented in case of less than 3 embryos found in the phenotype group.

6.3.2 Splice morpholinos efficiently affect *swap70a* pre-mRNA splicing

RT-PCR was performed to examine the efficiencies of splice MO1 and MO2 using embryos co-injected with splice MO 1 and MO2 (1.5ng each) (see Figure 6.7). Splice MO1 was designed to bind to exon1-intron1 junction resulting in inclusion of intron1 mRNA after splicing. Primers e1 and i1 were used to amplify mRNA with altered splicing, whereas primers e1 and e2 were used to amplify both mRNA with normal splicing and altered splicing. However, only wild type transcripts were amplified and thus only e1e2 amplification was shown in Figure 6.6. Splice MO2 was designed to induce deletion of exon 6 from transcripts. Primer e5 and e8 were used to amplify both wild type and altered mRNA, whereas primer e6 and e8 were used to amplify wild type transcripts only. The details of primer sets used were shown in Table 2.3. The amplicons were sequenced to confirm the altered pre-mRNA splicing (Figure 6.8).

As shown in Figure 6.7, splice MO1 efficiently induced the increase of e1i1 amplification and decrease of e1e2 amplification. The amplicons of e1i1 and e1e2 were confirmed by sequence analysis (Figure 6.8A and 6.8B). Similarly, splice MO2 efficiently caused deletion of exon 6 (Figure 6.7). There are two bands detected in e5e8 amplification, the upper band is from the wild type transcript and the lower band is from altered mRNA. Both bands were gel extracted and purified for sequence analysis. According to Figure 6.8C and 6.8D, exon 6 was completely deleted from the lower band. In case of amplification using e6 and e8 primers, reduced amount of wile type transcript with presence of exon 6 was detected in the MO injected embryos (Figure 6.7 and 6.8E).

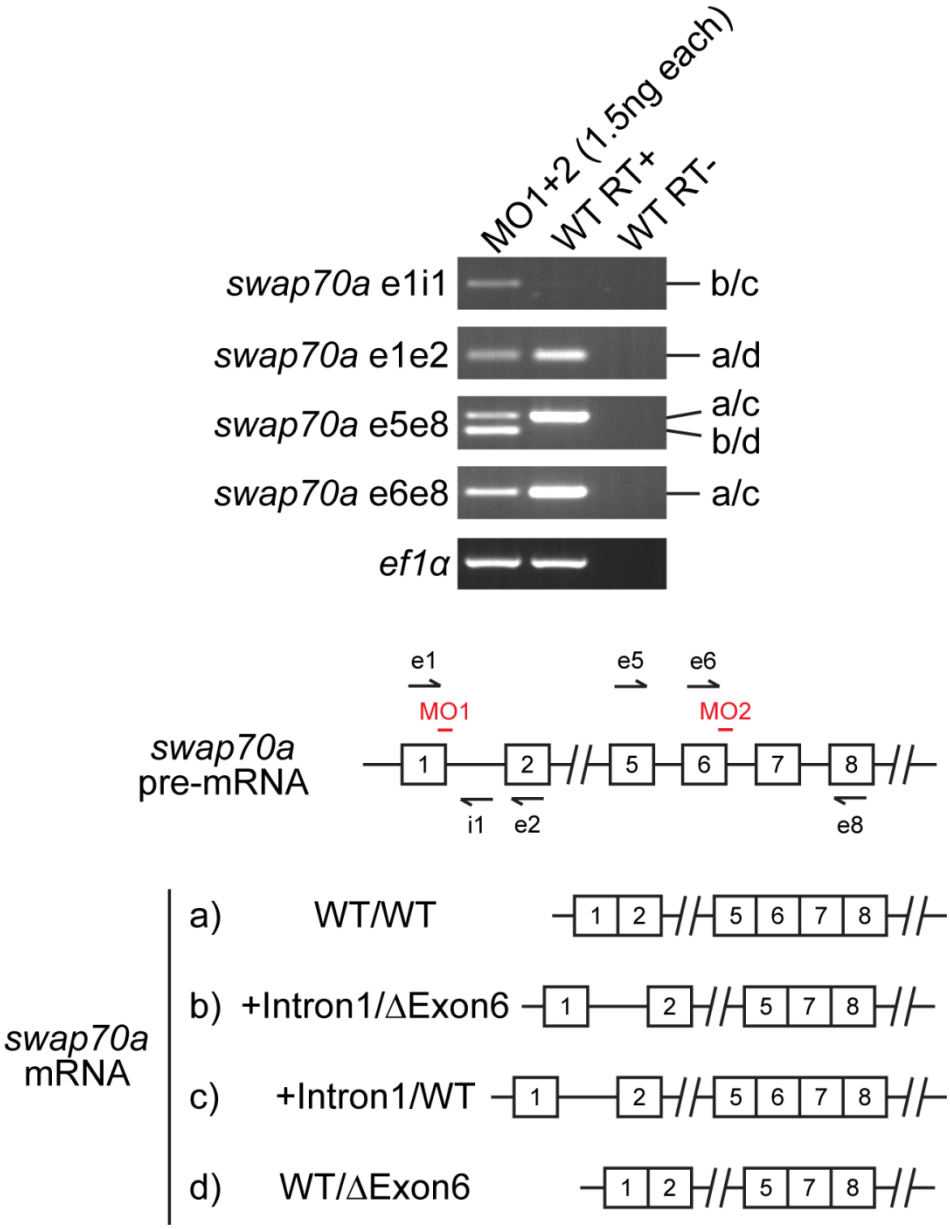


Figure 6-7 Efficiencies of Splice Morpholinos. (A) RT-PCR analysis. The analysis was performed using embryos co-injected with 1.5ng splice MO1 and 1.5ng splice MO2. PCR reaction with 35 cycles using wild type embryos with reverse transcription (RT+) and without reverse transcription (RT-) served as positive and negative control, respectively. The *ef1α* amplification with 30 cycles and under less UV exposure served as internal control. (B) Four possible outcomes of MOs co-injection.

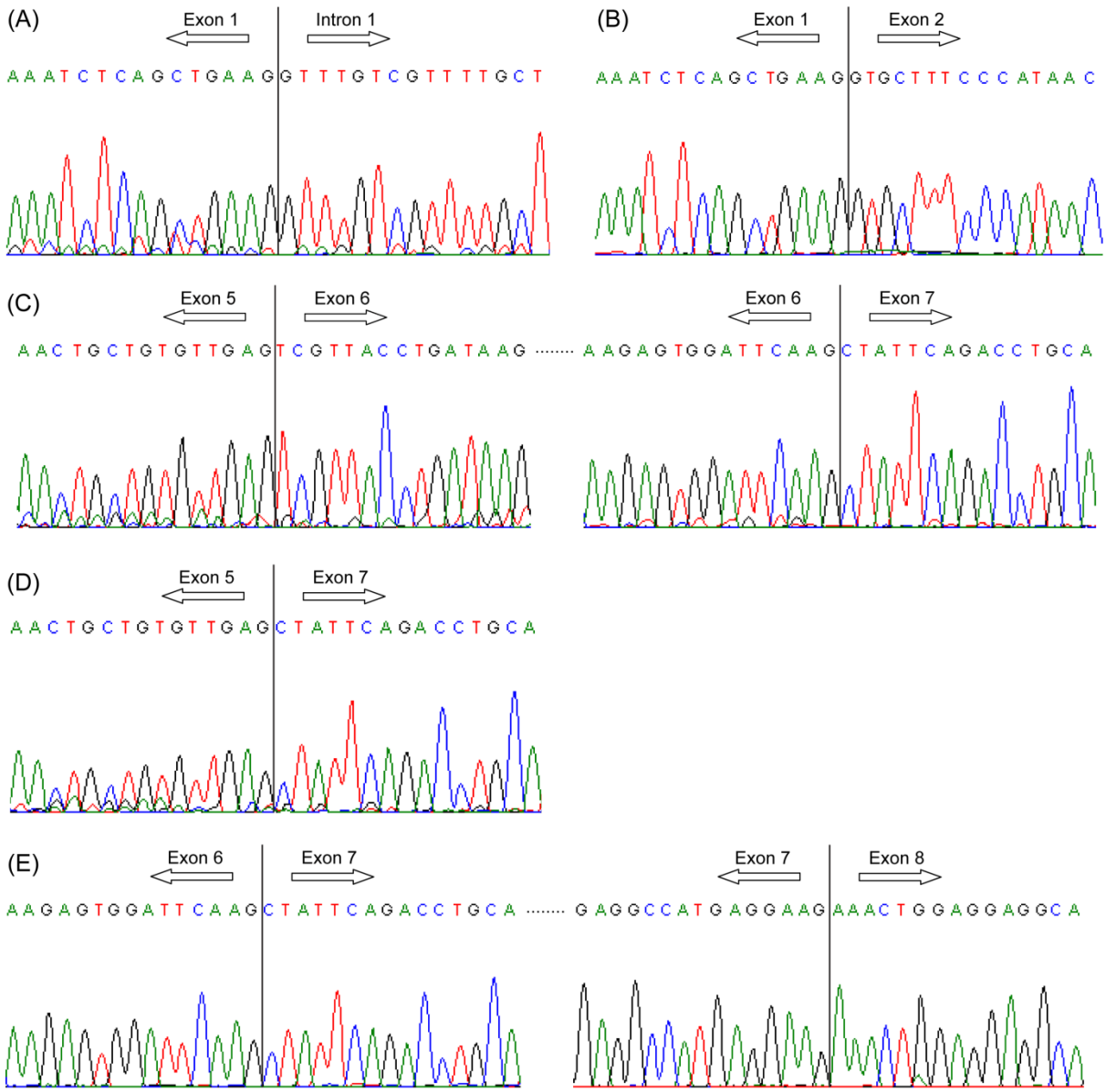
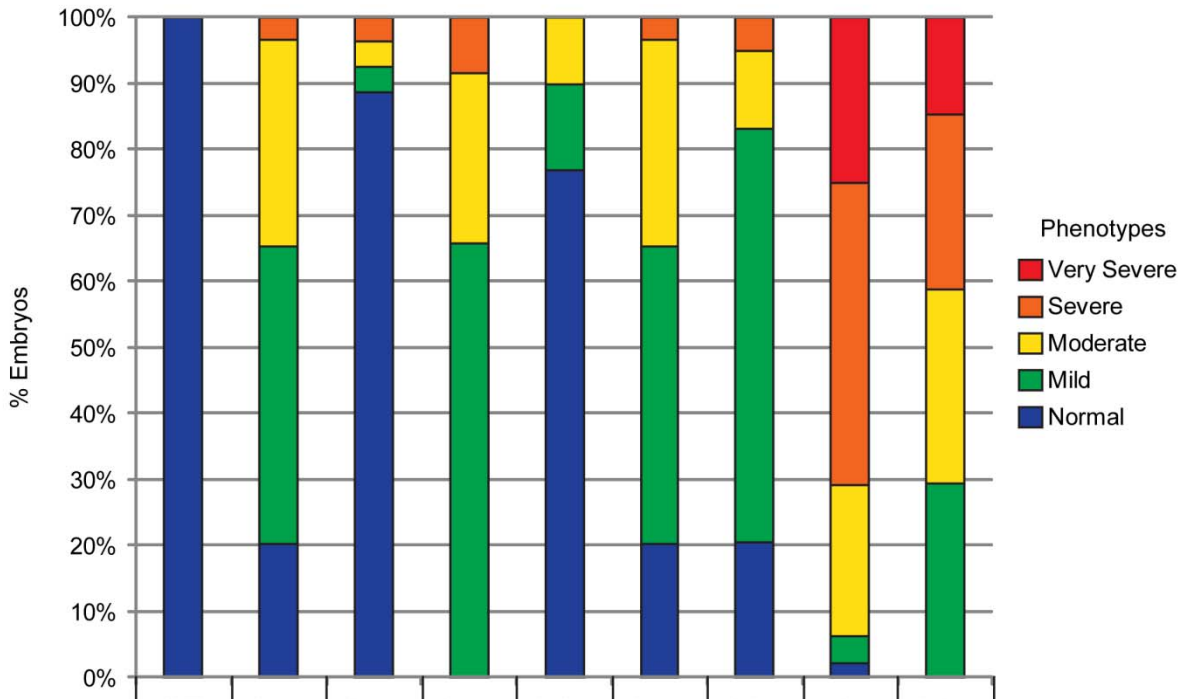


Figure 6-8 Sequence Analyses. The amplicons shown in Figure 6.5A were sequenced. (A) Exon1-intron1 boundary. Intron1 was included into mature mRNA by splice MO1-mediated altered splicing. (B) Exon1-exon2 boundary. Wild type mRNA was exclused intron 1 by normal splicing. (C) Exon5-exon6 and exon6-exon7 boundaries. Wild type mRNA was included exon 6 by normal splicing. (D) Exon5-exon7 boundary. Exon 6 was deleted in mRNA by splice MO2-mediated altered splicing. (E) Exon6-exon7 and exon7-exon8 boundaries. Exon 6 was included into mature mRNA by normal splicing.

6.3.3 Splice morpholino-induced defects which are specific to knockdown of *swap70a*

The specificities of splice MO1 and MO2 were determined using two strategies, co-injection of two non-overlapping MOs and co-injection of splice MO1 and *swap70a* mRNA. For the first strategy, as MO1 and MO2 target to different region of the same *swap70a* transcript, synergistic phenotypes in lower amount MO injected embryos were predicted to be observed. Embryos were co-injected with splice MO1 and MO2 (1.5ng each). For the rescue experiment, embryos were co-injected with 250pg of *in vitro* transcribed GFP-tagged *swap70a* mRNA, that on its own only caused a mild (13%) and moderate (10%) phenotypes (see figure 6.2 and 6.3), and either 2.5ng or 5.0ng of splice MO1.

According to Figure 6.9, co-injection of two splice MO clearly revealed the synergistic effects. Although both MOs were injected at lower dosage, 1.5ng each, the phenotypic bar chart is highly similar to 2.5ng splice MO1 injection alone with higher survival rate. Also, the phenotypes with shorter body axes (Figure 6.10) were more or less the same as injection of splice MO1. These results suggest that both splice MO1 and MO2 are specific to *swap70a* transcripts. Furthermore, splice MO1-induced defects were successfully rescued by *in vitro* transcribed GFP-tagged mRNA (Figure 6.9). For the first rescue experiment (co-injection of 250pg mRNA and 2.5ng splice MO1), moderate malformation phenotypes were reduced while more mild malformation phenotypes were achieved. In comparison, co-injection of 250pg mRNA and 5.0ng splice MO1 rescued not only severe and very severe phenotypes, but also the mortality (from 44% to 61%; see Figure 6.9). The partial rescues observed strongly suggest that the observed phenotypes are indeed due to the specific knockdown of *swap70a*.



	WT	2.5ng MO1 *	2.5ng MO2 *	1.5ng MO1 + 1.5ng MO2	250pg mRNA *	2.5ng MO1 *	250pg mRNA + 2.5ng MO1	5.0ng MO1 *	250pg mRNA + 5.0ng MO1
No. of Embryos	54	100	56	37	40	100	71	112	56
Survival %	100%	89%	95%	95%	98%	89%	83%	44%	61%

Figure 6-9 Specificities of Splice Morpholinos. Co-injection of 1.5ng splice MO1 and 1.5ng splice MO2, co-injection of 250pg mRNA and 2.5ng splice MO1, and co-injection of 250pg mRNA and 5.0ng splice MO1 were performed. The number of surviving embryos was measured and the phenotypes of alive embryos were examined at 24 hpf stage. Based on the seriousness of the phenotypes, embryos were grouped into normal (blue), mild (green), moderate (yellow), severe (orange), and very severe (red). Images of embryos co-injected are shown in Figure 6.10. * indicates the identical data sets presented earlier. Data sets of 2.5ng MO1, 5.0ng MO1, and 2.5ng MO2 were taken from Figure 6.4 and data set of 250pg mRNA was taken from Figure 6.2.

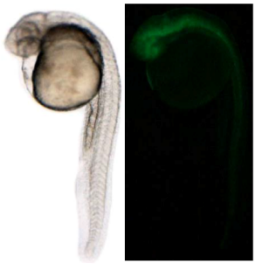
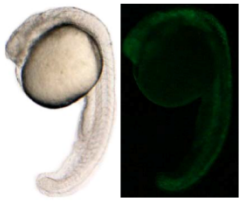
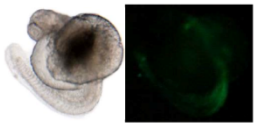
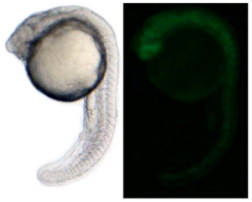
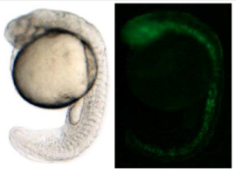
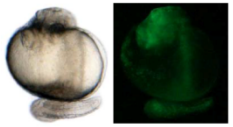
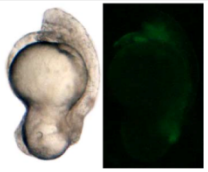
	Un-injected Wild Type Embryos	Co-injected Embryo Phenotypes				
	Normal	Normal	Mild	Moderate	Severe	Very Severe
MO1+MO2 1.5ng each	 100% 19.2±0.73	0%	 65.7% 17.6±1.23	 25.7% 14.9±1.64	 8.6% 10.0±2.00	0%
250pg mRNA + 2.5ng MO1	 100% 22.1±1.28	 20.0% 19.9±0.79	 62.7% 19.3±1.38	 11.9% 16.1±1.35	 5.1% 10.0±0.00	0%
250pg mRNA + 5.0ng MO1	 100% 18.2±0.77	0%	 29.4% 13.4±0.97	 29.4% 10.8±1.16	 26.5% 10.3±1.00	 14.7% 9.2±1.79

Figure 6-10 Synergy between Splice MO1 and MO2 as well as Partial Rescue of MO-mediated Phenotypes through Co-injection with *swap70a* RNA. Co-injection of 1.5ng splice MO1 and 1.5ng splice MO2, co-injection of 250pg mRNA and 2.5ng splice MO1, and co-injection of 250pg mRNA and 5.0ng splice MO1 were performed separately in three experiments. Injected embryos at 24hpf were grouped into normal (blue), mild (green), moderate (yellow), severe (orange), and very severe (red) according to the phenotypes. Success of GFP-tagged mRNA injection was confirmed by viewing injected embryos under fluorescence microscopy. Un-injected wild type embryos served as the experimental control. The percentages indicate the number of embryos (see Figure 6.9) found in each phenotype group. The axes of embryos were measured from head to tail under microscopy with 10X magnification and were presented as mean ± standard derivation with unit, mm.

6.3.4 Knockdown of *swap70a* results in gastrulation defects, delayed development and shortened body axis

According to previous analyses, splice MO1 showed high specificity and efficiency using 2.5ng. Thus, 2.5ng of splice MO1 was further used to determine the functions of *swap70a* during development. Un-injected wild type and Injected embryos were examined at different developmental stages (3, 6, 10, 13, 18, 24, 48, and 72 hpf). Major phenotypes with survival rates at different stages are shown in Figure 6.11.

At the earliest stage tested (3 hpf), no obvious phenotype was observed comparing MO injected and wild type embryos. At 6 hpf, slight delayed development which was possibly caused by the injection was observed. However, at 10hpf, MO-injected embryos exhibited incomplete gastrulation and higher mortality compared to wild type embryos that had reached bud tail stage. At 13 and 18 hpf, MO-injected showed a clear delay in development while wild type embryos had 9 somites at 13 hpf and 16 somites at 18hpf. MO-injected embryos, however, had only 6 somites at 13 hpf and 12 somites at 18hpf. And at 24 hpf, MO-injected embryos showed curved tails formed with obvious short body axes. At later stages (48 hpf), only one third of injected embryos survived which showed curved backs and shorter body axes. Furthermore, 7% of splice MO1 injected embryos had no red blood cell formation at 72 hpf.

6.3.5 Preliminary analysis of AUG morpholino injection reveals underdeveloped embryos and high lethality

As shown in Figure 6.1, *swap70a* was maternally expressed. Therefore, in order to determine whether maternal *swap70a* is essential to early development, AUG MO was designed to bind to the translational start sequence of the mRNA transcript, resulting in the block of translation and thus lack of protein production. 20ng of *swap70a* AUG MO was injected into each embryo.

As shown in Figure 6.11, there was no obvious effect observed at 3 hpf and slight delay in development at 6 hpf as for splice MO1 injected embryos. However, at bud stage (10 hpf), AUG MO injected embryos exhibited more severe phenotypes in gastrulation. The shape was like 70%-epilody stage with oblong embryos (Kimmel *et al.*, 1995). From 10 hpf to 13 hpf, half of the embryos died and the remaining

embryos had no eye, somite, tail formed. Furthermore, at 18 hpf, only one embryo was alive but it died later after 24 hpf.

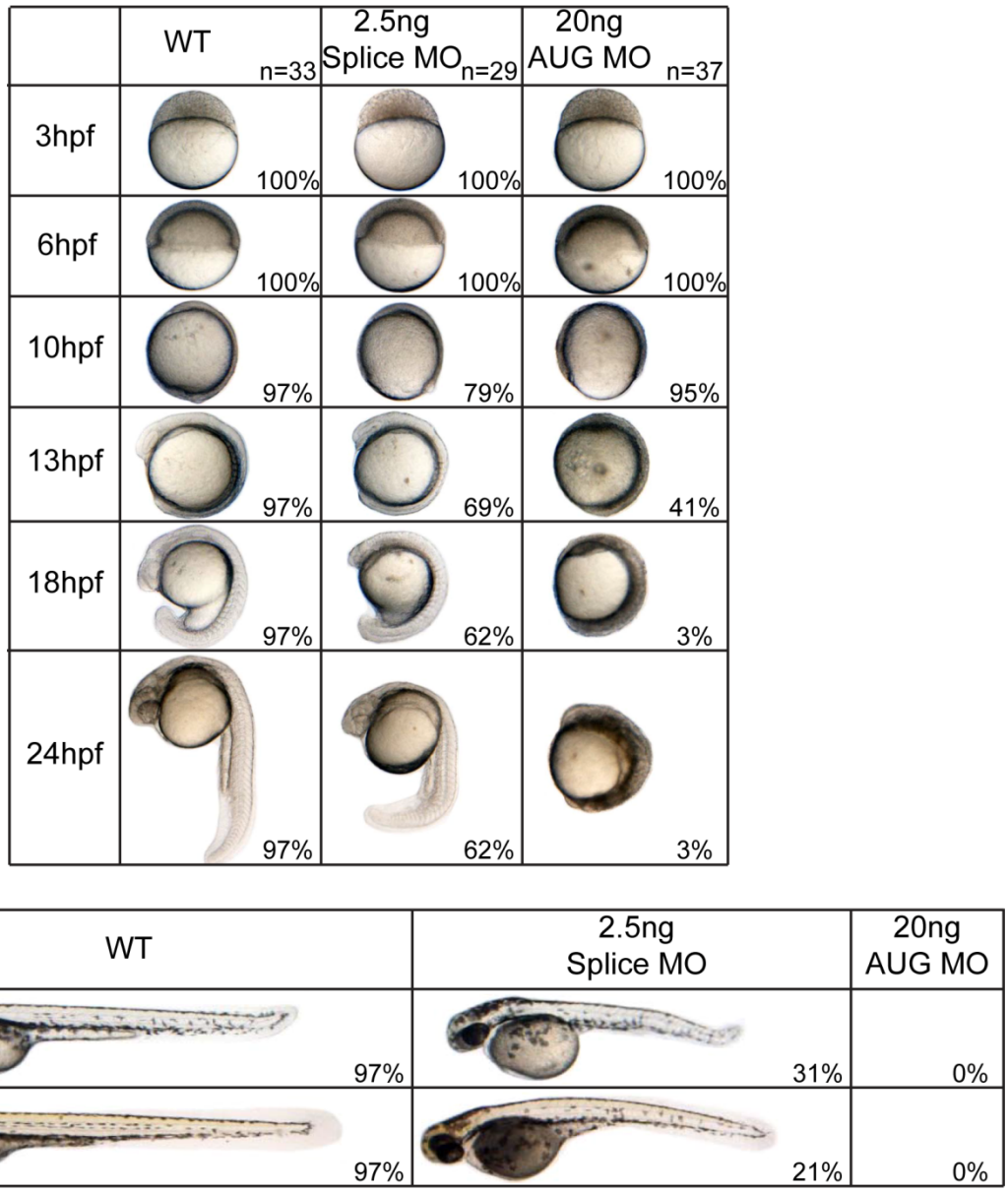


Figure 6-11 Knockdown of *swap70a* during Zebrafish Development. 2.5ng of splice MO1 and 20ng of AUG MO were injected into embryos at the 2-4 cell stage, separately. Un-injected wild type embryos served as experimental control. The phenotypes shown are from the same experiment. At different developmental stages (3, 6, 10, 13, 18, 24, 48, and 72 hpf), the phenotypes of each group embryos were examined under microscopy and the major phenotypes are shown in this figure. The percentages indicate the survival rates of embryos at different developmental stages.

7. Discussion

7.1 Varied number of def6/swap70 genes was determined in vertebrates and invertebrates

In zebrafish (*Danio rerio*), there are five genes belonging to the *def6/swap70* gene family, *swap70a*, *swap70b*, *def6a*, *def6b*, and *def6-like*, with different expression levels during embryogenesis. As shown in Figure 4.7 and 4.8, long branch lengths of *def6-like* group indicated that they are fast-evolving gene compared to others. The gene *def6b* (previously called as *LOC570940*) was annotated based on two contigs (Zv8_scaffold 2548.7 and Zv8_scaffold 2748.8) in zebrafish Zv8 genome project. But there is a gap in between two contigs, resulting in incorrect sequence annotation, and during this study, *def6b* was discontinued in the NCBI database. However, according to the sequence analysis shown here, *def6b* sequence is 100% matched to contig FP067424.7 which is shown in the recently released zebrafish Zv9 genome project. According to Ensembl release 59 (August 2010), the genomes of tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*), and medaka (*Oryzias latipes*) were assembled in 2003, June 2005, and October 2005, respectively. Therefore, the fact that only 4 *def6* paralogues had been identified in the other four teleost species genomes might be due to incomplete genome sequencing and annotations. Alternatively, the other teleost species might exhibit further gene loss during evolution. According to Gillis *et al*. (2009), they found that zebrafish has *GATA1a* and *GATA1b* whereas other four teleosts used in the tree analysis here have *GATA1a* only. Also, zebrafish *GATA1b* has a long branch length indicating the more diverged sequence compared to *GATA1a*. The tree topology of vertebrate *GATA1* is highly similar to that of vertebrate DEF6. This suggests that it is common for zebrafish to have one paralogue gene more than others. Consequently, the presence or absence of *def6b* in other teleost species could be determined once the updated genomes have been released in the future.

In the genomes of human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*) and frog (*Xenopus tropicalis*), only two *def6/swap70* family members were identifies, i.e. *DEF6* and *SWAP70*. As shown in Table 7.1, although the genome annotations have not been completed yet, it clearly shows that 33 tetrapod species contain DEF6 and SWAP70. And there is no *def6-*

like, *swap70b*, and *def6b* identified. Taken together with information of only one gene identified in invertebrates, a molecular evolution model is proposed here to explain the varied number of genes found in different species.

Common Name	Species	*DEF6*	*SWAP70*
Cow	*Bos taurus*	++	++
Marmoset	*Callithrix jacchus*	++	++
Dog	*Canis familiaris*	+	++
Guinea Pig	*Cavia porcellus*	++	++
Sloth	*Choloepus hoffmanni*	+	+
Armadillo	*Dasypus novemcinctus*	-	+
Kangaroo rat	*Dipodomys ordii*	++	+
Lesser hedgehog tenrec	*Echinops telfairi*	+	+
Horse	*Equus caballus*	++	+
Hedgehog	*Erinaceus europaeus*	+	-
Cat	*Felis catus*	++	+
Gorilla	*Gorilla gorilla*	++	++
Elephant	*Loxodonta africana*	++	++
Macaque	*Macaca mulatta*	++	++
Mouse Lemur	*Microcebus murinus*	+	+
Wallaby	*Macropus eugenii*	+	+
Turkey	*Meleagris gallopavo*	++	++
Opossum	*Monodelphis domestica*	++	++
Microbat	*Myotis lucifugus*	+	+
Pika	*Ochotona princeps*	++	++
Platypus	*Omithorhynchus anatinus*	+	+
Rabbit	*Oryctolagus cuniculus*	++	++
Bushbaby	*Otolemur garnettii*	+	+
Chimpanzee	*Pan trolodytes*	++	++
Orangutan	*Pongo pygmaeus*	++	++
Hyrax	*Procavia capensis*	++	+
Megabat	*Pteropus vampyrus*	+	+
Rat	*Rattus norvegicus*	++	++
Shrew	*Sorex araneus*	+	+
Squirrel	*Spermophilus tridecemlineatus*	-	+
Pig	*Sus scrofa*	++	+
Zebra Finch	*Taenipygia guttata*	++	++
Tarsier	*Tarsius syrichta*	+	+
Tree Shrew	*Tupaia belangeri*	+	+
Dolphin	*Tursiops truncatus*	++	+
Alpaca	*Vicugna pacos*	+	+

The information is from Ensembl database, version release 59 (August 2010).
-: No gene identifed yet.
+: Gene identified with incomplete sequence.
++: Gene identified with complete sequence.

Table 7.1 List of Identification of DEF6 and SWAP70 in Tetrapod Species.

7.2 A Molecular evolution model

There are many phylogenetic analyses, comparative analyses and gene phylogenies supporting the presence of two rounds of whole genome duplications (WGDs) in the ancestral vertebrate. Two rounds of WGD contributed to the genomic complexities among vertebrate species. Also there is an actinopterygians lineage specific third WGD commonly proposed. Actinopterygians which is the ancestral teleost evolved into fish species including zebrafish, fugu, and medaka (reviewed in Sato & Nishida, 2010). In contrast, there is gene duplication proposed for invertebrates rather than WGDs, an example is the duplication of *Pax6* in arthropod species (reviewed in Callaerts *et al*, 2006). Therefore, in case of *def6/swap70* gene family, as there is only one gene identified in invertebrate, we hypothesised that there was one *def6/swap70* ancestral gene present before speciation of vertebrate and invertebrate. In the vertebrate lineage, there were two rounds of WGD and thus *def6/swap70* was duplicated into 4 genes. Given that there was a teleost-specific WGD, we have to predict that there were 8 *def6* paralogues formed. Alternatively, if one copy was lost after the 2nd WGD, the 3rd WGD in teleost would have resulted in 6 *def6* paralogues (as shown in Figure 7.1). Over time, one gene was lost in zebrafish and two genes were lost or one was lost and one has not yet identified in other teleost fish species. As tetrapods and teleosts shared a common ancestral vertebrate before speciation, tetrapods also had 3 genes in their genomes. One was lost later, resulting in only 2 genes, DEF6 and SWAP70, identified in tetrapods in present study. The molecular evolution hypothesis is summarised in Figure 7.1.

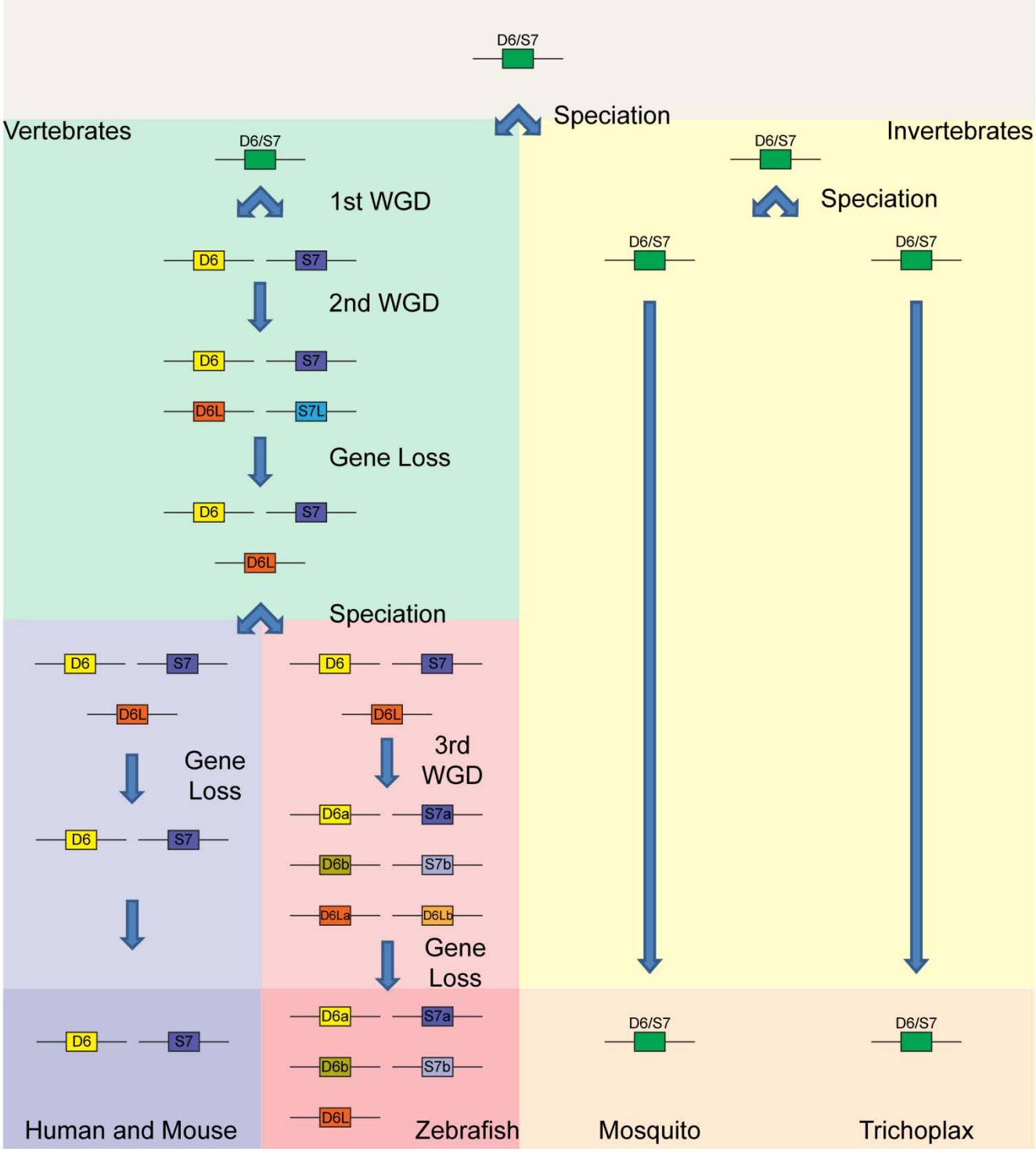


Figure 7-1 Hypothesis Model for Molecular Evolution of *def6/swap70* Gene Family. D6/S7 indicates the gene *def6/swap70*. D6, S7, D6L, and S7L represent *def6*, *swap70*, *def6-like*, and *swap70-like*, respectively. D6a, D6b, S7a, S7b represent def6a, def6b, swap70a, and swap70b, respectively, in teleost species. D6La and D6Lb represent *def6-likea* and *def6-likeb*. Three rounds of whole genome duplications (WGDs) and hypothesised gene loss are shown in the model.

7.3 The Molecular evolution model is supported through synteny analysis

According to the molecular evolution hypothesis (Figure 7.1), *DEF6* and *SWAP70* were predicted to derive from a simple ancestral *def6/swap70* during 1st WGD and *def6-like* was predicted to be duplicated from *DEF6* during 2nd WGD. Further, *def6a* and *def6b* as well as *swap70a* and *swap70b* were duplicated from *DEF6* and *SWAP70* during the 3rd WGD. The gene duplications as a result of WGD were confirmed by synteny analyses, such as *Cyp19* duplication (Chiang *et al*., 2001), neurexin gene duplication (Rissone *et al*., 2006), *Period* gene family (Wang, 2008), and basic helix-loop-helix transcription factors family (Wang *et al*., 2009). Also, synteny analysis is importantly involved in comparative genomics to determine the vertebrate genome evolution (Postlethwait *et* al., 1998 and Postlethwait *et al*., 2000). Consequently, synteny analysis was performed to further support the molecular evolution hypothesis. As shown in Figure 7.2B, the gene, *def6a* showed no syntenic relationship to any chromosome. Nevertheless, the 1st WGD resulting in the formation of *DEF6* and *SWAP70* (Figure 7.2A), the 2nd WGD resulting in the formation of *def6-like*, and the teleost specific WGD resulting in the formation of *def6b* (Figure 7.2B) as well as *swap70a* and *swap70b* (Figure 7.2C) were confirmed by the synteny analyses. As a result, 3 rounds of WGD contributed to the formation of *def6/swap70* gene family in zebrafish.

According to Hufton *et* al. (2008), speciation of zebrafish after 3rd WGD had the highest rate of synteny loss among vertebrate species. It could be the possible reason that *def6a* does not show syntenic relationship. Nevertheless, the observed synteny shown in Figure 7.2 strongly supports the molecular evolution model presented here.

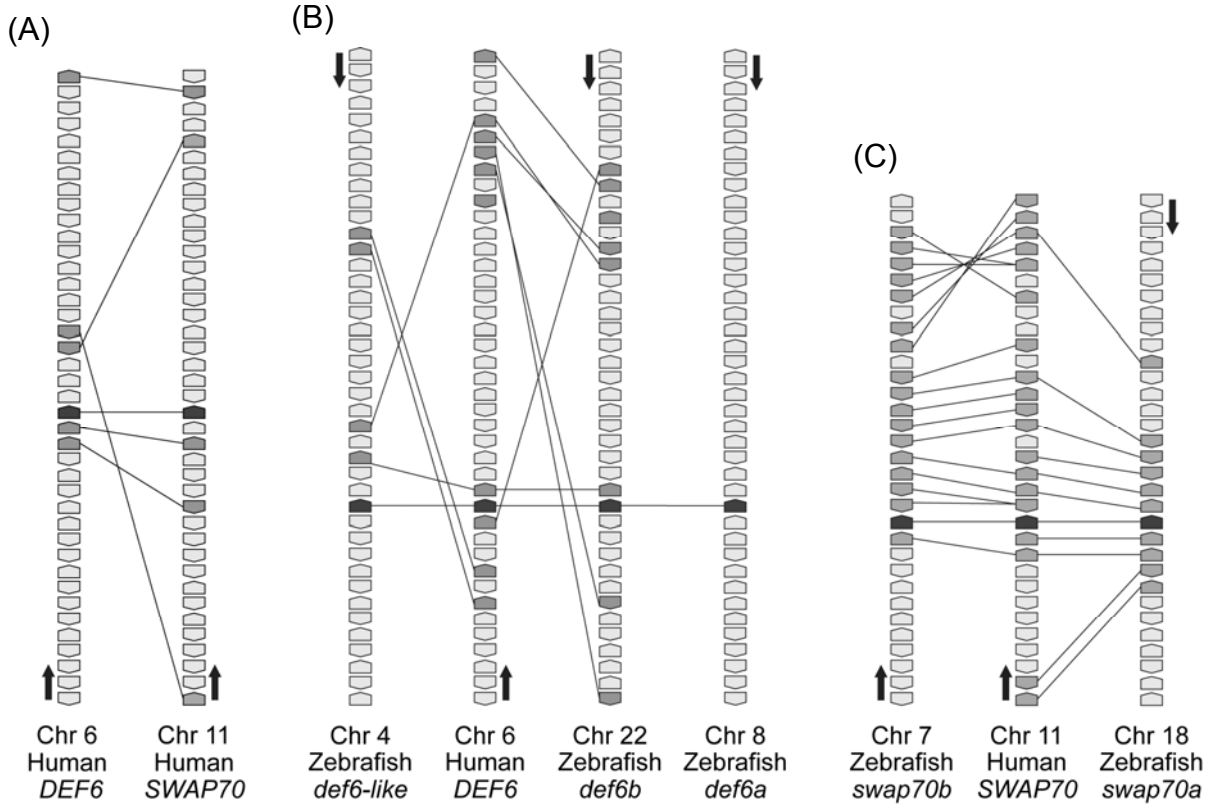


Figure 7-2 Synteny Analysis. Information was acquired from Genomicus version 58.01 which is integrated with Ensembl release 58. Dark boxes indicate the genes in *def6*/*swap70* family. Grey boxes with black lines indicate the different homology genes found in different species chromosomes. Light grey boxes indicate the genes which have no homology relationship to other species chromosomes. The orientations of the boxes indicate the directions of the open reading frames. Black arrows indicate the direction of chromosome towards centromere for human and chromosome central for zebrafish. (A) Syntenic relationship between human *DEF6* and *SWAP70*. (B) Syntenic relationship among human *DEF6* and zebrafish *def6-like*, *def6b*, and *def6a*. (C) Syntenic relationship among human *SWAP70* and zebrafish *swap70a* and *swap70b*.

7.4 Lamprey and hagfish genome sequencing and annotations could be used to test the molecular evolution model

As Escriva *et al*. mentioned in 2002, the gene duplication during early vertebrate evolution can be determined through analysis of lamprey and hagfish genomes. Gillis *et al*. (2009) summarised the recent knowledge of species evolution (see Figure 7.3). Both Escriva *et al*. (2002) and Gillis *et al*. (2009) indicated that the divergence timing for lamprey and hagfish speciation is still uncertain. But it is sure that speciation of lamprey and hagfish was after the 1st WGD. As a result, in the case of *def6*/*swap70* gene family, lamprey and hagfish might have two *def6*-related paralogues. One is closer to *def6* and one is closer to *swap70*. If speciation of lamprey and hagfish had undergone after 2nd WGD, there might be four *def6*-related genes identified. The molecular evolution hypothesis could be further supported once the updated lamprey and hagfish genome information have been released.

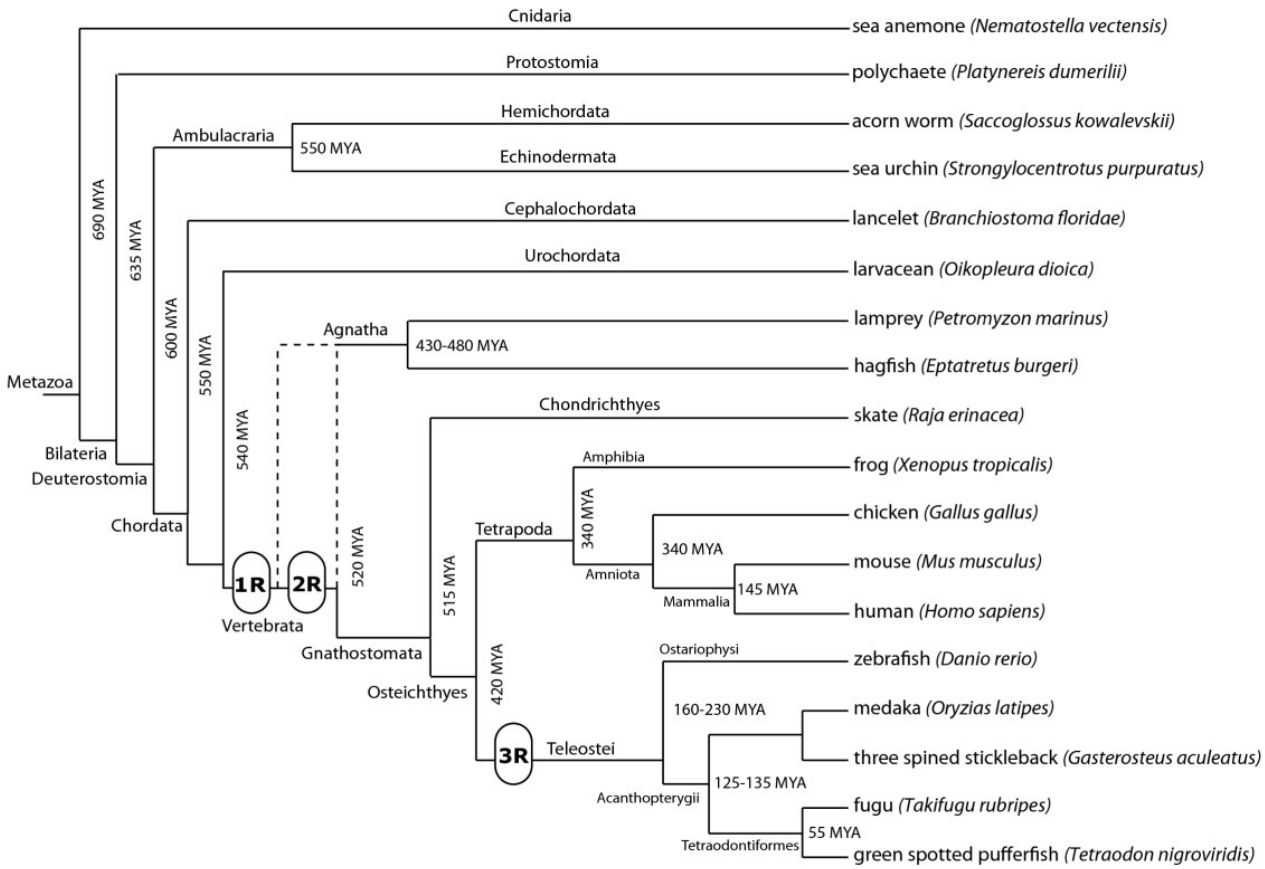


Figure 7-3 Relationship and Divergence Times of Vertebrates and Invertebrates. Divergence times were estimated in millions of years ago (MYA). The first round (1R) of whole genome duplication (WGD), the second round (2R) of WGD, and the teleost-specific third round (3R) WGD are represented by rounded rectangles. The dotted line represents the uncertainty of the divergence time for 2R of WGD. Adapted from Gillis *et al*., 2009.

7.5 Present *SWAP70* is much closer to the ancestral *def6/swap70* gene than *DEF6* to *def6/swap70*

The exon-intron structures of *def6-like* (Figure 4.4) are more similar to *swap70* orthologues (Figure 4.1 and 4.2). This phenomenon reveals that before 2nd WGD and duplication of *def6-like* and *def6* occurred, the ancestral *def6* gene possibly had 12 exons in total. Along the evolution, exon 7 and exon 8 in *def6* were fused together and thus the present def6 contains only 11 exons (Figure 4.3). Moreover, the multiple amino acid alignments (Figure 5.2 and 5.3) as well as pairwise analysis shown in Figure 5.4 indicated that statistically close relationship of amino acids of def6/swap70 and swap70 were determined. In other words, *def6* orthologues were faster evolving than *swap70* orthologues.

7.6 PH-DHL domain arrangement which is the key signature of *def6/swap70* family is a conserved ancestral structure in metazoa evolution

Rossman *et al*. (2005) reviewed that there are 69 unique Dbl proteins including itam1 and vav1, identified in human containing DH-PH domain arrangement. It could be hypothesised that the atypical PH-DHL domain arrangement in def6/swap70 group was evolved from one of the ancestral Dbl proteins during evolution. However, this hypothesis is absolutely contradicted with the analyses presented in this thesis. PH-DHL domain arrangement is highly supported to exist in *def6*/*swap70* family as early as the simplest animal, Tricoplax. Therefore, the special PH-DHL domain is the key signature of *def6*/*swap70* genes. In addition, it is common to select a specific conserved region for tree construction. For example, tree analysis of type XXVII collagen genes containing multiple domains was constructed based on a single conserved C-propeptide domain only (Christiansen *et al*., 2009). But, in case of def6/swap70 family, conserved region alone or in combinations failed to produce good neighbour-joining and maximum likelihood trees with high bootstrap values supporting in most of the tree branches separation. Therefore, all the tree analyses shown in this thesis are based on full length amino acid sequences, revealing the high conservation of full length amino acids throughout the evolution. And thus the formation of multiple domain structures of *def6* and *swap70* are not the result of domain fusions from different sequences. It can be concluded that *def6*/*swap70* family is a novel GEF family with PH-DHL domain arrangement different to Dbl family.

7.7 Extremely conserved amino acids in *def6/swap70* family reveal the importance of amino acid sequences towards structure and function

As shown in Figure 4.5 and 4.6, SWAP70 orthologues and DEF6 orthologues show high conservation in amino acid sequences, suggesting the structures and folding of SWAP70 and DEF6 are highly similar among different vertebrate species. The I-TASSER 3D prediction shown in Figure 4.10 also supported this hypothesis. The conserved folding possibly contributes to highly similar functions in different species. Remarkably, the N-terminal end from alignment position 1 to 103 and C-terminal end from 530 to 559 are extremely conserved in DEF6 orthologues (Figure 4.6). These two regions possibly contribute to the actual contact points resulting in formation of highly conserved folding.

7.8 *swap70a* may have a more important role than other *def6/swap70* members in early embryogenesis

As Kane & Kimmel (1993) indicated that zebrafish midblastula transition (MBT) begins at cycle 10 just before 3 hpf. They injected labelled UTP to 8-cell stage embryo to determine the accumulation of newly synthesised RNA from 2 hpf to 7 hpf. A measurable increase of labelled RNA was firstly detected at 3 hpf. It means that the zygotic transcription was activated and new transcripts started to be produced. Therefore, detections of *swap70a* and *def6a* mRNA at 1hpf indicated that they were maternally expressed. However, at 1 hpf, *swap70a* expression is the highest compared to the other *def6* paralogue genes in zebrafish, suggesting that *swap70a* is the key protein involving in early embryogenesis (see Figure 6.1). As shown in Figure 6.11, the injection of 2.5ng *swap70a* splice MO1 shows a gastrulation defect and the preliminary analysis of 20ng *swap70a* AUG MO injection shows higher mortality and a more severe gastrulation defect at 10 hpf.

Nevertheless, there is no antibody available to detect maternal expressed proteins and to determine the efficiency of *swap70a* AUG MOs. The mortality cannot be sure whether it is caused by *swap70a* knockdown or off-taget effects. Thus, it is necessary to raise antibody and to quantify the efficiencies of AUG MO-mediated gene knockdown using western blotting. Also, as the mortality rate of *swap70a* AUG MO is very high, co-injection with p53 MO should be performed to inactivate the p53 apoptotic cell death pathway caused by MO off-target effect (Robu *et al*., 2007) or knockdown of *swap70a*. In addition, *in situ* hybridisations, MO-mediated gene

knockdown, and gene over-expression for other *def6* paralogues in zebrafish should be performed in the future.

7.9 *swap70a* may be required for convergence and extension movements during gastrulation

Both splice MO1-mediated and AUG MO-mediated *swap70a* knockdown shows obvious defects in gastrulation at bud stage, 10 hpf (Figure 6.11). AUG MO injected embryos showed high mortality rate as well. It may be caused by the severe defect during gastrulation and thus embryos cannot survive. It is well known that convergence and extension (CE) cell movements, which are regulated by Wnt/PCP pathway, are critical for gastrulation (reviewed in Roszko *et al*., 2009). It is tempting to speculate that *swap70s* may be involved in Wnt/PCP pathway which regulates CE cell movements.

7.10 Phenotypes of *swap70a* are similar to those of *def6a*, suggesting that they are involved in similar signal transduction pathways

Comparing to Martin, *PhD thesis* (2007) and Goudevenou, *PhD thesis* (2010), the *swap70a* MO-induced defects are highly similar to *def6a* MO-induced defects. At early developmental stages, both MO-injected embryos showed gastrulation defects. At later stages, disrupted eye, midbrain-hindbrain boundary and otic vesicle were identified as well as tail formations were also affected. But the *def6a* morphant resulted mainly in curved tails with broad, flattened and irregular shaped somites whereas *swap70a* morphants mainly resulted in curved tail formation only. Therefore, *swap70a* and *def6a* may be involved in similar pathways. Because *def6a* was proven to act downstream of the non-canonical Wnt5b signaling pathway and to function with Wnt11 in parallel or overlapping pathways (Goudevenou, *PhD thesis*, 2010), the roles of *swap70a* may be similar to *def6a* in non-canonical Wnt signaling. To further test this hypothesis, co-injection of *swap70a* and *def6a* MO, rescue of *swap70a* morphants with *def6a* mRNA and *vice versa* could be performed to determine whether they act in parallel or overlapping pathways. Similarly, co-injection of *Wnt5b/11* MO and *swap70a* MO, rescue of *Wnt5b/11* morphants with *swap70a* mRNA could be performed to determine whether *swap70a* is involved in *Wnt5b/11* signalling pathways. Also, some marker probes need to be used for *in situ*

hybridization to further dissect the signaling pathways and to determine whether *swap70a* affects the cell fates.

7.11 *swap70a* may be involved in Wnt11 signalling contributing to eye development and in RhoA activation for correct number of otoliths formation

The coordination of Wnt/PCP pathway through Wnt11, Fz5, and Wnt/β-catenin pathway through Wnt8b and Fz8a is required for eye development (Cavodeassi *et al*., 2005). Wnt8b and Fz8 inhibit the eye specification, whereas Wnt11 and Fz5 promote development of eye field. As *swap70a* knockdown showed undeveloped eyes and *swap70a* over-expression showed misshapen eyes, *swap70a* may act downstream of Wnt11 signalling pathway.

Panizzi *et al*. (2007) discovered a new role of Rho guanine exchange factor (GEF) which is development of ciliated epithelia. Importantly, they found that the activation of Rho GTPase via a guanine nucleotide exchange factor, Arhgef11, which contains typical DH-PH domain arrangement, is required for the formation of the normal number of ear otoliths. In comparison, *swap70a* knockdown and overexpression also induced abnormal number of ear otoliths formation. It is possible that swap70a directly or indirectly activates RhoA to regulate otolith formation like Arhgef11 but in different signalling pathways.

As eyes and ears are sensory organs, their developments require extensive development of nervous system. According to transcriptomic analysis for astrocytes, neurons, and oligodendrocytes development in mouse (Cahoy *et al*., 2008), the expression of SWAP70 increases when astrocytes and oligodendrocytes were activated. Takada & Appel (2010) also proved that *swap70a* is expressed in oligodendrocytes and not Schwann cells in zebrafish. Therefore, it is clear that *swap70a* is involved not only in signalling pathways for induction of eyes and ears development, but also in promoting nervous system development.

7.12 Conclusion

Within the *def6*/*swap70* gene family, invertebrate species have one gene, tetrapod species have two genes, teleosts species have four to five genes. The amino acid sequences among family members are extremely conserved, resulting in highly similar predicted structures. Bioinformatics analysis strongly suggests that the atypical PH-DH domain arrangement is the key signature for this GEF family.

The *swap70a* has an essential role in zebrafish embryogenesis. Gain-of function analysis of *swap70a* showed misshapen eyes, abnormal numbers of otolith formation, and tail formation defects. Loss-of function analysis of *swap70a* showed gastrulation defects, delay in development, underdeveloped brain, eyes, and ears, abnormal number of otolith formation, and tail formation defects. It is hypothesised that *swap70a* acts downstream of Wnt/PCP signalling pathway through Wnt11 and is involved in the non-canonical Wnt pathways parallel to *def6a*.

8. References

Abascal, F., Zardoya, R. and Posada D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 1:2104-2105.

Alberto Rissone, A., Monopoli, M., Beltrame, M., Bussolino, F., Cotelli, F. and Arese, M. (2007). Comparative Genome Analysis of the Neurexin Gene Family in *Danio rerio*: Insights into Their Functions and Evolution. *Molecular Biology Evolution. 24: 236-252.*

Bill, B.R., Petzold, A.M., Clark, K.J., Schimmenti, L.A. and Ekker, S.C. (2009). A Primer for Morpholino Use in Zebrafish. *Zebrafish*. 6:69-77.

Biswas, P.S., Bhagat, G. and Pernis, A.B. (2010). IRF4 and its regulators: evolving insights into the pathogenesis of inflammatory arthritis? *Immunological Reviews*. 233:79–96.

Borggrefe, T., Keshavarzi, S., Gross, B, Wabl, M and Jessberger, R. (2001). Impaired IgE response in SWAP-70-deficient mice. *European Journal of Immunology.* 31:2467–2475.

Borggrefe, T., Masat, L, Wabl, M, Riwar, B, Cattoretti, G. and Jessberger, R. (1999). Cellular, intracellular, and developmental expression patterns of murine SWAP-70. *European Journal of Immunology. 29:*1812–1822.

Borggrefe, T., Wabl, M., Akhmedov, A.T. and Jessberger, R. (1998). A B-cell-specific DNA Recombination Complex. *The Journal of Biological Chemistry*. 273:17025–17035.

Bornberg-Bauer, E., Rivals, E. and Vingron, M. (1998). Computational approaches to identify leucine zippers. *Nucleic Acids Research*. 26:2740-2746.

Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S. and Jones, D.T. (2005). Protein structure prediction servers at University College London. *Nucleic Acids Research*. 33:W36-38.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., *et al*. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *The Journal of Neuroscience*. 28:264 –278.

Callaerts, P., Clements, J., Francis, C. and Hens, K. (2006). Pax6 and eye development in Arthropoda. *Arthropod Structure & Development*. 35:379-391.

Cavodeassi, F., Young, F.C-B.R.M., Tada, M.L.C.M. and Wilson, S.W., Allende, M.L. and Houart, C. (2005). Early stages of zebrafish eye formation require the coordinated activity of Wnt11, Fz5, and the Wnt/β-catenin pathway. *Neuron*. 47:43–56.

Chiang, E.F-L., Yan, Y-L., Guiguen, Y., Postlethwait, J. and Chung, B-C. (2001). Two *Cyp19* (P450 Aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. *Molecular Biology and Evolution.* 18:542–550.

Christiansen, H.E., Lang, M.R., Pace, J.M. and Parichy, D.M. (2009). Critical early roles for *col27a1a* and *col27a1b* in zebrafish notochord morphogenesis, vertebral mineralization and post-embryonic axial growth. *PLoS ONE.* 4:e8481. doi:10.1371/journal.pone.0008481.

Eisen, J.S. and Smith, J.C. (2008). Controlling morpholino experiments: don't stop making antisense. *Development*. 135:1735-1743.

Escriva, H., Manzon, L., Youson, J. and Laudet, V. (2002). Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. *Molecular Biology and Evolution.* 19:1440-1450.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunesekaran, R., Ceric, G., Forslund, K., *et al*. (2010). The Pfam protein families database. *Nucleic Acids Research*. 38:D211-222.

Fukata, M., Nakagawa, M. and Kaibuchi, K. (2003). Roles of Rho-family GTPases in cell polarisation and directional migration. *Current Opinion in Cell Biology*. 15:590–597.

Fukui, Y., Tanaka, T., Tachikawa, H. and Ihara, S. (2007). SWAP-70 is required for oncogenic transformation by v-Src in mouse embryo fibroblasts. *Biochemical and Biophysical Research Communications*. 356:512–516.

Fukui, Y., Wakamatsu, I, Tachikawa, H, Okamura, Y., Tanaka, T. and Ihara, S. (2007). Activity of β3-β4 loop of the PH domain is required for the membrane targeting of SWAP-70. *IUBMB Life.* 59:99-103.

Gillis, W.Q., St John, J., Bowerman, B. and Schneider, S.Q. (2009). Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evolutionary Biology.* 9:207.

Goudevenou, K. (2010) PhD Thesis: *Def6 acts downstream of the non-canonical Wnt signalling pathway and is required for convergence and extension movements*.

Gross, B., Borggrefe, T., Wabl, M., Sivalenka, R.R, Bennett, M., Rossi, A.B. and Jessberger, R.(2002). SWAP-70-deficient mast cells are impaired in development and IgE-mediated degranulation. *European Journal of Immunology*. 32:1121–1128.

Guindon, S. and Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 52:696-704

Hall, A. (1998). Rho GTPases and the Actin Cytoskeleton. *Science.* 279:509.

Heasman, J. (2002). Morpholino oligos: making sense of antisense? *Developmental Biology*. 243:209–214.

Hilpela, P., Oberbanscheidt, P., Hahne, P., Hund, M, Kalhammer, G., Small, J.V. and Bahler, M. (2003). SWAP-70 identifies a transitional subset of actin filaments in motile cells. *Molecular Biology of the Cell.* 14:3242–3253.

Hoffmann, R. and Valencia, A. (2004). A Gene Network for Navigating the Literature. *Nature Genetics.* 36:664.

Hotfilder, M., Baxendale, S., Cross, M.A. and Sablitzky, F. (1999). Def-2, -3, -6 and -8, novel mouse genes differentially expressed in the haemopoietic system. *British Journal of Haematology.* 106:335-344.

Ihara, S., Oka, T. and Fukui, Y. (2005). Direct binding of SWAP-70 to non-muscle actin is required for membrane ruffling. *Journal of Cell Science*. 119:500-507.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. 292:195-202.

Kane, D.A. and Kimmel, C.B. (1993). The zebrafish midblastula transition. *Development*. 119:447-456.

Kumar, S., Dudley, J., Nei, M. and Tamura, K. (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*. 9: 299-306.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., *et al*. (2007). ClustalW and ClustalX version 2. *Bioinformatics*. 23: 2947-2948.

Lee, Y-N, Tuckerman, J., Nechushtan, H., Schutz, G., Razin, E. and Angel, P. (2004). c-Fos as a regulator of degranulation and cytokine production in FcεRI-activated mast cells. *The Journal of Immunology*. 173:2571–2577.

Letunic, I. Doerks, T. and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Research*. 37:D229-32.

Lupas, A., Van Dyke. and M., Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*. 24:1162-1164.

Mammalian Gene Collection Program Team. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences U.S.A.* 99:16899-16903.

Martin, P. (2007) PhD Thesis: *Analysis of Def6, a novel guanine nucleotide exchange factor showing dynamic regulation in vitro and playing an essential role in zebrafish development*.

Masat, L., Caldwell, J., Armstrong, R., Khoshnevisan, H., Jessberger, R., Herndier, B., Wabl, M. and Ferrick, D. (2000). Association of SWAP-70 with the B cell antigen receptor complex. *Proceedings of the National Academy of Sciences U.S.A.* 97:2180–2184.

Masat, L., Liddell, R.A., Mock, B.A. Kuo, W.L., Jessberger, R, Wabl, M., Morse III, H.C. (2000). Mapping of the SWAP70 gene to mouse Chromosome 7 and human Chromosome 11p15. *Immunogenetics.* 51:16–19.

Mavrakis, K.J., McKinlay, K.J. Jones, P. and Sablitzkya, F. (2004). DEF6, a novel PH-DH-like domain protein, is an upstream activator of the Rho GTPases Rac1, Cdc42, and RhoA. *Experimental Cell Research*. 294:335– 344.

Muffato, M., Louis, A., Poisnel, C-E., Roest Crollius, H. (2010). Genomicus:a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics*. 26:1119-1121.

Murugan, A.K., Ihara, S., Tokuda, E., Uematsu, K., Tsuchida, N. and Fukui, Y. (2008). SWAP-70 is important for invasive phenotypes of mouse embryo fibroblasts transformed by v-Src. *IUBMB Life*. 60: 236–240.

Oberbanscheidta, P., Balkowb, S., Kuhnla, K., Grabbeb, S. and Bahler, M. (2007). SWAP-70 associates transiently with macropinosomes. *European Journal of Cell Biology.* 86:13–24.

Ocana-Morgner, C., Wahren, W. and Jessberger, R. (2009). SWAP-70 regulates RhoA/RhoB-dependent MHCII surface localization in dendritic cells. *Blood.* 113: 1474-1482.

Panizzi, J.R., Jessen, J.R., Drummond. I.A. and Solnica-Krezel, L. (2007). New functions for a vertebrate Rho guanine nucleotide exchange factor in ciliated epithelia. *Development*. 134:921-931.

Pernis, A.B. (2009). Rho GTPase-mediated pathways in mature CD4+ T cells. *Autoimmunity Reviews*. 8:199–203.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y-L, Kelly, P.D., Chu, F., Huang, H., Hill-Force, A. and Talbot, W.S. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes.*Genome Research.* 10:1890-1902.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y-L, Kelly, P.D., Chu, F., Huang, H., Hill-Force, A. and Talbot, W.S. (2000) Zebrafish Comparative Genomics and the Origins of Vertebrate Chromosomes. *Genome Research*. 10.1890-1902.

Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A, Donovan, A, Egan, E.S., Force, A., Gong, Z., et al. (1998). Vertebrate genome evolution and the zebrafish gene map. *Nature Genetics.* 18:345-349.

Raftopoulou, M. and Hall, A. (2004). Cell migration: Rho GTPases lead the way *Developmental Biology*. 265:23– 32.

Rice, P., Longden, I., Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*. 16:276-277.

Robu, M.E., Larson, J.D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S.A. and Ekker, S.C. (2007). p53 activation by knockdown technologies. *PLoS Genetics*. 25: e78.

Rossman, Der, C.J. and Sondek, J. (2005). GEF means go: turning on Rho GTPases with guanine nucleotide exchange factors. *Nature Reviews Molecular Cell Biology*. 6:167-180.

Roszko, I., Sawada, A., and Solnica-Krezel, L. (2009). Regulation of convergence and extension movements during vertebrate gastrulation by the Wnt/PCP pathway. *Seminars in Cell & Developmental Biology.* 20:986–997.

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 5:725-738.

Sato, Y. and Nishida, M. (2010). Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environmental Biology of Fishes*. 88:169–188.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences U.S.A.* 95:5857-5864.

Shinohara, M., Terada, Y., Iwamatsu, A., Shinohara, A., Mochizuki, N., Higuchi, M., Gotoh, Y., Ihara, S., Nagata, S., Itoh, H., Fukui, Y and Jessbergerk, R. (2002) SWAP-70 is a guanine-nucleotide exchange factor that mediates signalling of membrane ruffling. *Nature*. 416:759-763.

Sivalenka, R.R. and Jessberger, R. (2004). SWAP-70 regulates c-kit-induced mast cell activation, cell-cell adhesion, and migration. *Molecular and Cellular Biology*. 24:10277–10288.

Sivalenka, R.R., Sinha, M. and Jessberger, R. (2008) SWAP-70 regulates mast cell FceRI-mediated signalling and anaphylaxis. *European Journal of Immunology*. 38:841–854.

Spudich, G.M. and Fernández-Suárez, X.M. (2010). Touring Ensembl: a practical guide to genome browsing. *BMC Genomics*. 11:295.

Takada, N. and Appel, B. (2010). Identification of genes expressed by zebrafish oligodendrocytes using a differential microarray screen. *Developmental Dynamics*. 239:2041-2047.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution.* 24:1596-1599.

The UniProt Consortium (2010). The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research*. 38:D142-D148.

Tybulewicz, V.L.J. and Henderson, R.B. (2009). Rho family GTPases and their regulators in lymphocytes. *Nature Reviews Immunology*. 9:630-644.

Wakamatsu, I., Ihara, S. and Fukui, Y. (2006). Mutational analysis on the function of the SWAP-70 PH domain. *Molecular and Cellular Biochemistry.* 293:137–145.

Wang, H. (2008). Comparative Analysis of Period Genes in Teleost Fish Genomes. *Journal of Molecular Evolution*. 67:29–40.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25:1189-1191.

Welcha, H.C.E, Coadwell, W,J, Stephensa, L.R. and Hawkinsa, P.T. (2003). Phosphoinositide 3-kinase-dependent activation of Rac. *FEBSLetters*. 546:93-97.

Yong Wang, Y., Chen, K., Yao, Q., Zheng, X.D. and Yang, Z. (2009). Phylogenetic Analysis of Zebrafish Basic Helix-Loop-Helix Transcription Factors. *Journal of Molecular Evolution*. 68:629–640.

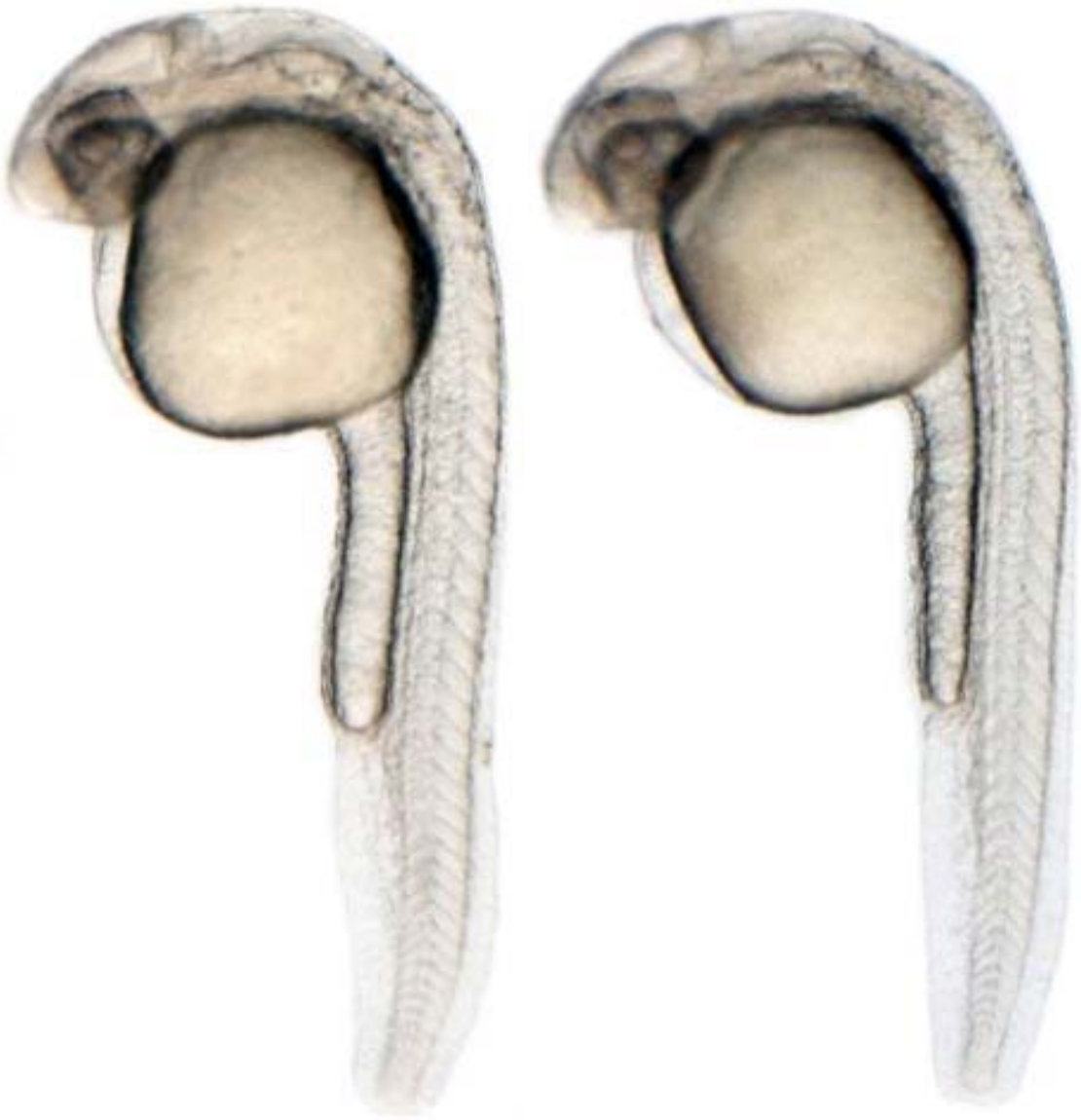
Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 9:40.

Zhang, Y. (2009). I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins*. S9:100-113.

Zheng, Y. (2001). Dbl family guanine nucleotide exchange factors. *TRENDS in Biochemical Sciences*. 26:724-32.

Appendices

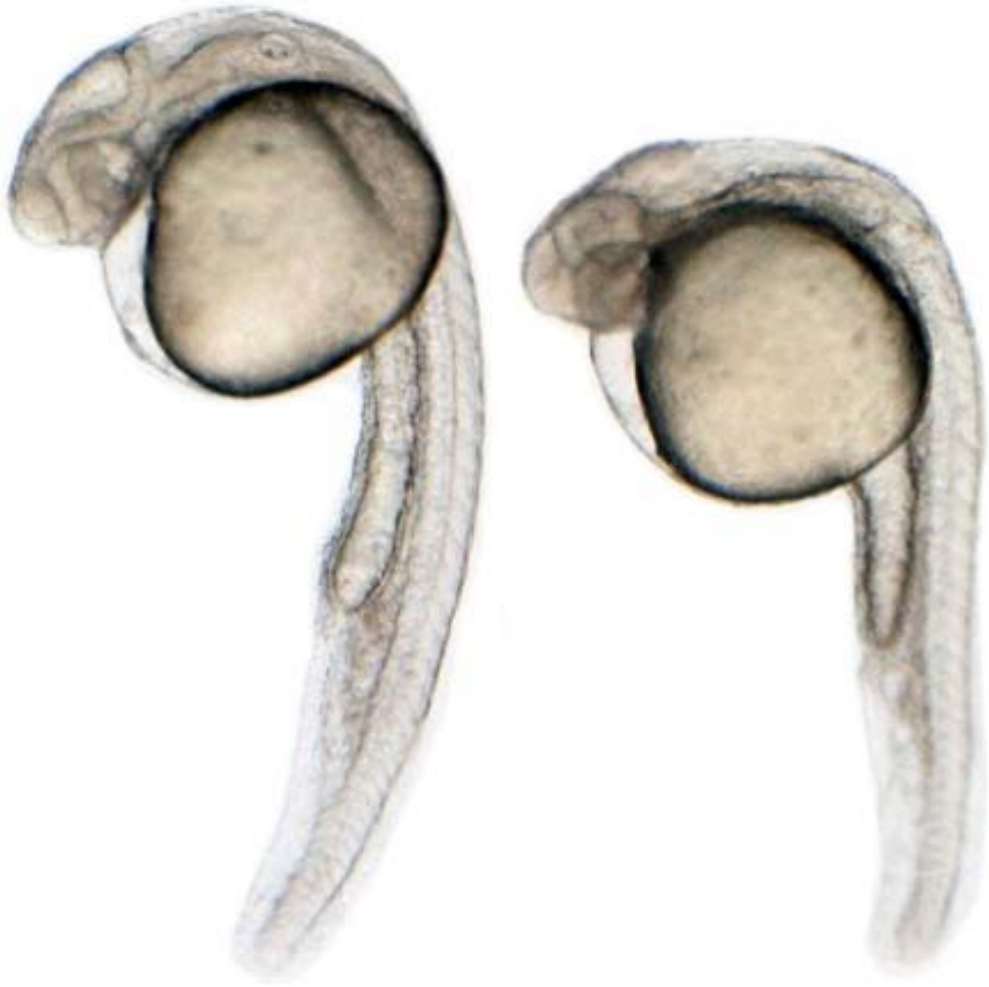
Appendix I - Enlarged 250pg mRNA Injected Embryo Phenotypes



Wild Type

Normal

Appendix I - Enlarged 250pg mRNA Injected Embryo Phenotypes (Continuous)



Mild Moderate

Appendix II - Enlarged 500pg mRNA Injected Embryo Phenotypes

Wild Type

Normal

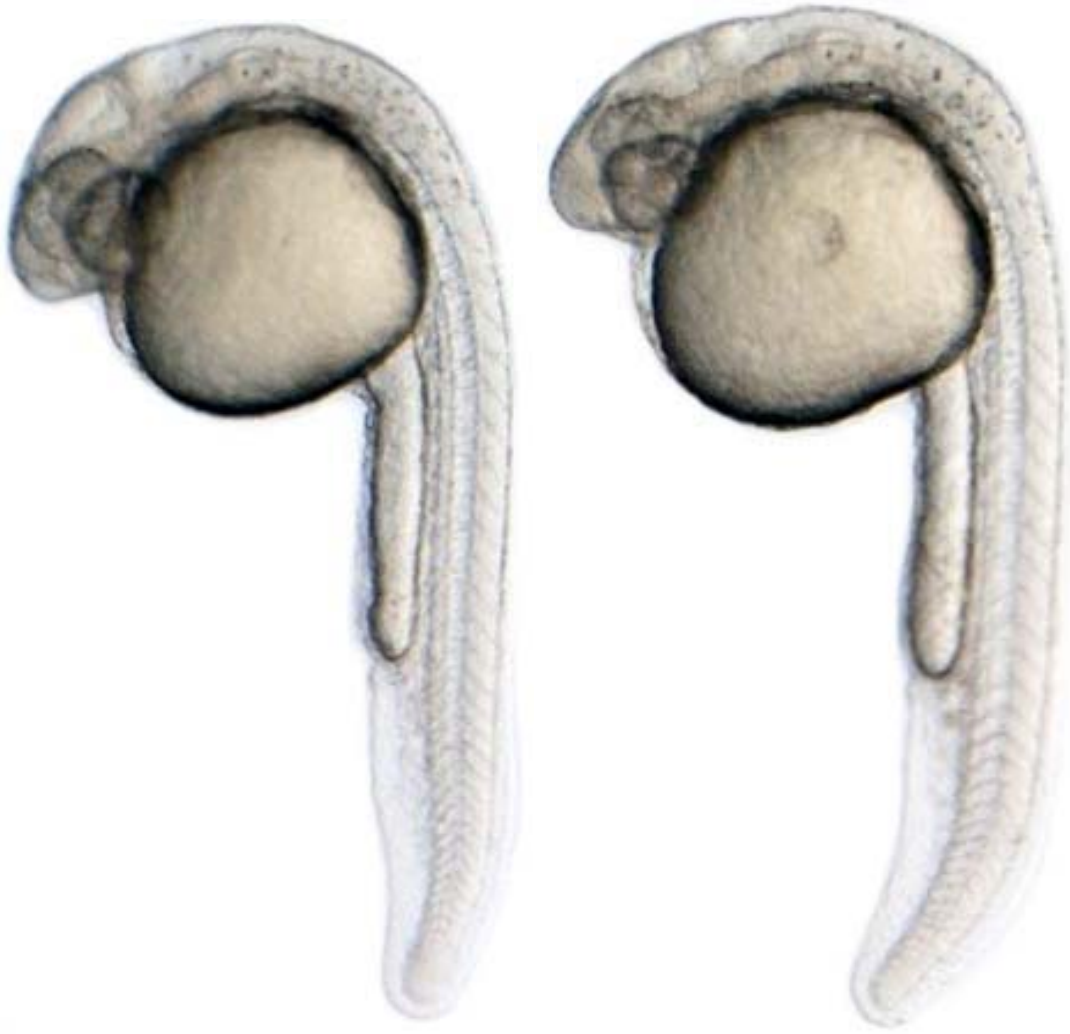
Appendix II - Enlarged 500pg mRNA Injected Embryo Phenotypes (Continuous)

Mild

Moderate

Severe

Appendix III - Enlarged 750pg mRNA Injected Embryo Phenotypes



Wild Type

Normal

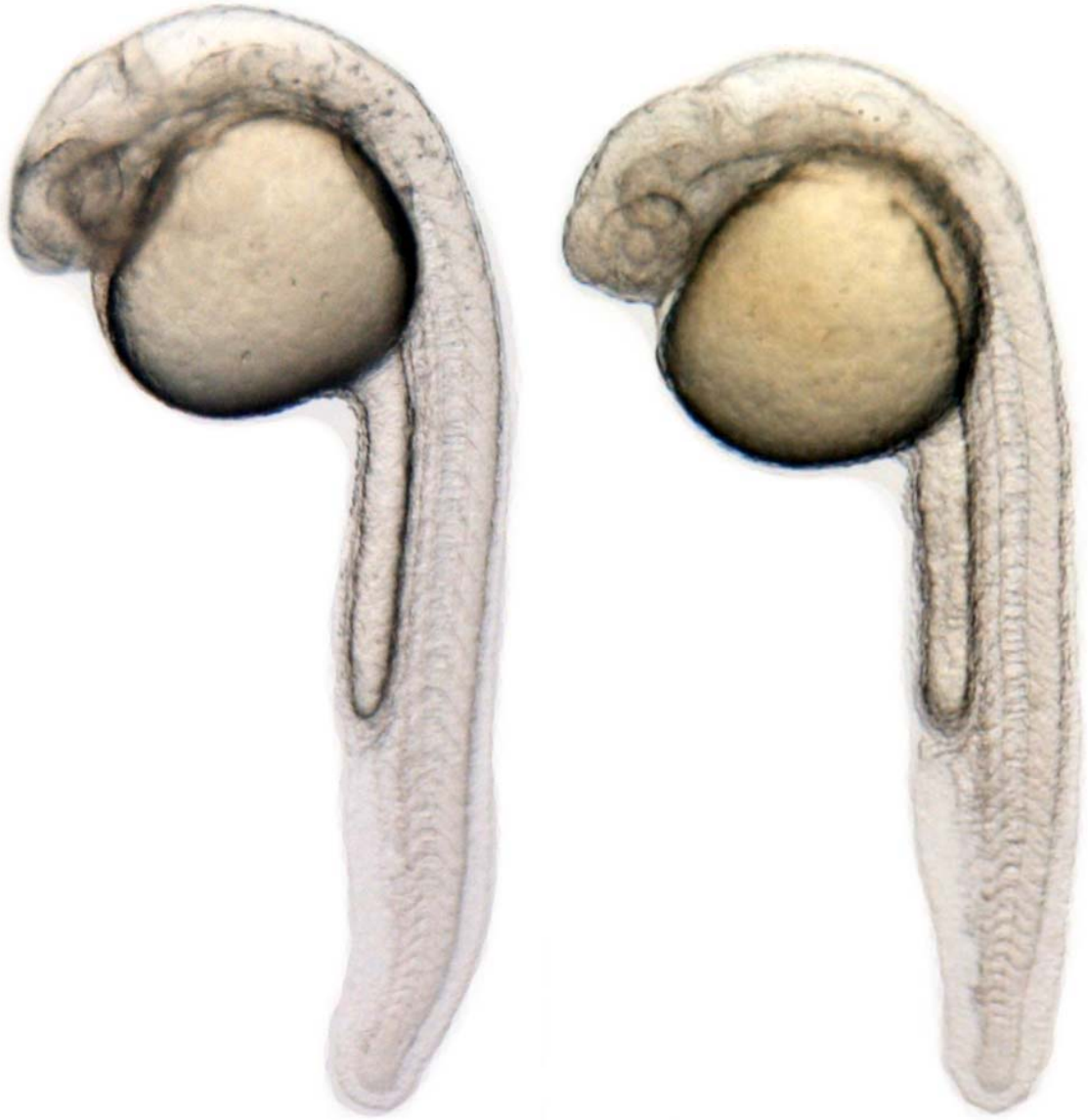
Appendix III - Enlarged 750pg mRNA Injected Embryo Phenotypes (Continuous)

Mild

Moderate

Severe

Appendix IV - Enlarged 2.5ng Splice MO1 Injected Embryo Phenotypes

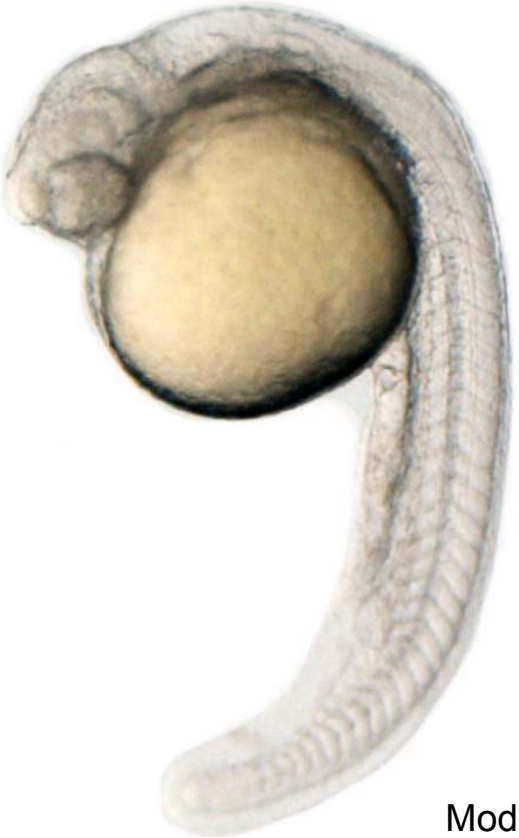


Wild Type

Normal

Appendix IV - Enlarged 2.5ng Splice MO Injected Embryo Phenotypes (Continuous)

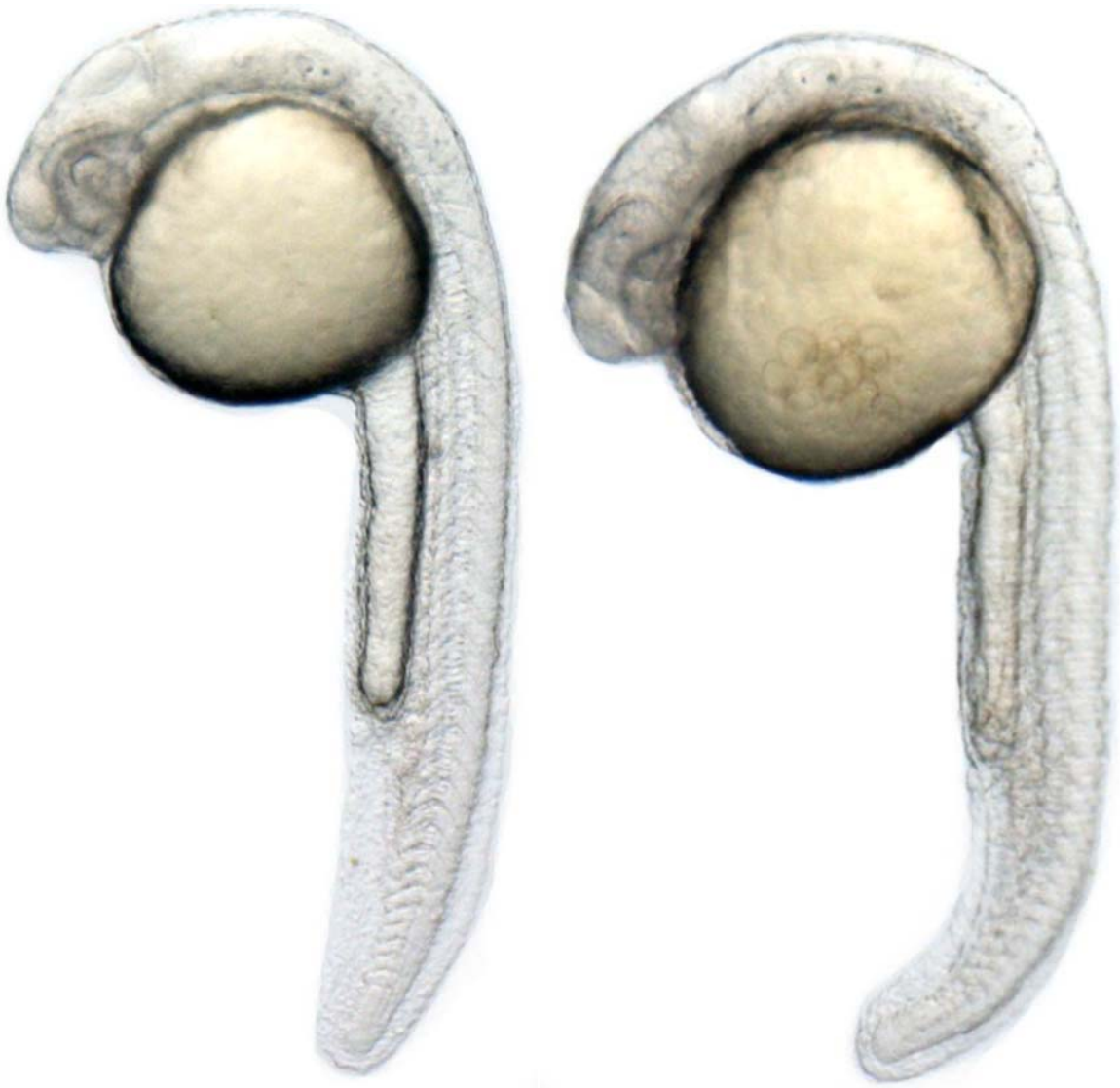
Mild



Moderate

Severe

Appendix V - Enlarged 5.0ng Splice MO Injected Embryo Phenotypes



Wild Type

Normal

Appendix V - Enlarged 5.0ng Splice MO Injected Embryo Phenotypes (Continuous)

Mild

Moderate

Severe

Very Severe

Appendix VI - Enlarged 7.5ng Splice MO1 Injected Embryo Phenotypes

Wild Type

Moderate

Severe

Very Severe