Regulatory elements, protein function and evolution of the *actinodin* genes

Daniel Moses
Master's Thesis

Department of Biology, Faculty of Science and
Faculty of Graduate and Postdoctoral Studies,
University of Ottawa

Abstract

Small fibrils termed actinotrichia are involved with the growth and structure of the fin fold during fin development in fish. The *actinodin* (*and*) genes are required for actinotrichia formation, and the loss of these genes from the genomes of tetrapods has been implicated in the tetrapod-specific loss of actinotrichia, loss of a fin fold and the concurrent evolution of paired fins into limbs. This study focuses on the function of the *and* genes and their role in actinotrichia formation. The results reveal *cis*-acting regulatory elements required for *and1* expression in the fin epithelium. Furthermore, it is shown that the And proteins display similarities to the secreted signaling molecule, Ecrg4, implying a possible role in cell differentiation during fin fold development. In the final section of this report, I use a genomic analysis to show that the *and* genes were lost from otherwise well-conserved syntenic loci in fish and tetrapod genomes. These results suggest possible causes for the evolutionary loss of *and* genes and the associated developmental changes that may have permitted fin to limb evolution.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1R – 1st Round of whole genome duplication
2R – 2nd Round of whole genome duplication
3R – 3rd Round of whole genome duplication
AER – Apical ectodermal ridge
adipoql2 *– adiponectin a*
and *– actinodin*
AP – Anterior-posterior
ap4m1 *– adaptor-related protein complex 4*
bmp *– bone morphogenic protein*
bp – basepair
cdx *– caudal-related homeobox*
Ch – Chromosome
cops6 *– constitutive photomorphogenic homolog subunit 6*
DEPC – Diethylpyrocarbonate
dlx *–distalless-related homeobox*
dpf – days post fertilization
DV – Dorsal-ventral
ecrg4 *– esophageal cancer related gene 4*
eGFP – enhanced Green fluorescent protein
en *– engrailed*
ER – Endoplasmic reticulum
ET37 – Enhancer trap 37
fgf *– fibroblast growth factor*
FW – Forward primer
hox *– homeobox*
hpf – hours post fertilization
H3P – Phosphorylated histone-3
HRP – Horseradish peroxidase
igf *– insulin growth factor*
kb – kilobasepair
KCL – Potassium chloride
kDa – kiloDalton
LPM – Lateral plate mesoderm
mcm7 *– minichromosome maintenance deficient 7*
Mb – Megabasebair
MSA – Multiple sequence alignment
myeov2 *– myeloma overexpressed 2*
mya – million years ago
otos *– otospiralin*
PAGE – Polyacrylamide gel electrophoresis
PC – Proprotein convertase
PCR – Polymerase chain reaction
PD – Proximal-distal
per1a *– period homolog 1a*

pmx *– paired-related homeobox*
RA – Retinoic acid
RBH – Reciprocal best hit
REV – Reverse primer
RX(K/R)R – *arginine-x-(lysine* or *arginine)-arginine*
RXRR – *arginine-x-arginine-arginine*
SDS - Sodium dodecyl sulfate
shh *– sonic hedgehog*
stag3l3 *– stromal antigen 3-like 3*
TBST - Tris-buffered saline and tween 20
TEM – transmission electron microscopy
vamp2 *- vesicle-associated membrane protein 2*
ZPA – Zone of polarizing activity

Acknowledgements

I would firstly like to thank my supervisor Dr. Marie-Andrée Akimenko for providing the opportunity for me to learn and grow as a member of her lab. Her guidance, expertise, and support were invaluable to my experience as a student. I am especially grateful to her for all of the insightful discussions about the *actinodin* project, as well as her openness to allow me to explore various techniques throughout the course of this study.

I would also like to thank my committee members, Dr. Marc Ekker and Dr. Tom Moon. I greatly appreciate their useful comments and input throughout the course of my studies.

In addition, I would like to thank all of my colleagues in the Akimenko lab, including Shirine Jeradi, Ali Al-Rewashdy, Kevin Lin, Stephanie McMillan, and Robert Lalonde for their continuous help and positivity. I would like to send a special thanks to Jing Zhang for her patience, assistance, and leadership in the Akimenko lab. I would like to send another special thanks to Vishal Saxena for providing care and maintenance for the zebrafish and aquatic facilities.

Furthermore, I am grateful for the assistance provided by the Honour's and summer students who worked on the *actinodin* project including Helena Bleeker, Erica Gifford, Weishi Liu, Samuel Drapeau and Léa Guinot. I would also like to thank all of the other Honour's, co-op, scholarship, and summer students for their hard work in the Akimenko lab. I appreciate all the kindness and help from other colleagues, professors and lab technicians at the University of Ottawa Department of Biology.

Finally, I would like to thank my friends and family for their constant support.

Funding for this research was provided by the Natural Science and Engineering Council (NSERC) to Dr. Marie-Andrée Akimenko.

Chapter 1 - General introduction

1.1. Overview

The vertebrate limb has been used as a paradigm for the study of organogenesis. The cellular and molecular interactions that orchestrate limb development have provided insight for the types of mechanisms required during complex developmental processes (Zeller et al., 2009). Furthermore, an understanding of limb development is beneficial for the treatment of congenital limb defects among other medical applications. The limb is a useful tool for biological study because it can be experimentally manipulated without causing death to the organism. Model tetrapod species that have commonly been used for studies on limb development include the mouse, chick, and *Xenopus*. Interestingly, the early stages of tetrapod limb and fish paired fin development are controlled by highly conserved genetic mechanisms. For this reason, zebrafish pectoral fin development has become a model for the study of early events in limb/fin development (reviewed in (Mercader, 2007)).

In addition to similar developmental mechanisms, the fossil record supports the hypothesis that the tetrapod paired fore- and hind- limbs are homologous to the fish paired pectoral and pelvic fins, respectively (Callier et al., 2009; Clack, 2012). The evolution of limbs from fins, as well as the evolution of lungs, were some of the key events required for the vertebrate transition from aquatic to terrestrial environments during the Devonian over 350 million years ago (mya) (Clack, 2012). Currently, with a broad understanding of the developmental mechanisms required for fin and limb development, researchers have begun to uncover precise genetic changes that contributed to the evolution of fins into limbs (for example see (Schneider et al., 2011). In this study, I focus on the *actinodin* (*and*) genes which are

expressed in the fins and are specific to fish genomes (Zhang et al., 2010). It is hypothesized that the loss of the *and* genes from tetrapod genomes caused a change in the developmental program of the appendages which contributed to fin-to-limb evolution.

1.2. Early stages in fin and limb development

During vertebrate embryogenesis, retinoic acid (RA) is important for anterior-posterior (AP) patterning of the body axis and is also required for limb and fin development (Begemann et al., 2001; Grandel et al., 2002; Niederreither et al., 1999; Niederreither et al., 2000). The combination of RA and other genetic factors at specific levels along the AP axis will induce the outgrowth of the appendage from the trunk of the body (Duboc and Logan, 2011; Mercader et al., 2006). The limb or fin primordium takes on the form of a bud of mesenchyme cells that protrudes from the trunk and is covered by a layer of ectoderm cells. The mesenchyme cells of the bud are derived from the lateral plate mesoderm (LPM). The ectoderm overlying the mesenchyme is initially equivalent to the ectoderm covering the trunk of the embryo (Altabef et al., 1997).

The three-dimensional axes of the paired fin and limb buds are organized by three distinct signaling centres. The apical ectodermal ridge (AER) organizes development along the proximal-distal (PD) axis, the zone of polarizing activity (ZPA) controls patterning along the AP axis of the bud, and the non-AER ectoderm controls the development of the dorsal-ventral (DV) axis. The formation of the AER is concomitant with the early genetic signaling events in the bud. In mice, chicks, and zebrafish, one of the earliest factors expressed in the bud mesenchyme is *fibroblast growth factor 10* (*fgf10*) (Agarwal et al., 2003; Mercader et al., 2006; Ng et al., 2002). Functional analyses have shown that via signal transduction pathways, mesenchymal *fgf10*

activates *fgf8* expression in the overlying ectoderm (Galceran et al., 1999; Kawakami et al., 2001; Norton et al., 2005). The interactions between mesenchyme and ectoderm establish the formation of the AER, a thickening of ectoderm located at the distal edge of the bud along the AP axis. Formation of the AER is typical of vertebrate limb development (Fernandez-Teran and Ros, 2008; Tabin and Wolpert, 2007) and has been documented in the paired fins (Bouvet, 1978; Wood, 1982) and the median fins of teleosts (Dane and Tucker, 1985). A key function of the AER is to maintain *fgf* expression in the bud which is required for cell proliferation and cell survival (reviewed by (Fernandez-Teran and Ros, 2008), (Mercader, 2007)). Ultimately, the AER-induced proliferation of mesenchyme cells permits the outgrowth of the bud along the PD axis.

The gene *sonic hedgehog* (*shh*) encodes a signaling molecule that is localized to the posterior margin of the bud where it establishes a polarity across the AP axis (Riddle et al., 1993). In tetrapod limbs, this posterior margin, termed the ZPA, releases SHH as a morphogen which forms a gradient across the bud and differentially controls the transcription of downstream target genes (Bénazet and Zeller, 2009). Posterior localization of *shh* in the paired fin buds has been identified in teleosts and cartilaginous fish where it plays a similar posteriorizing role (Dahn et al., 2007; Krauss et al., 1993; Yonei-Tamura et al.). In regard to polarizing the DV axis, tetrapods and zebrafish exhibit similar gene expression patterns of *wnt7a* in the dorsal ectoderm (Norton et al., 2005; Parr and McMahon, 1995) and *bone morphogenic protein 7a* (*bmp7a*) and *engrailed 1* (*en1*) in the ventral ectoderm (Ahn et al., 2001; Hatta et al., 1991). During limb development *Wnt7a* versus *Bmp7a* and *En1* specify dorsal and ventral morphologies, respectively (Logan et al., 1997; Pizette et al., 2001; Rodriguez-Esteban et al., 1998). Patterning along the three axes of the bud is coordinated and the activity of the signaling centres impact

each other. For example, *bmp7* is expressed in the ventral ectoderm up to the presumptive level of the AER and is implicated in specification of the location of the AER (Robert, 2007). Additionally, signals from the ZPA maintain the overlying ectoderm, while in turn the AER is required for ZPA activity (Harfe, 2011; Todt and Fallon, 1987).

1.3. Differences in early fin and limb development

The early events in fin and limb development are evidently very similar. Despite these striking developmental similarities, fins and limbs display obvious morphological differences in adulthood. One of the earliest morphological divergences concerns the fate of the ectoderm of the bud. The AER is active throughout the development of the tetrapod limb, however during early stages of fin development, the cells of the AER elongate distally to form a structure termed the fin fold which will give rise to the fin rays (Fig 1.1) (Dane and Tucker, 1985; Wood, 1982). The emergence of a fin fold has not been observed in tetrapod limbs and has been theorized as a significant determinant of the morphological differences between adult fins and limbs (Thorogood, 1991; Yano and Tamura, 2013).

1.4. Tetrapod limb morphology and later stages of limb development

1.4.1. Morphology of adult limbs

The general organization of the limb skeleton is conserved in tetrapod species. In order from proximal to distal, the skeleton segments are divided into the stylopod, zeugopod, and autopod (Fig 1.1b). In typical tetrapods, the stylopod consists of the humerus or femur bones, the zeugopod corresponds to the radius/ulna or the fibula/tibia, and the autopod corresponds to the carpals or tarsals, metacarpals or metatarsals, and phalanges also known as the digits.

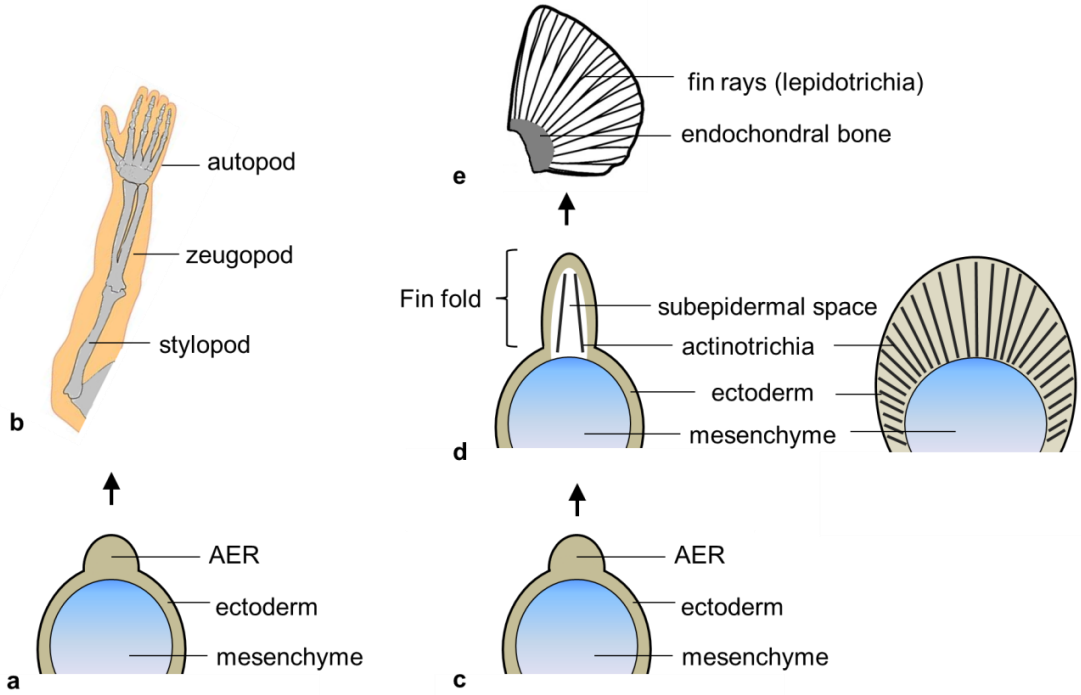


Fig 1.1. Different fates of the ectoderm during fin and limb development. a, The limb bud initiates as a grouping of mesenchyme cells and a sheet of overlying ectoderm cells. The apical ectodermal ridge (AER) is condensed at the distal tip and permits outgrowth of the mesenchyme in the distal direction. **b,** The adult limb consists of endochondral bones (grey) throughout the entire organ. The endochondral bone differentiates from mesenchyme cells. Persistent activity of the AER allows for the outgrowth of the mesenchyme which will result in the pattern of endochondral bone shown here. **c,** The fin bud initiates with the same basic structure as the early limb bud. **d,** Fin bud with fin fold shown as a cross sectional view and a lateral view on the left and right sides, respectively. The actinotrichia fibrils (straight black lines) appear and push the distal ectoderm away from the mesenchyme resulting in a distinct fin fold. **e,** The adult fin consists of endochondral bone at the proximal region, and fin rays (lepidotrichia) constitute the majority of the structure (adult fin image adapted from (Collar et al., 2008)).

1.4.2. Later stages of limb development

Experiments performed by Saunders and Rubin (Rubin and Saunders, 1972; Saunders, 1948) revealed that the limb develops in a proximal to distal sequence. In these experiments, the authors surgically removed the chick limb bud AER at an early stage in development. The result was that only the proximal limb elements, for example only the stylopod, developed (Rubin and Saunders, 1972; Saunders, 1948). Furthermore, when the AER was removed at later stages, the stylopod and zeugopod formed although the autopod was absent (Rubin and Saunders, 1972; Saunders, 1948). This revealed that proximal elements were specified early, and that the AER was required for further development of the distal elements. Additionally, Saunders et al. (Saunders et al., 1968) showed that when a forelimb AER is transplanted to the hindlimb mesenchyme, the result is the formation of a hindlimb. Additionally, when a hindlimb AER is transplanted to the forelimb mesenchyme, a forelimb arises (Saunders etal., 1968). Overall, these experiments revealed that the mesenchyme is competent for the differentiation into the proper limb elements, although it requires the AER for a permissive and instructive signal to continue development along the PD axis (Fernandez-Teran and Ros, 2008; Sheth et al., 2013).

The mesenchyme cells of the bud will differentiate into bone, tendons and different types of vasculature tissues (Duboc and Logan, 2011). The bones of the tetrapod limb are endoskeleton structures. Endoskeleton forms by the process of endochondral ossification which involves the formation of a cartilage precursor skeleton and the subsequent replacement of this cartilage tissue by bone-matrix tissue (Safadi et al., 2009). Members of the *homeobox* (*hox*) gene family are downstream targets of the three signaling centres. *Hox* genes are required for the patterning of the endoskeleton elements along the PD axis. Expression of *Hoxa9, Hoxa11,* and *Hoxa13* are correlated with formation of the stylopod, zeugopod, and autopod, respectively (Yokouchi et al.,

1991). Additionally, the 5'-genes of the *Hoxd* cluster, including *Hoxd9* to *Hoxd13*, are differentially distributed along the PD and AP axes and mediate normal bone patterning (Nelson et al., 1996; Shubin et al., 2009; Woltering and Duboule, 2010).

1.5. Fish fin morphology and later stages of fin development

1.5.1. Morphology of adult fins

The typical structure of a pectoral or pelvic fin consists of an endoskeleton base and an exoskeleton consisting of fin rays that extend from the distal end (Fig 1.1e). The zebrafish paired fin endoskeleton is composed of radial bones that are confined to the proximal domain of the fin (Grandel and Schulte-Merker, 1998). The chondrichthyes, or cartilaginous fish, do not form endoskeleton bone although they have cartilaginous radial-like structures termed pterygiophores which are similarly confined to the proximal domain of the paired fins (See Fig 1.2 for phylogenetic relationships) (De Iuliis and Pulera, 2011). In the paired fins of zebrafish and chondrichthyes, multiple radials or pterygiophores articulate with the adjacent girdle (De Iuliis and Pulera, 2011; Grandel and Schulte-Merker, 1998). Additionally, in teleosts and chondrichthyes, the fin rays are the prominent structures of the fin. Teleost fin rays are composed of dermal bone elements termed the lepidotrichia (Géraudie and Landis, 1982), which are a characteristic feature of the ray-finned fish. Dermal bone is formed through the process of intramembranous ossification, by which cells aggregate and differentiate into osteoblasts that directly synthesize a new bone matrix (Safadi et al., 2009). Unlike endochondral ossification, dermal bone development does not involve a cartilaginous intermediate. Each lepidotrichia consists of two concave, symmetrical hemirays that face each other and enclose vasculature, nerve, and connective tissues. Located at the distal tips of the lepidotrichia are tufts of relatively

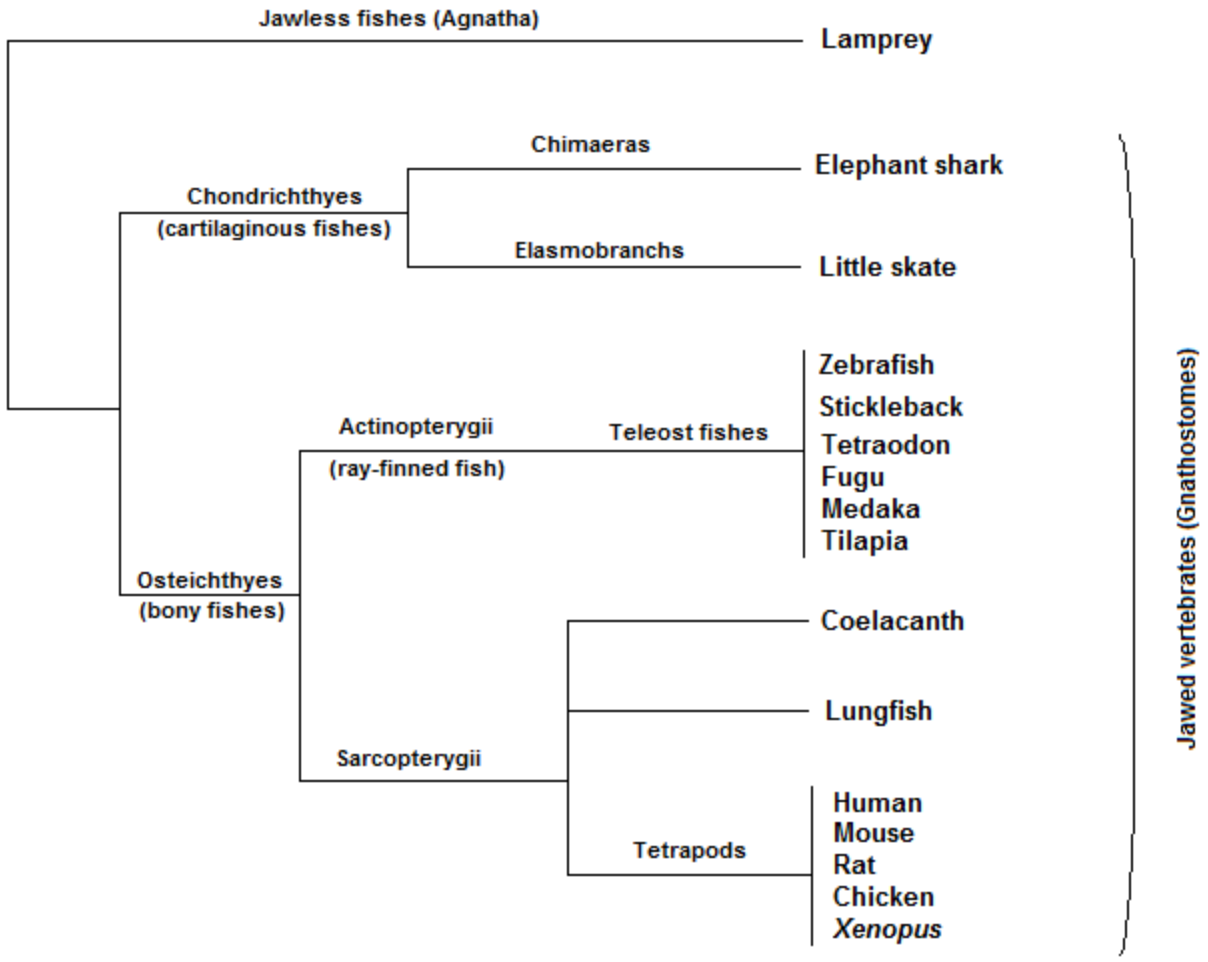


Fig 1.2. Vertebrate phylogenetic tree. Jawed vertebrates (gnathostomes) can be divided into two major groups, the chondrichthyes and the osteichthyes. The osteichthyes can be further divided into two groups, the actinopterygii (ray-finned fish) and the sarcopterygii (lobe-finned fish and tetrapods). Examples of extant species within each group are listed on the right side of the tree. *And* genes have been identified in the genomes of species within the chondrichthyes and actinopterygii lineages, but are absent from tetrapod genomes.

small unmineralized fibrils termed actinotrichia (Akimenko et al., 2003; Marí-Beffa et al., 1989). It is possible that actinotrichia support the distal ectoderm and provide flexibility to the distal fin ray in adult fish. The fin rays of the chondrichthyes are straight, semi-rigid, collagenous rods termed the ceratotrichia (Chandross, 1982; De Iuliis and Pulera, 2011; Lingham-Soliar, 2005a; Lingham-Soliar, 2005b).

The aquatic sarcopterygians, or the lobe-finned fish which include the extant coelacanth and lungfish, are phylogenetically closer relatives to the tetrapods versus other fish species (Fig 1.2) (Nelson, 1994). The morphology of the lobed fins is regarded as being an intermediate to the teleost fins and tetrapod limbs (Lecointre and Guyader, 2006). The lobed fin is similar to the limb such that it has a large endoskeleton that binds to the skeletal girdle via a single articulation, defined as a monobasal appendage (Nelson, 1994). However, the lobed fin has fin rays at the distal apex reminiscent of a teleost fin.

In addition to the paired fins, fish species possess median, unpaired fins. The median fins of the zebrafish include the dorsal, caudal, and anal fins (Fig 1.3b). The teleost dorsal and anal fins are supported by proximal endoskeleton elements (Crotwell et al., 2004), and exhibit the conventional lepidotrichia fin rays with actinotrichia at their distal tips. Like the paired fins, the median fin lepidotrichia occupy the majority of the fin structure. Similarly, the chondrichthyes median fins are largely supported by their fin rays, the ceratotrichia (De Iuliis and Pulera, 2011).

1.5.2. Later stages of fin development

Mesenchymal cells of the fin bud are competent for undergoing endochondral ossification, as is the case in the limb bud mesenchyme. Orthologs of the *hox* genes required for endoskeleton development in the limb are active during fin endoskeleton development

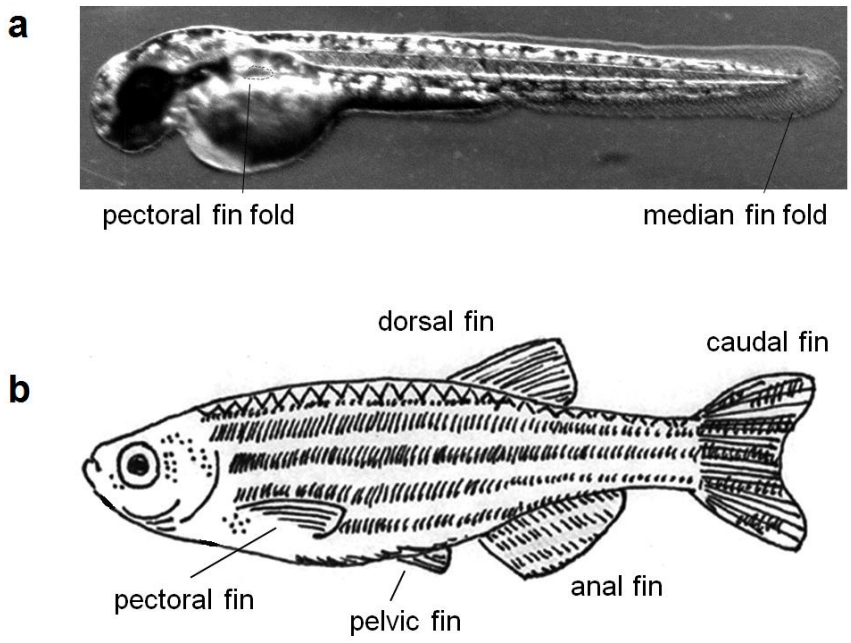


Fig 1.3. Fin folds of the zebrafish embryo and the derived fins in adults. a, 48 hours post fertilization (hpf) embryo showing the median fin fold encompassing the caudal region of the tail, and the pectoral fin fold on the lateral side posterior to the eye. The pelvic fin fold appears later during development. **b,** There are five fin types in the adult zebrafish: the paired pectoral and pelvic fins derive from paired fin buds, and the three single fins all derive from the median fin fold (adult fish image adapted from (Jackman and Stock, 2006)).

(Ahn and Ho, 2008). During later stages of limb development, there is a separation of the domains of *Hoxa11* and *Hoxa13* which correlates to zeugopod and autopod development, respectively (Nelson et al., 1996; Yokouchi et al., 1991). During early fin development, these genes are divided along the PD axis of the fin mesenchyme in a similar manner as the tetrapod expression pattern. However, in zebrafish, medaka, and a basal actinopterygian species, the paddlefish, *hoxa11* and *hoxa13* expression is no longer separated, but overlaps during later developmental stages (Davis et al., 2007; Takamatsu et al., 2007). The overlap of the *hoxa* genes in fins but not in limbs likely contributes to the morphological differences between fin and limb endoskeleton (reviewed by (Woltering and Duboule, 2010)).

In teleost fins, when few endoskeleton elements have been specified, the cells of the AER begin to fold on themselves and elongate distally (Dane and Tucker, 1985) resulting in two opposing layers of ectodermal cells. This apical structure is termed the fin fold (JR Hinchliffe, 1980) and has not been observed in limb development. Gene expression analyses have shown that markers of the AER are present at the distal ectoderm of the fin fold of pectoral fins, also referred to as the cleft cells in the median fin fold (reviewed by (Yano and Tamura, 2013), (Dane and Tucker, 1985)). However, it is debated whether the AER and fin fold are capable of performing homologous functions (Grandel and Schulte-Merker, 1998; Yano et al., 2012). Using light microscopy, Wood (1982) showed that in the killifish pectoral fin mesenchyme, endochondral condensations are established at the time the AER transforms into a fin fold (Wood, 1982). It has also been proposed that once the AER transforms into a fin fold, proliferation is reduced in the mesenchyme (Capdevila and Belmonte, 2001; Géraudie, 1978). Overall, these previous reports suggest that during early stages of fin development, the AER

permits endoskeleton differentiation and mesenchyme outgrowth. However, once the fin fold is established, there is an apparent loss of the signaling events provoked by the AER.

Between the two ectodermal layers of the fin fold, the dermal fin rays structures are synthesized. Moreover, only dermal skeleton and not endoskeleton bone development, occurs in the fold tissue. The pectoral fin fold is located distally to the endochondral disk at the base of the fin (Yano et al., 2012). The median fins are derived from a median fin fold that runs along the entire dorsal and ventral midline of the fish embryo (Fig 1.3a). At the zebrafish larval stage, the fin fold tissues regress in between the locations of the adult median fins (Van Eeden et al., 1996). Interestingly, the tadpoles of frogs, an amphibian and basal species in the tetrapod lineage, have a median fin fold that never develops fin rays and regresses (Doherty et al., 1998).

The formation of the fin fold is coordinated with the synthesis of the unmineralized extracellular fibrils termed actinotrichia. The actinotrichia are the earliest extracellular elements established in the fold and the lepidotrichia subsequently calcify in a proximal to distal direction (Géraudie and Landis, 1982). The actinotrichia fibrils remain at the distal tip of the lepidotrichia throughout development and adulthood. Actinotrichia fibrils are small structures and may seem negligible in the adult fins. However, previous studies have provided evidence for the requirement of the actinotrichia towards the initiation of fin ray development (Santamaria et al., 1996; Thorogood and Wood, 1987; Wood and Thorogood, 1984; Zhang et al., 2010). Additionally, it has been hypothesized that the evolutionary loss of actinotrichia resulted in the loss of fin ray development in tetrapod limbs (Yano and Tamura, 2013; Zhang et al., 2010). Due to their importance for fin development and their implications in regards to fin-to-limb evolution, the genetic mechanisms related to actinotrichia formation and function are of major focus in the remainder of this study.

1.5.3. Morphology and function of the actinotrichia

At the initial stages of fin fold development, the actinotrichia fibrils are oriented along the proximal-distal axis, and form two parallel arrays, where each array runs along the basal surface of each opposing ectodermal layer of the fin fold (Fig 1.1d) (Bouvet, 1974; Géraudie, 1977). Actinotrichia have varying widths depending on the species, yet as observed by transmission electron microscopy (TEM) they consistently display a banding appearance (Dane and Tucker, 1985; Geraudie and Meunier, 1980; Géraudie, 1977; Montes et al., 1982) and taper towards the ends (Becerra et al., 1983; Géraudie and Landis, 1982). The fin fold and actinotrichia together grow distally, and the function of the actinotrichia is likely to provide a structural support for the fin fold. Furthermore, the actinotrichia are used as a substratum for the migration of the mesenchyme cells from the base of the fin fold into the subepidermal space of fold (Thorogood and Wood, 1987; Wood and Thorogood, 1984). A subset of these mesenchymal cells will differentiate into the lepidotrichia (Géraudie and Landis, 1982). It has been suggested that actinotrichia exert a morphogenic influence on the development of the lepidotrichia (Santamaria et al., 1996; Santamaría and Becerra, 1991). The actinotrichia can be considered an important morphological aspect of fin development that is conducive to fin ray ontogenesis.

1.5.4. Molecular composition of actinotrichia

Actinotrichia are composed of a molecular substance called elastoidin. Elastoidin was first described from the ceratotrichia of chondrichthyes fins (Krukenberg, 1885). They share the same chemical composition as the actinotrichia and display the same regular banding pattern as

observed by TEM, thus the ceratotrichia and actinotrichia are considered to be homologous (Chandross and Bear, 1979; Géraudie, 1988; Géraudie and Landis, 1982; Goodrich, 1904; Marí-Beffa et al., 1989). Elastoidin is composed of collagens (Durán et al., 2011; Woodhead-Galloway and Knight, 1977) plus a non-collagenous substance that is rich in tyrosine (Damodaran et al., 1956; Gross and Dumsha, 1958) which is suggested to facilitate the crosslinking of the collagenous components (Chandross and Bear, 1979; Kimura and Kubota, 1966; Kimura and Kubota, 1969). Elastoidin first gained interest by researchers for the investigation of collagen-related structures (Chandross and Bear, 1979; Hukins et al., 1976). More recently, the proteins encoded by the *actinodin* (*and*) genes were identified as the non-collagenous, rich-in-tyrosine component of elastoidin (Zhang et al., 2010). In the previous report by Zhang et al. (2010), it was revealed that the *and* genes are likely required for actinotrichia synthesis.

1.6. The *actinodin* gene family

1.6.1. Discovery and expression pattern of actinodin *genes*

The *and* genes were initially identified in a screen for genes differentially expressed during zebrafish caudal fin regeneration (Padhi et al., 2004). In the zebrafish genome, there are four paralogs of *and* (*and1, 2, 3, 4*) whose expression correlates spatially and temporally with the formation of the actinotrichia in both the median fin fold and the pectoral fin bud (Zhang et al., 2010). In the median fin fold at 24 hours post fertilization (hpf), all zebrafish *and* genes are first expressed in epithelial cells with the exception of the distal-most cleft cells corresponding to the AER. Transcripts for *and1* and *and2* were not detected in the AER of the pectoral fin buds, but were present in the remaining epidermal cells of the pectoral fin fold just prior to and during actinotrichia formation starting at 36 hpf. Transcripts of *and* genes were also detected in the

median and pectoral fin fold mesenchyme cells once they have begun to invade the fold. Immunofluorescence from an Actinodin-binding antibody directly overlaps with the actinotrichia fibrils in adult fins (Zhang et al., 2010), and more recently this pattern has been observed in the embryonic actinotrichia (unpublished results from JZ). The gene and protein expression patterns indicated a correlation between *and* genes and the actinotrichia. Additionally, prolonged staining following *in situ* hybridization revealed additional expression domains of *and1* and *and2* in the brain, heart, and branchial arches (Zhang et al., 2010). The potential functions of *and* genes in these additional tissues remain unclear.

1.6.2. Characteristics of the Actinodin proteins

All zebrafish *and* genes encode proteins with a leading signal peptide suggesting they are secreted (Fig 1.4). Orthologs of *and* genes in other fish species encode a conserved region of approximately 60-amino acids near the N-terminus, though the function of this domain is unknown. And1 and And2 proteins possess respectively eight and ten repeated 9-amino acid motifs suggested to form loops in an elongated domain (Andrade et al., 2001; Zhang et al., 2010). Furthermore, in And1 and And2 there are two occurrences of a consensus sequence, RXRR, where X is any amino acid, which are conserved recognition sites for cleavage reactions catalyzed by convertase enzymes in the secretory pathway (Fig 1.4) (Seidah and Chrétien, 1999). And3 and And4 proteins each encode one putative convertase recognition site and three and four repeated 9-amino acid motifs, respectively (Fig 1.4). Furthermore, the average tyrosine content of the zebrafish And proteins is 6.2% (ranging from 8.2% in And1 to 5.2% in And4). This is consistent with previous studies that showed that the non-collagenous constituent of ceratotrichia has more than 6% tyrosine content (Gross and Dumsha, 1958; Zhang et al., 2010).

Fig 1.4. Characteristics of the zebrafish Actinodin polypeptide sequences. Pairwise alignment of And1 and And2 (above) and And3 and And4 (below). Conserved amino acids are coloured in red. Signal peptide (blue), convertase cleavage sites (green) and repeated motifs (yellow) are highlighted for each sequence. Exon boundaries are indicated by blue lines. This figure is adapted from (Zhang et al., 2010).

1.6.3. Functions of actinodin1 *and* actinodin2

Zhang et al. (2010) showed that morpholino-mediated knockdown of either *and1* or *and2* did not cause any noticeable phenotypes. However, co-injection of both *and1* and *and2* morpholinos resulted in a variety of defects of the fin folds. Firstly, actinotrichia are absent from the developing fin folds of double-morphant embryos. The median and pectoral fin folds were poorly developed illustrating the supporting function of the actinotrichia (Zhang et al., 2010). These results suggested the requirement of And proteins for proper structure and growth of the fin fold. Furthermore, mesenchymal cell migration into the fold was distorted in the median fins, or absent in the pectoral fins, of the double-morphants. These mesenchymal cells include the progenitors for lepidotrichia formation. It is possible that without actinotrichia, the mesenchymal cells do not properly migrate into the fold, and in turn cannot properly develop into the fin rays (Thorogood and Wood, 1987; Zhang et al., 2010).

Using *in situ* hybridization, Zhang et al. (2010) analyzed the expression pattern of genes that have been previously implicated in the evolution of limbs. In comparison with wild type fin buds, the expression pattern of *hoxd13a* and *shha* expanded to the anterior of the bud in *and1/and2* double-morphants. An expansion of the expression pattern of these genes during limb development is associated with patterning of the digits. Mouse mutants for *Gli3* similarly exhibit an expansion of *Shh* and *Hoxd13* (Büscher et al., 1997). Mouse and human mutants for *Gli3* typically display polydactyly and similar digits (Hui and Joyner, 1993). Interestingly, *and* genes are absent from the genomes of tetrapods, although they are present in genomes of extant fish species including teleosts and the phylogenetically basal elephant shark, *Callorhinchus milii*. Zhang et al. (2010) propose that the loss of *and* genes in an ancestral tetrapod species may have been concurrent with two novel evolutionary traits associated with limbs. Firstly, the

mesenchymal precursors for lepidotrichia formation do not properly invade the fin fold in *and1/and2* morphants. The loss of *and* genes from a tetrapod ancestor may have contributed to disrupted fin fold development, and in effect the loss of fin rays from limbed species (Zhang et al., 2010). Secondly, *and1/and2* morphant gene expression in the fin bud resembles the polydactylous *Gli3* mutants. Zhang et al. (2010) propose that the loss of actinodin genes in an ancestral tetrapod species may have contributed to gene expression patterns characteristic of a polydactyl limb. Consistently, the fossil record shows that the earliest primitive aquatic tetrapods were polydactylous (Coates and Clack, 1990; Long and Gordon, 2004; Zhang et al., 2010).

1.7. Fin-to-limb evolution

1.7.1. Phylogenetic distribution of vertebrate fins and limbs

Figure 1.2 shows the phylogenetic distribution of major vertebrate groups. According to the fossil record, the median fins evolved at the base of the vertebrate lineage (Zhang and Hou, 2004) during the Cambrian over 500 mya. Paired appendages are specific to the jawed vertebrates, or the gnathostomes. The two major subdivisions of the gnathostomes are the chondrichthyes, also known as the cartilaginous fish, and the osteichthyes, also known as the bony vertebrates (Fig 1.2). One significant evolutionary novelty of the osteichthyes is the presence of endoskeleton tissue (Lecointre and Guyader, 2006). A major clade within the osteichthyes is the actinopterygii, or ray-finned fish, which includes the teleost fish. Another major osteichthyes clade is the sarcopterygians which includes the lobe-finned fish and the tetrapods. Paired fins were characteristic to the ancestral sarcopterygian species. Paired limbs evolved from the ancestral sarcopterygian paired fins specifically in the tetrapod lineage.

And genes have previously been identified in the genomes of chondrichthyes fish and actinopterygian fish (Zhang et al., 2010). The *and* genes are evidently present only in the genomes of species with lepidotrichia, actinotrichia, or ceratotrichia types of fin rays. It can be hypothesized that *and* genes were present in the genome of the common ancestor to all gnathostome species, predating the divergence of chondrichthyes and osteichthyes. Tetrapod genomes do not have *and* genes, indicating that *and* genes were lost at some point during sarcopterygian evolution. The notion of whether the aquatic sarcopterygians possess *and* genes in their genomes will be addressed later in this study.

1.7.2. Effects of developmental and morphological changes in fin-to-limb evolution

Thorogood (1991) proposed that the morphological changes observed during the evolution of fins into limbs can be partially explained by the timing of the transformation of the AER into a fin fold during development (Thorogood, 1991). In teleosts, the AER is initially active and is suggested to participate in the specification of the proximal endoskeleton elements for the fin. Early in fin development, the AER transforms into the fold, resulting in the termination of endoskeleton outgrowth and the initiation of dermal skeleton development of the fin rays. As mentioned previously, the proximal limitation of the fin endoskeleton is typical of teleosts (Grandel and Schulte-Merker, 1998). The proximal limitation of the endoskeleton was similarly observed by Saunders and Rubin (1948) when prematurely removing the AER of the chick limb (Rubin and Saunders, 1972). Based on Thorogood's model, the AER is active for a long duration in the lobed fins of sarcopterygian fish resulting in an elongated endoskeleton. Later during lobed fin development, the AER transforms into a fin fold resulting in fin rays at the apex (Thorogood, 1991). In tetrapods, the AER does not transform into a fin fold and

endoskeleton growth proceeds throughout digit development (Barrow; Fernandez-Teran and Ros, 2008). Thorogood's theory was experimentally tested by Yano et al. (2012) who showed that prolonged AER activity in the zebrafish pectoral fin results in delayed fin fold formation and endoskeleton expansion (Yano et al., 2012).

As mentioned above, different patterns of *hox* gene expression likely account for different endoskeleton morphologies in fins and limbs. Freitas et al. (2012) addressed this hypothesis by overexpressing *hoxd13a* in the zebrafish pectoral fin (Freitas et al., 2012). In limbs, the distal expression of *Hoxd13* is associated with digit formation. Overexpression of *hoxd13a* in the fin bud resulted in enhanced mesenchyme proliferation and endoskeleton outgrowth (Freitas et al., 2012). Furthermore, the fin fold was diminished and *and1* genes were downregulated following *hoxd13a* overexpression. Both experiments by Freitas et al. (2012) and Yano et al. (2012) resulted with endoskeleton expansion and a diminished fin fold. However these experiments establish two opposite interpretations. On one hand, the loss of the fin fold and prolonged AER activity may have caused extended endoskeleton growth in the ancestral tetrapod (Yano et al., 2012). On the other hand, prolonged *hox* expression (extended endoskeleton growth) in an ancestral tetrapod species may have caused the reduction of the fin fold (Freitas et al., 2012; Schneider et al., 2011). Taken together, it is not clear whether the loss of the fin fold is a cause or an effect of extended endoskeleton growth. In any case, results from the experiments performed by Zhang et al. (2010), Yano et al. (2012) and Freitas et al. (2012) suggest that the reduction of the fin fold may have been a permissive step in the direction of fin to limb evolution (Freitas et al., 2012; Yano and Tamura, 2013; Yano et al., 2012; Zhang et al., 2010).

The presence or lack of a fin fold is suggested to be a major determinant of appendage developmental fate. The genetic mechanisms involved in actinotrichia formation appear to play

an important role in fin fold development and growth. The *and* genes have been shown to play a role in this process, although the exact mechanisms of *and* function for actinotrichia synthesis is not clear. The study below focuses on the *and* genes and their contributions to actinotrichia formation and function.

1.8. Summary of background information and objectives

The onset of *and* gene expression occurs in the fin fold epithelium (Zhang et al., 2010). These early stages of *and* expression likely contribute to the initiation of actinotrichia formation and fin fold development. In turn, uncovering the regulatory mechanisms that control *and* gene expression can strengthen our understanding of the mechanisms required for actinotrichia and fin fold formation. Furthermore, it is possible that changes in these regulatory mechanisms may have contributed to the loss of the actinotrichia and fin fold in an ancestral tetrapod species.

Zhang et al. (2010) reported that *and* genes encode the non-collagenous component of the actinotrichia fibrils. It was previously suggested that in ceratotrichia, which are homologous to actinotrichia, the non-collagenous component induces cross-linking of collagens resulting in fibril formation (Chandross and Bear, 1979; Kimura and Kubota, 1966; Kimura and Kubota, 1969). However, the function of the non-collagenous component has not been explored in detail. Zhang et al. (2010) documented various motifs inherent to the proteins encoded by the *and* genes, for example the presence of putative cleavage sites that may signal for the fragmentation of the protein (Fig 1.4). Preliminary data has shown that there are similarities between the And protein amino acid sequence and the amino acid sequence of another protein termed Esophageal cancer related gene 4 (Ecrg4) (M. Andrade-Navarro, unpublished data). Secretion and fragmentation of the Ecrg4 protein is required for its function (Dang et al., 2012; Ozawa et al.,

2011). Due to the sequence similarities, it is possible that secretion and fragmentation of the And protein is required for its function towards actinotrichia synthesis and fin fold development.

The loss of *and* genes from the genomes of tetrapods has been implicated in the evolutionary transition of fins into limbs. Aquatic sarcopterygians, including the coelacanth and lungfish, possess fins with rays. Recently an ortholog of the *and* gene was documented in the genome of the coelacanth (*Latimeria chalumnae*) (Amemiya et al., 2013). The presence of an *and* ortholog in the coelacanth genome is consistent with the hypothesis that *and* genes were lost specifically from tetrapods. A better understanding of the developmental mechanisms and evolutionary patterns related to the *and* genes can provide insight towards the events that lead to the loss of *and* genes during the evolution of fins into limbs.

Objectives 1 and 2 of this study aim to characterize molecular functions of the *and* genes in regard to actinotrichia and fin fold development. Objective 3 aims to identify the evolutionary pattern of *and* gene loss from the tetrapod lineage. The objectives of this study are:

1) **a.** Characterize the *cis*-acting regulatory elements that promote *and1* gene expression in the fin fold epithelium.

b. Determine potential transcription factors that are capable of binding the DNA sequence of the *and1 cis*-acting regulatory elements.

2) **a.** Analyze and predict possible functions of And proteins based on motifs in the polypeptide sequence.

b. Detect fragmented And1 protein molecules *in vivo* using a flag-tag epitope fused to the And1 N-terminal domain.

3) **a.** Identify and characterize the *and* genes in the coelacanth genome.

b. Determine the pattern of conserved synteny of the *and* genes across vertebrate species.

The function of *and* genes are important for actinotrichia synthesis and likely contribute to fin fold development. Understanding these genes and how their functions have potentially been altered during evolution may elucidate the events that caused the loss of a fin fold during fin-to-limb evolution.

Chapter 2 – Materials and Methods

2.1. *In silico* analyses

A variety of online tools were used for obtaining and analyzing the sequences of DNA and proteins. Zebrafish genes and polypeptide sequences were obtained from the zebrafish model organism database (zfin.org) (Bradford et al., 2011). Zebrafish *and1-4* gene and polypeptide sequences were obtained from zfin or from the manuscript by Zhang et al. (Zhang et al., 2010). All other genes and polypeptide sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/index.html). Pairwise sequence alignments were performed with Tcoffee (http://tcoffee.crg.cat/apps/tcoffee/do:regular) (Notredame et al., 2000) or Emboss (http://www.ebi.ac.uk/Tools/psa/) (EMBL-EBI, Wellcome Trust Genome) where indicated. The *and1* genomic regions, or syntenic genomic regions including the genes *otos* and *myeov2*, of different vertebrates were extracted with the Ensembl genome browser (coelacanth: LatCha1-_dna range = scaffold JH126651.1:1-77734, human: GRCh37_dna range=chr2: 241046074-241093070, chicken: Galgal4_dna range=chr9: 4709937-4747226, zebrafish: Zv9_dna range=chr24: 27032525-27110853, *Xenopus*: JGI_4.2_dna range=scaffold GL172691.1: 2085511-2121310, rat: Rnor_5.0_dna range=chr 9: 99459831-99489582). Annotations of genomic regions were derived from Ensembl and repetitive DNA was masked with repeatmasker (http://www.repeatmasker.org/). Genomic regions were aligned and visualized with the program mVISTA (http://genome.lbl.gov/vista/customAlignment.shtml) (Frazer et al., 2004). Conserved synteny was identified with the Ensembl "region comparison" tool. Large-scale synteny was observed for the Acanthomorph *and3* loci using the Synteny Database (http://teleost.cs.uoregon.edu/synteny_db/) (Catchen et al., 2009). Proprotein convertase

cleavage sites were predicted using Neuropred (http://neuroproteomics.scs.illinois.edu/cgi-bin/neuropred.py) (Southey et al., 2006), and polypeptide molecular weights were predicted using Protein Prospector (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct). The coelacanth 'small' *and* genomic sequence was translated using the Expasy translate tool (http://web.expasy.org/translate/). Transcription factor binding sites were predicted with TRANSFAC (http://www.biobase-international.com/product/transcription-factor-binding-sites) using the vertebrates-only profile and the cut-offs were set to minimize false positives.

2.2. Subcloning of fin-epithelial enhancers

All subcloning was performed following standard procedures (John J. Sambrook, 2001). The 1947bp or the 876bp genomic DNA sequences directly upstream the zebrafish *and1* gene were previously isolated by members of our lab. Additionally in our lab, two constructs were previously made with either the *1947and1* or the *876and1* elements replacing the CMV promoter in the pEGFP-N1 plasmid (CLONTECH Laboratories, Inc., Mountain View, CA, USA). Subfragments were amplified by PCR from the *1947and1:pEGFP-N1* clone DNA using specific oligonucleotides (Table 2.1). PCR products were cloned into linearized 5kb plasmid vectors *876and1:pEGFP-N1*.

Table 2.1. Oligonucleotides for subcloning enhancer fragments for transient transgenic analysis.

oligonucleotide name	alternate name	sequence (5'-3')	restriction site underlined
1279and1 Forward (FW)	1.5Pand1 FW	GATCTGCAGTGACGA AGACGACAAA	*PstI*
1199and1 FW	1.4Pand1 FW	CTCGAGGTCTTAACT GTCAAGTGTGTTAAT	*Xho*I
1117and1 FW	1.35Pand1 FW	CTCGAGGAAACCCCA GACATTCCTACTC	*Xho*I
1063and1 FW	1.3Pand1 FW	CTCGAGCTAGGCTGA AATCTGTTTCGAGT	*Xho*I
1036and1 FW	1.25Pand1 FW	CGACTGCAGATTGTA GGCTTTCG	*Pst*I
1383and1 FW	1.6Pand1 FW	CTCGAGGATTTGACG GAGAACAGCACATC	*Xho*I
1176and1 Reverse (REV)	1.4Pand1 REV	GAATTCATTAACACA CTTGACAGTTAAGAC	*Eco*RI
1297and1 FW	FE+Pand1 FW	GATCTCGAGGATTTA TTCGATG	*Xho*I
1030and1 REV	FE+Pand1 REV	TGCGAATTCCTACAA TCAGC	*Eco*RI
932and1 REV	1.1Pand1 REV	GAATTCTGTCAATGTC AGAGTGTAAAAC	*Eco*RI
971and1 REV	1.2Pand1 REV	GAATTCACGGCCAAT CGCATGTGTA	*Eco*RI

2.3. Subcloning of cDNA with *flag-tag* sequence

Fragments of the *and1*cDNA were amplified by PCR from the *and1cDNA:pDrive* plasmid using specific oligonucleotides. The first 171bp of the *and1* cDNA was amplified by PCR (PCR#1 (Fig 2.1)) using the M13 forward primer and the DMOVPFLAG Reverse primer which has the Flag-tag (DYKDDDDK) DNA sequence (GACTACAAGGACGACGATGACAAG) at its 5' end (Table 2.2). The resulting PCR#1

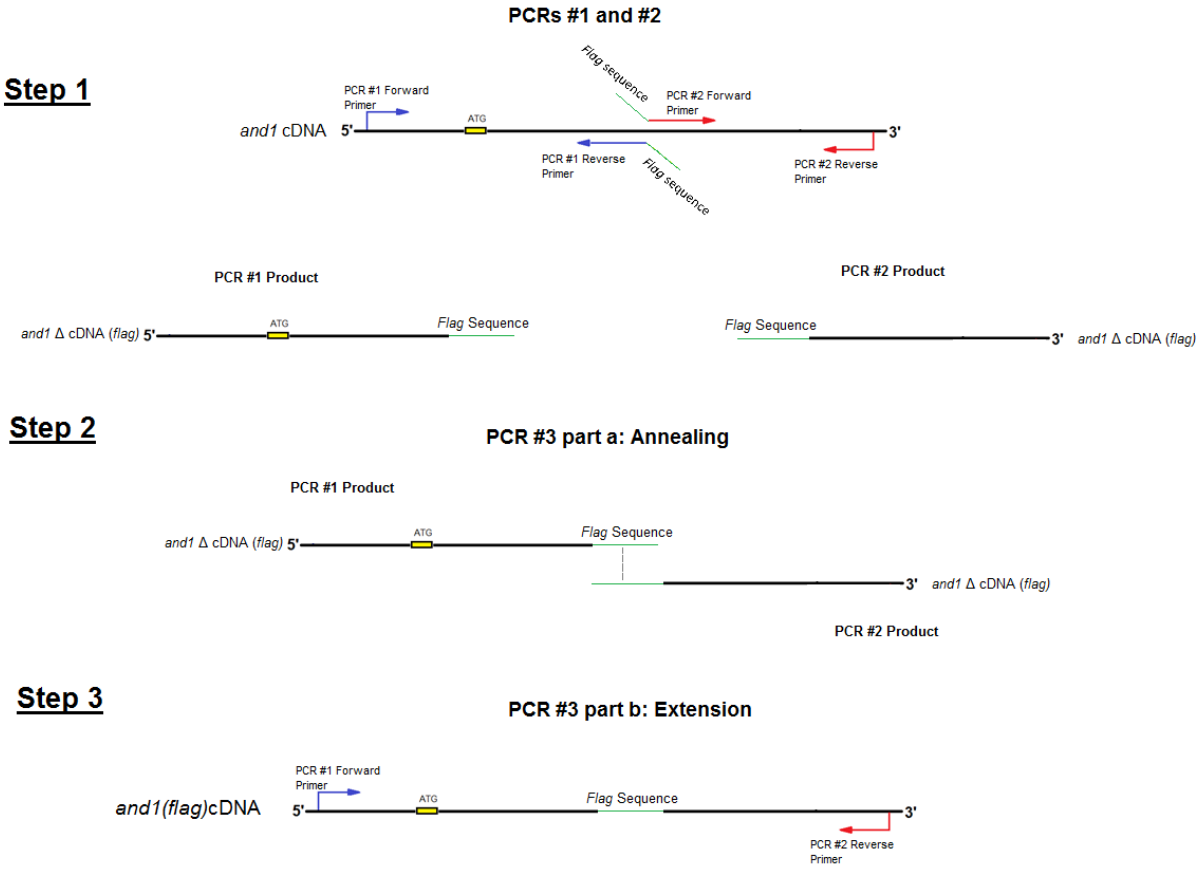


Fig 2.1. Overlapping PCR for site-directed mutagenesis. Insertion of the *flag* tag DNA sequence (GACTACAAGGACGACGATGACAAG) into *and1* cDNA by overlapping PCR. PCR #1 produces the 5' region of the *and1* cDNA with the *flag* flanking its 3' end. A separate PCR reaction, PCR # 2 produces the 3' region of the *and1* cDNA with the *flag* flanking its 5' end. The first step in PCR #3 anneals the complimentary *flag* sequences from the earlier PCR products. Using the forward primer from PCR #1 and the reverse primer from PCR #2, *and1(flag)*cDNA can be amplified.

Table 2.2. Oligonucleotides for overlapping PCR (insertion of flag tag sequence into *and1* cDNA).

oligonucleotide name	overlapping PCR (Fig 2.1) name	sequence (5'-3')
M 13 Fw	PCR#1 Forward	GTAAAACGACGGCCAG
DM OVP FLAG Rev	PCR#1 Reverse	CTTGTCATCGTCGTCCTT GTAGTCTCTCCTGTTGC GGATCAGTTTCTT*
DM OVP FLAG FW	PCR#2 Forward	GACTACAAGGACGACG ATGACAAGAACATCAGT TGGTACAAGCAGCAC*
and1 EcoO109I Rev	PCR#2 Reverse	TAGTATTGACCAGAACG TGGGTCCT

*****Flag sequence underlined

product consists of the first 171bp of *and1* cDNA plus the flag DNA sequence in frame at its 3' end (Fig 2.1). The nucleotide positions 172-644 of the *and1* cDNA were amplified by PCR (PCR#2 (Fig 2.1)) using the DMOVPFLAG Forward primer, which has the Flag sequence at its 5' end, and the and1EcoO109I Reverse primer (Table 2.2). The resulting PCR#2 product consists of *and1* cDNA nucleotides 172-644 with the Flag DNA sequence in frame at the 5'end (Fig 2.1). The PCR#1 and PCR#2 products were annealed, or "overlapped", at their complimentary Flag sequences. The annealed PCR#1 and PCR#2 strands were extended by PCR using the M13 forward primer and the and1EcoO109I Reverse primer (Fig 2.1). The resultant PCR product consists of the *and1* cDNA nucleotides 1-644 with the Flag sequence inserted in between the nucleotides 171 and 172, termed *and1(flag)1-644*. The nucleotide position 171 corresponds to amino acid residue 57 which is the last residue of the putative convertase cleavage site RNRR. Therefore, the flag sequence immediately follows the putative cleavage site. The *and1(flag)1-644* clone was ligated into the pDRIVE (QIAGEN Inc., Toronto, ON, Canada) vector. The

and1(flag) 1-644 was subsequently cloned into the full length *and1* cDNA using the restriction enzyme *EcoO109I* and screened for the correct orientation. The full length *and1(flag)* cDNA was cloned into the pCS2+ vector using *XhoI* and *SnaBI* restriction enzymes.

Plasmids containing cDNA of mouse *ecrg4(C-terminal Flag)* (*ecrg4(C-flag)*) or *ecrg4(N-terminal Flag)*(*ecrg4(N-flag)*) in the pCDNA3.1 vector were kind gifts from Lukasz Piszczek (Dr. Cornelius Gross Lab, EMBL, Monterotondo, Italy). *Ecrg4(C-flag)* and *ecrg4(N-flag)* were separately cloned into the pCS2+ vector using *XhoI* and *BamHI* restriction enzymes. Sequences of all clones were verified with DNA sequencing.

2.4. Animal care

All fish are bred and raised in the D' Iorio, University of Ottawa, zebrafish facility. The fish facility is maintained at 28 °C, with a photo-period of 14 h of light and 10 h of darkness. Fish are fed and tanks are cleaned regularly (Westerfield, 2000). Animal care and experiments were performed according to the CCAC guidelines.

2.5. Microinjection of fin epithelium enhancer DNA

Injection mixes were prepared just prior to injection. The injection mix consists of DNA diluted to a final concentration of 100 nanograms per microliter (ng/µl), Phenol Red at 0.5%, and the final volume adjusted with DEPC treated water. One cell-stage embryos were collected and soaked in bleach for 2 minutes followed by rinsing with zebrafish aquatic system water. Embryos were aligned on an agarose injection plate. Injections were performed using the NARISHIGE IM-300 microinjector, with a pressure of 100 psi and for 30 msec/injection. DNA was injected directly into the cells at the one- to two-cell stage.

2.6. Microinjection of RNA

The *and1* cDNA in the pCS2+ vector was transcribed into mRNA *in vitro* using the mMESSAGE mMACHINE (Ambion, Life Technologies Inc., Burlington, ON, Canada) kit. Aliquots of RNA were stored at -80 degrees Celsius (°C). RNA injection mixes were prepared immediately before microinjections and consisted of RNA diluted to 100ng/µl, Phenol red at 0.5%, potassium chloride (KCl) at 0.02 Molar (M), and final volume adjusted with DEPC water. The RNA injection mix was kept on ice during preparation and the injection procedures. The procedure for preparing and microinjecting embryos is the same as above. *EGFP* RNA, at RNA concentration of 100ng/µl, was co-injected with *and1(flag)* RNA as a control. Embryos injected with *and1(flag)* RNA which were used for protein extraction and Western blotting were not co-injected with *eGFP*. With the same procedures described above, *ecrg4(C-flag)* and *ecrg4(N-flag)* were transcribed *in vitro,* prepared for microinjections, and microinjected as a control for *and1(flag)* protein analyses (results not shown).

2.7. Screening

Fish injected with putative fin epithelial enhancer DNA fused to *eGFP* were screened at 48 hpf for green fluorescence. A Leica MZ FLIII microscope with an integrated FLUOIII filter system was used to visualize the samples. Photographs of samples were taken with a digital Sony 3CCD Color Video Camera linked with AxioVision microscopy software. Fluorescent pictures of a whole-mount embryo or caudal fin requires an average exposure time of 600ms. All embryos showing at least one fluorescent cell in the fins were photographed. Total numbers of microinjected embryos and total numbers of fluorescent embryos were tallied for statistical

analysis. Embryos injected with *1947and1:eGFP* or *2.5kband1:eGFP* which showed relatively high levels of fluorescence were kept to be raised. Fish injected with *eGFP* RNA were screened for green fluorescence at 2-2.5 hpf (Nusslein-Volhard and Dahm, 2002). A subset of embryos injected with *and1(flag),* mouse *ecrg4(N-flag),* or *eGFP* RNA were monitored for phenotypic effects up to 7 dpf. Embryos injected with *and1(flag)* and mouse *ecrg4(N-flag)* RNA were used for protein extractions.

2.8. Transgenic lines

Embryos injected with *1947and1:eGFP* or *2.5kband1:eGFP* were raised. Between the times of sexual maturity to the time of maximum breeding potential, approximately 3-14 months of age (Westerfield, 2000), fish were bred and their offspring screened at 48hpf for green fluorescence in the fin folds. Fluorescent offspring were raised and monitored as transgenic lines. A fish was considered as negative for transgenesis whenever 300 of their offspring were obtained and no fluorescence was observed.

2.9. Protein extraction

Protein extraction was carried out on batches of embryos microinjected with *and1(flag)* RNA, embryos microinjected with *ecrg4 (N-flag)* RNA, and non-injected wild type embryos. Embryos were collected at desired stages. The embryos were dechorionated and anesthetized in tricaine at 0.02% (w/v) (Westerfield, 2000). The full yolk region was amputated with the bevel of a small syringe as a blade. Embryos were directly homogenized in 2x sodium dodecyl sulfate sample buffer (SB) (125 mM Tris-HCl [pH 7.5], 4% SDS, 20% Glycerol, 10% ß-mercaptoethanol, 0.004% Bromophenol Blue, DEPC water) using the following concentrations:

10 embryos at 5hpf /40µl SB; 20 embryos at 20hpf/80µl SB; 20 embryos at 48hpf/80µl SB. Samples were crushed with a glass pestle, followed by centrifugation at 4400 rpm at 4 °C for 10 minutes. The resultant middle proteinous fraction was isolated from the sample tube and stored at -80 °C. The above steps were all performed over ice.

2.10. Western blotting

15µl of the homogenized protein sample was loaded in the SDS-polyacrylamide gel electrophoresis (PAGE) gel (5% stacking and 10% resolving gel) using the BIO-RAD (BIO-RAD Laboratories Inc., Mississauga, ON, Canada) minigel apparatus. 10µl of broad range prestained SDS-PAGE standard (BIO-RAD Laboratories Inc., Mississauga, ON, Canada) and 10µl of Bench Mark protein ladder (Life Technologies, Burlington, ON, Canada) were loaded for molecular weight determination. The protein content was electrophoretically transferred from gels to nitrocellulose membranes according to the procedures in the Protein Blotting Guide (BIO-RAD Laboratories Inc., Mississauga, ON, Canada). The total protein profile of the embryos was first visualized by staining the membrane with Ponceau S solution. Ponceau S was rinsed off with 1X TBST. Membranes were then blocked in 5% milk (w/v) in 1X TBST for 1h at room temperature. The membrane was then immersed with the primary antibody (anti-Flag, Sigma-**Aldrich** #F7425) diluted at 1:1000 (Objective 2, Fig 4.5) or 1:4000 (Objective 2, Fig 4.6) and incubated overnight at 4 °C, following by washing with 1X TBST. The immunoblots were then incubated at room temperature for 1h with anti-rabbit-HRP IgG (Amersham, GE, Baie d'Urfe, QC, Canada) diluted at 1:10000 in 5%milk in 1X TBST. Blots were developed using Luminata Forte Western HRP substrate (Millipore, Billerica, MA, USA), and visualized on Kodak X-omat LS Film (Sigma-**Aldrich**, St. Louis, MO, USA). Stripping and reprobing

membranes was performed as described in abcam technical protocols (www.abcam.com). Immunoblotting for reprobing was performed as described above, though with alternate primary antibodies including anti-2-F11 (anti-Actinodin) diluted at 1:100 or anti-Actin (Sigma-**Aldrich** #A2066) diluted at 1:1000.

Chapter 3 - Objective 1: Characterization of fin epithelium-specific *cis*-acting regulatory elements of the *and1* gene

3.1. Introduction

3.1.1. Cis-acting regulatory elements in the developing fins and limbs

The timing and location of gene expression is controlled by *cis*-acting regulatory elements. The complex networks that integrate gene function with the control of gene expression are essential for the development of an organism. While the functions of many genes and transcription factors involved in fin and limb development are well characterized, their *cis*-acting regulatory elements have not been thoroughly investigated (Rabinowitz and Vokes, 2012). The formation of digits is one of few morphological examples that have been the focus of several studies in regard to *cis*-acting regulatory elements and their impact on the evolution from fins to limbs (Schneider et al., 2011; Shubin et al., 2009). The genetic regulatory mechanisms that control other important morphological features of fins and limbs, for example the AER or the fin fold, are less clear. A better understanding of *cis*-acting regulatory elements that control aspects of fin and limb development can help distinguish important developmental mechanisms and possible evolutionary changes.

3.1.2. Gene regulation of and1 in the fin bud

In the fin, actinotrichia formation is associated with the emergence of a fin fold, or a turning point in fin versus limb development. The *and* genes have previously been shown as key structural contributors to the actinotrichia (Zhang et al., 2010). The loss of the *and* genes from the genomes of tetrapods implies a cause for the loss of actinotrichia, and is probably related to

the loss of a fin fold in limbed species. The genetic *cis*-acting regulatory mechanism that mediates *and* gene expression can be understood as an important contributor to actinotrichia formation as well as a genetic mechanism that separates fin fold formation from the developmental program of a tetrapod limb.

Preliminary work from our lab uncovered a *cis*-acting regulatory element located immediately upstream of the *and1* gene in zebrafish. The element is contained in a 1947 basepair (bp) genomic sequence that was cloned into an enhanced Green fluorescent protein (eGFP) reporter construct. The *1947and1*:*eGFP* construct was sufficient for driving tissue-specific eGFP expression in the epithelium of the zebrafish pectoral fin and median fin folds as early as 30 hpf and persists throughout early stages of development (Fig 3.1). The eGFP expression induced by the *1947and1* fragment recapitulated *and1* gene expression in the epithelium of the pectoral and median fins, and like *and1*, eGFP was excluded from the AER and cleft cells (Zhang et al., 2010).

3.1.3. Fin epithelium-specific gene expression during development

Previous reports have identified the ectoderm as the first tissue in the fin bud that contributes to actinotrichia formation and the first tissue that expresses *and1* in the fin (Durán et al., 2011; Zhang et al., 2010). Due to its early epithelium-specific expression, the *1947and1* element likely contributes to the onset of *and* gene expression during fin fold morphogenesis. While enhancers can be up to hundreds of bps in length, shorter DNA sequences, of approximately 10-20 bps within the enhancer, are recognized by the DNA-binding transcription factors that will promote gene expression. In the section below, it is shown that two short elements of 54bps and 61bps embedded in the *1947and1* element can induce significant

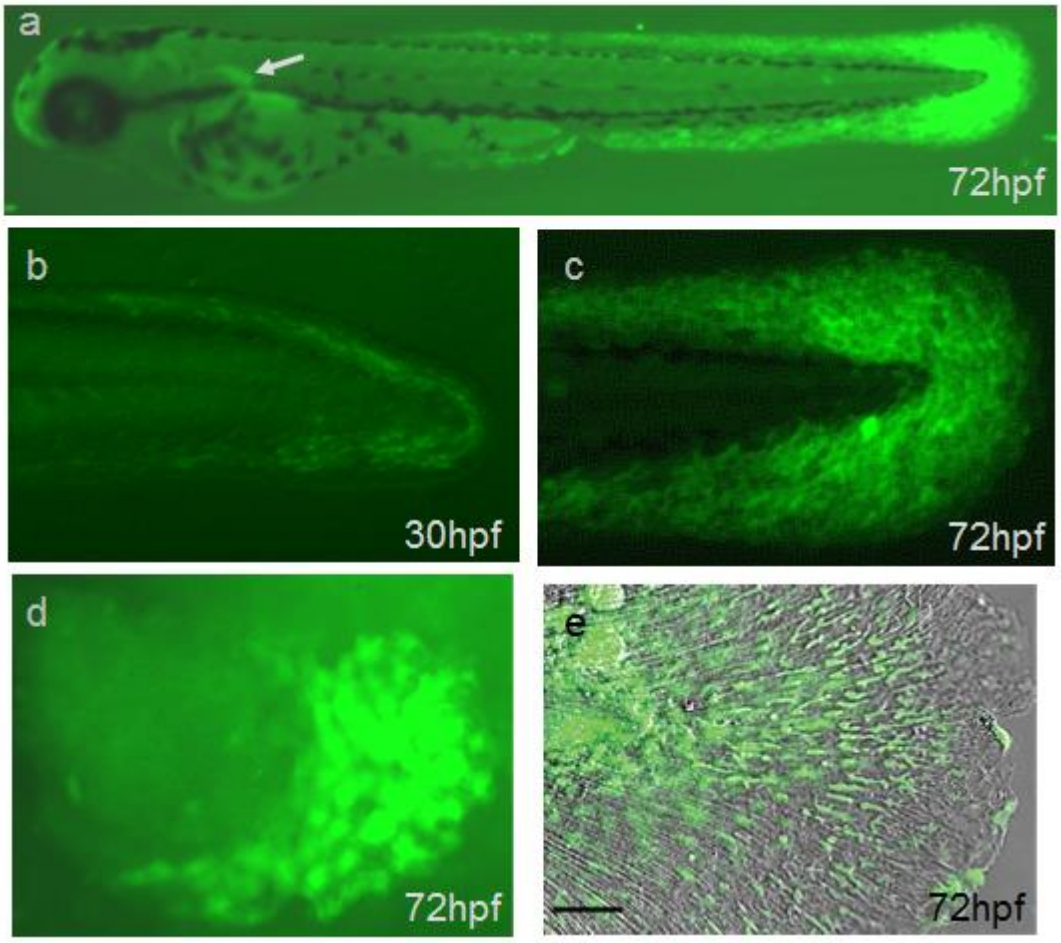


Fig 3.1. The zebrafish transgenic line for *1947and1:eGFP* expresses fin epithelium-specific eGFP. a, Full body of 72hpf *1947and1:eGFP* transgenic embryo expressing eGFP in the pectoral fin fold (arrow) and the median fin fold (far right). **b,** Detail of median fin fold at 30hpf showing the onset of transgenic eGFP expression. eGFP expression is only in epithelial cells as mesenchyme cells are yet to migrate into the fin fold. **c, d,** Close-up of median (c) and pectoral (d) fin folds at 72hpf with expression of eGFP in the circular-shaped epithelium cells. **e,** Median fin fold detail of ET37 transgenic embryo with eGFP expression specifically in the elongated mesenchymal cells invading the fin fold at 72hpf (adapted from Zhang et al, 2010). Note the different morphology of the mesenchyme cells in e and the epithelium cells expressing eGFP in a-d.

fin epithelium-specific reporter gene expression upon transient transgenic analysis in zebrafish embryos. Consensus DNA sequences in the enhancer elements are identified as potential binding sites for transcription factors such as Dlx3, Pmx1, Cdx2 and Hoxa13. These factors have reported functions in fin or limb development and can provide insight towards the genetic pathways that govern fin fold formation through the control of *and1*.

3.2. Results

3.2.1. Identification of the and1 *promoter and a fin epithelium-specific enhancer*

A genomic sequence of 1947bp was previously cloned from the region directly upstream the first exon of *and1* (Ensembl gene ID: ENSDARG00000088514). Although different transcripts have been recorded for *and1* (variations between genome databases Ensembl.org and Zfin.org), the *and1* gene transcript and exons referred to in this study are based on those documented by Zhang et al. (Appendix 3A) (Zhang et al., 2010). A zebrafish transgenic line was produced that stably expresses *1947and1*:*eGFP*, and displays eGFP in the epithelium of the developing fins (Fig 3.1). Previously in our lab, it was shown that the 3'-most 876bp within the *1947and1* element cannot in isolation drive eGFP expression in zebrafish embryos (Fig 3.2). However, this 876bp region is located within the first kilobasepairs (kb) directly upstream the *and1* first exon, and likely contains the *and1* minimal promoter. Consistently, this 876bp sequence possesses a promoter sequence as identified with the online tool, Eukaryotic promoter prediction (http://www.fruitfly.org/seq_tools/promoter.html), as well as exhibits sequence conservation with the upstream region of zebrafish *and2* (Ensembl gene ID: ENSDARG00000079302). Thus, the fin-epithelium enhancer element is located in the 1071bps between the 5' boundaries of the *1947and1* element and the 876bp promoter region (Fig 3.2).

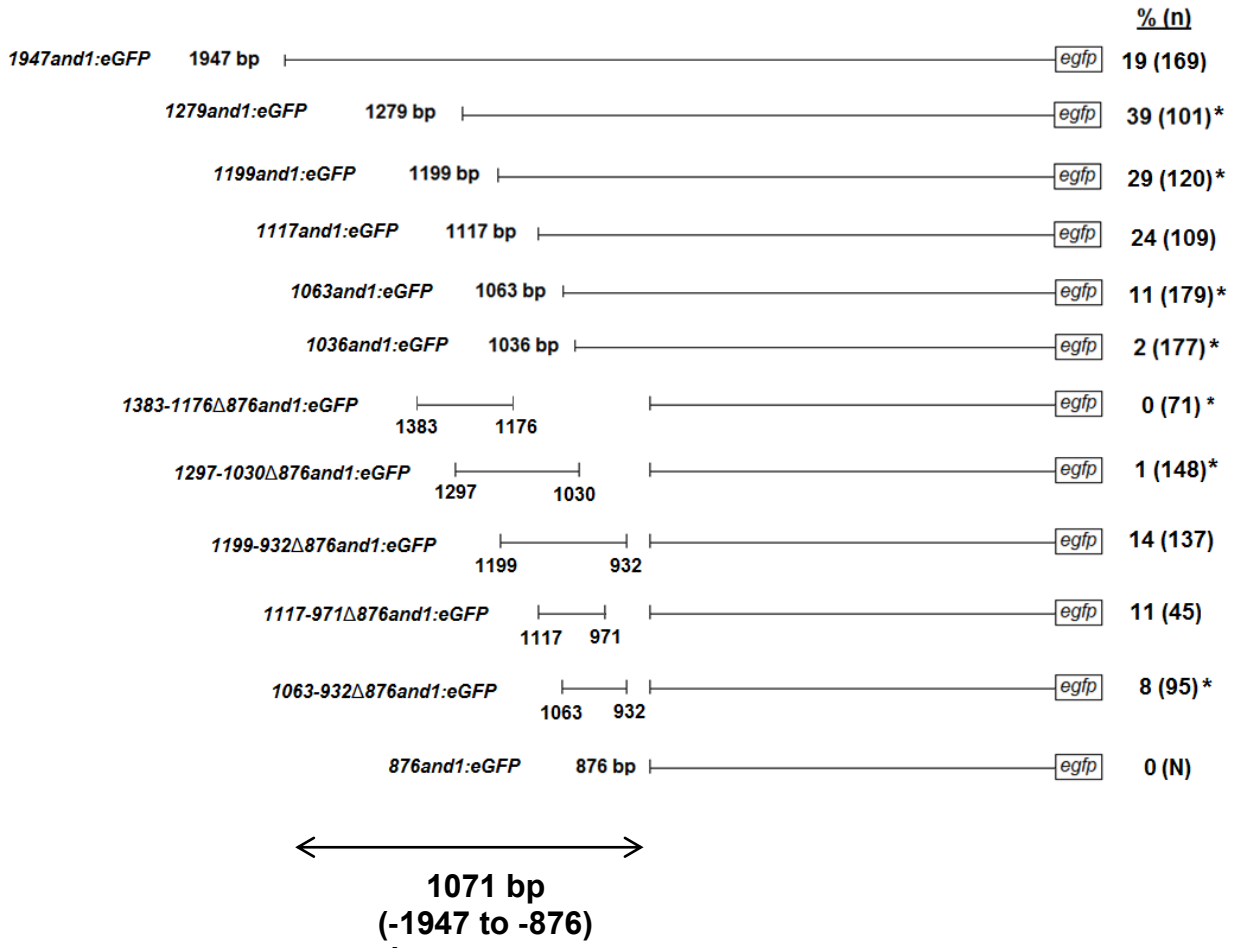


Fig 3.2. Schematic representation of the subfragments cloned from *1947and1* and fused to an *eGFP* reporter gene. The title of each clone is listed on the far left side of each schematic. The numbers represent the basepair (bp) distance upstream from the *and1* first exon. On the right side, the number in parentheses indicates the total number of microinjected embryos observed per respective construct (indicated as n). The number to the left of the parenthesis indicates the percentage (%) of total observed embryos that displayed eGFP expression in the fin folds at 48hpf. Asterisks indicate whether the number of eGFP-expressing embryos was significantly different than the *1947and1:eGFP* construct. The putative enhancer elements are located in a 1071bp region indicated at the bottom.

Different subfragments of the 1071bp sequence were cloned and inserted upstream the 876bp minimal promoter and subsequently tested for enhancer capabilities[1].

3.2.2. Deletion fine mapping of enhancer sequences and transient transgenic analysis

Figure 3.2 shows the DNA subfragments of the *1947and1* element that were cloned to *eGFP* and subsequently microinjected at 100ng/µl into 1-cell stage wild type zebrafish embryos. The transient transgenic expression of eGFP in the fin fold was observed in embryos that were directly microinjected. Observations were recorded at 48 hpf following microinjections because at this time point strong expression has been observed for *and1* transcripts (Zhang et al., 2010) and for eGFP in the *1947and1*:*eGFP* transgenic line. Transient transgenic analysis was used because it is a much faster method than producing transgenic lines for each construct. However, there are drawbacks when using a transient transgenic analysis in comparison to analyzing transgenic lines expressing the same reporter construct. For example, in embryos microinjected with *1947and1*:*eGFP* , the average number of cells transiently expressing eGFP was much lower than the relatively ubiquitous eGFP expression in the fin epithelium of *1947and1*:*eGFP* transgenic lines (Figs 3.1, 3.3). Furthermore, only 19%, or 32 embryos of *1947and1*:*eGFP*-microinjected embryos (n=169) displayed any eGFP expression, whereas all offspring of a

[1] Various members of our lab contributed to the subcloning of the *1947and1* subfragments, including N.C., H.B., L.G., W.L., and myself. My personal contributions to this project included: Subcloning of the fragments *1199-932Δ876and1:eGFP, 1383-1176 Δ876and1:eGFP,* designing constructs and primers for *1117-971 Δ876and1:eGFP, 1063-932 Δ876and1:eGFP, 1199and1eGFP, 1117and1eGFP, 1063and1eGFP,* as well as microinjecting all constructs shown (Figs 3.2, 3.3, 3.4), compiling data for the expression profile of each construct (Figs 3.2, 3.3, 3.4), and statistical analyses.

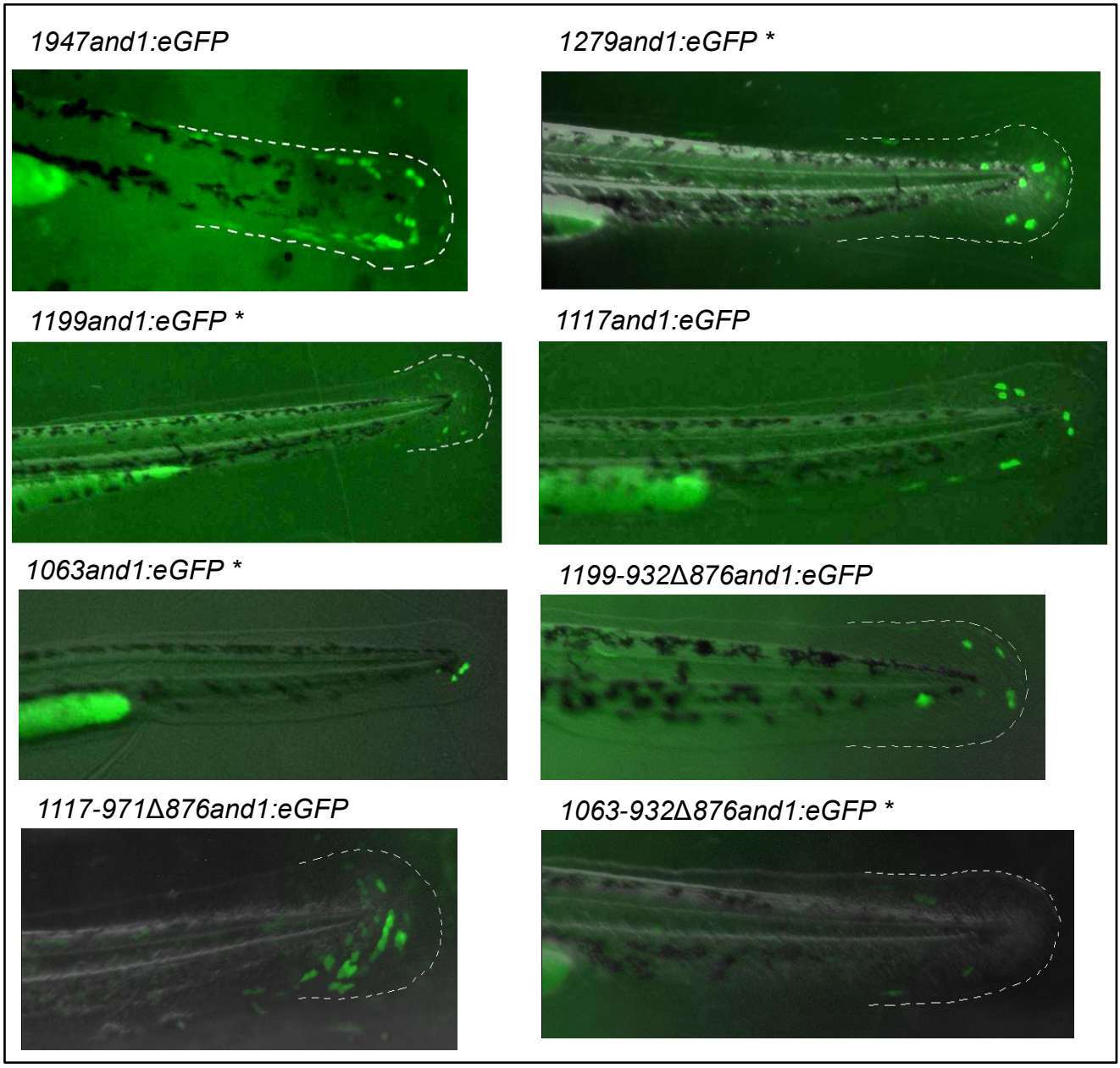


Fig 3.3. Transient transgenic eGFP expression in the median fin fold at 48 hpf in zebrafish embryos microinjected with different subfragments of the *1947and1* enhancer element.

Each image was taken of an embryo microinjected with the construct titled above the image. The expression pattern varied considerably between different constructs and between embryos injected with the same construct. The images represent the average or typical expression pattern observed for each construct. An asterisk beside the construct name indicates a significantly

different number of eGFP-expressing embryos when compared to *1947and1:eGFP*. Dotted lines encompass the distal border of the median fin fold.

transgenic line theoretically exhibit transgene expression (Nusslein-Volhard and Dahm, 2002) showing that the efficiency of gene expression controlled by *1947and1* is greatly reduced when microinjected (Figs 3.2, 3.3).

3.2.3. Statistical approach to transient transgenic analysis

When testing the constructs shown in figure 3.2, it is expected that deleting DNA from the *1947and1* element potentially deletes important enhancer elements, and eGFP expression in the fin epithelium would be lost. Some constructs were capable of driving very low eGFP expression when compared to the transient expression induced by *1947and1*, but expression was not completely absent (Figs 3.2, 3.3). Low eGFP expression was defined as having relatively few numbers of cells expressing eGFP (Appendix 3B), and a lower percentage of microinjected embryos that express eGFP, in comparison to *1947and1*:*eGFP*-microinjected embryos. This low eGFP expression may have been caused by the deletion of important enhancer elements, although, as mentioned above, low eGFP may also be the effect of low reporter expression efficiency when using transient transgenic analysis. To determine whether the low eGFP was a result of the loss of enhancer elements or due to the low efficiency of transient gene expression, a statistical analysis was used to test each construct in comparison to the *1947and1* element. The number of embryos positively expressing eGFP was scored from the total number of embryos microinjected per construct. As observed for transient expression of *1947and1:eGFP*, 32 eGFP-expressing embryos out of 169 microinjected embryos (32/169) is the expected ratio for constructs retaining all enhancer elements. Using a chi-squared test, it was determined whether each construct drives a similar expression profile to the expected value derived from

1947and1:eGFP , or whether the number of eGFP expressing embryos was significantly different (Appendix 3C).

3.2.4. Two short sequences may be sufficient for fin epithelium-specific reporter expression

The construct *1117and1:eGFP* does not drive significantly different reporter expression from the *1947and1:eGFP* clone indicating that this fragment likely retains the enhancer elements (Fig 3.4). A deletion of DNA to change *1117and1:eGFP* into *1063and1:eGFP* is the smallest observed deletion that veers a non-significant construct into a significant construct. Thus, it is concluded that important enhancer elements are lost from the *1063and1:eGFP* construct. The DNA deleted in this example is located from positions -1064 to -1117 upstream the *and1* first exon and this fragment likely contains important enhancer elements (Fig 3.4). Two other constructs that do not induce significantly different expression from *1947and1:eGFP* are *1199-932Δ876and1:eGFP* and *1117-971Δ8766and1:eGFP* (Fig 3.4). Expectedly, both of these constructs contain the important region mentioned above, -1064 to -1117. However, the construct *1297-1030 Δ876and1:eGFP* gives significantly lower GFP expression despite that it also contains the fragment -1064 to -1117 (Fig 3.4). Therefore in addition to -1064 to -1117, there must be a second important element located within the three non-significant constructs (*1117and1:eGFP, 1199-932Δ876and1:eGFP, 1117-971Δ876and1:eGFP*) but not in *1297-1030 Δ876and1:eGFP* which is significantly different. A deletion of DNA to change *1117-971Δ876and1:eGFP* into *1297-1030 Δ876and1:eGFP* is the smallest observed deletion that veers a non-significant construct into a significant construct while retaining the -1064 to -1117 region. The DNA deleted in this example is located from basepairs -971 to -1031 upstream the *and1* first exon (Fig 3.4). Overall, these results support that the expression pattern observed in

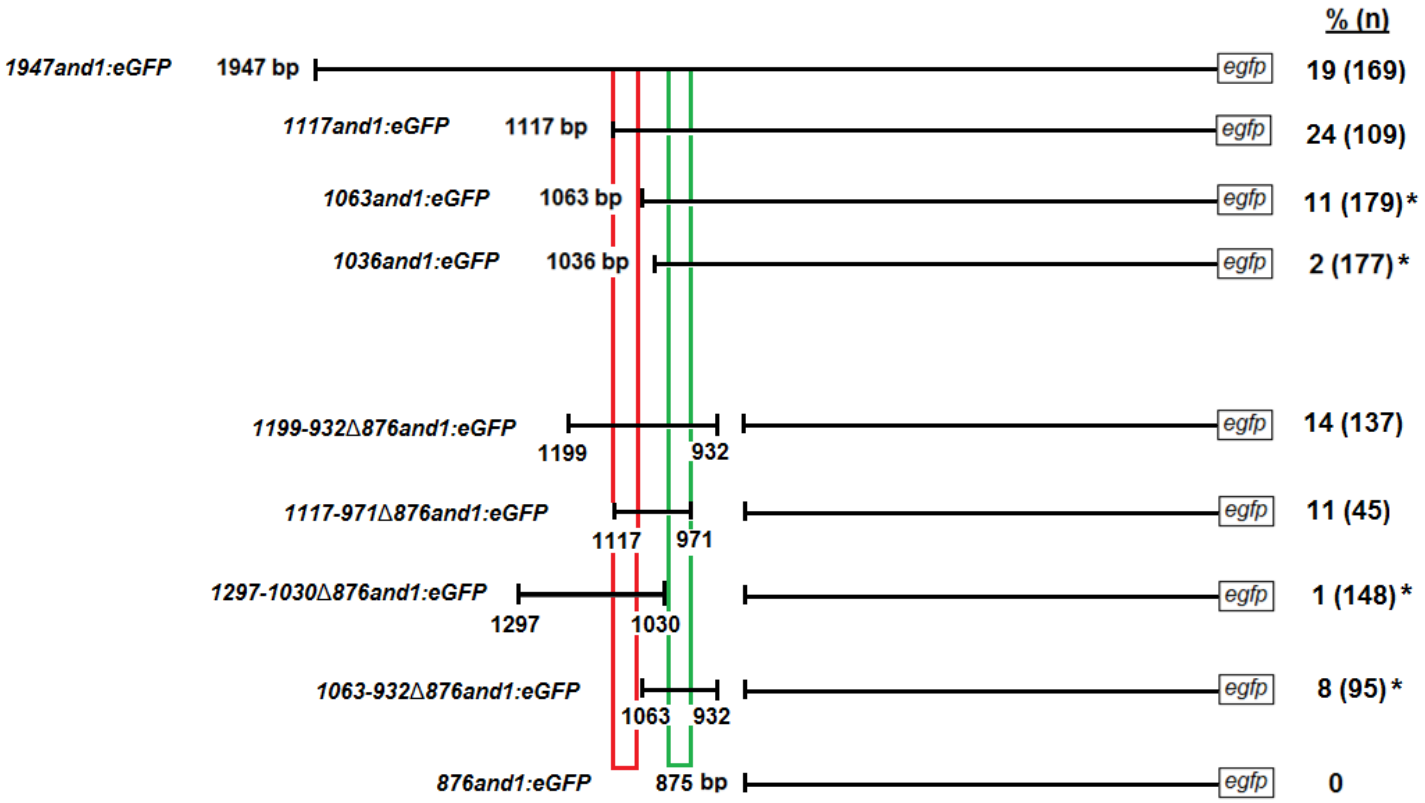


Fig 3.4. Schematic representation of the specific enhancer elements required for reporter expression comparable to that of the *1947and1:eGFP* construct. Constructs retaining two sequences located from -1064 to -1117 (highlighted in red) and from -971 to -1031 (highlighted in green) upstream the *and1* gene drive significantly similar reporter gene expression profiles as the *1947and1:eGFP* construct. The absence of at least one of these two sequences results in significantly decreased numbers of embryos capable of reporter gene expression. Construct names, sequence boundaries, and numbers are repeated from Figure 3.2. Significantly lower percentage values are indicated by an asterisk.

the *1947and1:eGFP* construct is caused by the presence of two necessary elements located upstream the *and1* first exon at the locations -1064 to -1117 and -971 to -1031. Interestingly, two constructs, *1279and1:eGFP* and *1199and1:eGFP*, have significantly higher values of eGFP-expressing embryos in comparison *1947and1:eGFP* (Fig 3.2). It is not clear how these constructs drive an increase in eGFP expression, though it is possible that in these cases silencer elements were removed from the *1947and1* element.

3.2.5. Prediction of transcription factor binding sites to the fin epithelium enhancer elements

TRANSFAC, an online tool that uses a wide knowledge base of published data to predict transcription factor binding sites at specific DNA sequences (http://www.biobase-international.com/product/transcription-factor-binding-sites) (Dunham et al., 2012), was used to analyze the enhancer elements discussed above. Tables 3.1 and 3.2 show the predicted transcription factors that bind to the -1064 to -1117 and -971 to -1031 regions, respectively. Both regions possess potential binding sites for transcription factors with known function in fins or tetrapod limb and tail, which are homologous to the fish paired and median fins. Some notable transcription factors include Dlx3, Pmx1, Cdx2 and Hoxa13 (Tables 3.1, 3.2, and see Appendix 3D for Hoxa13 binding sites). In the -971 to -1031 region there is an overlap of the predicted binding sites for Pmx1 and Dlx3 and these factors may compete for binding to this region. While some of these factors bind to degenerate DNA sequences, their regulatory functions have been described in vertebrate development which will be discussed more thoroughly below.

Table 3.1. Predicted transcription factor binding sites to the -1063 to -1117bp sequence upstream the first exon of *and1*. Results are shown only for factors with the highest possible core and matrix score which is 1.

Position (Strand)	Core score	Matrix score	Sequence	Factor name
1106(-)	1	1	cATTCC	TEF-1
1091(-)	1	1	gttGGAAT	ZABC1
1085(-)	1	1	ATTAT	ZNF333
1084(-)	1	1	tTATTA	PMX1
1082(-)	1	1	ATTAT	ZNF333
1080(-)	1	1	tATAAA	CDX-2
1069(-)	1	1	tTTTCC	NFAT1
1069(-)	1	1	tTTTCC	NF-AT4
1069(-)	1	1	tTTTCC	NF-AT1

Table 3.2. Predicted transcription factor binding sites to the -971 to -1031bp sequence upstream the first exon of *and1*. Results are shown only for factors with the highest possible core and matrix score which is 1.

Position (Strand)	Core Score	Matrix score	Sequence	Factor name
1020 (-)	1	1	cATTCC	TEF-1
994 (+)	1	1	ATAAT	ZNF333
994 (+)	1	1	aTAATTac	dlx-3
993 (+)	1	1	TAATTa	PMX1
993 (-)	1	1	tAATTA	PMX1

3.3. Discussion

The objective of this section was to characterize the fin epithelium-specific enhancer elements of the *and1* gene, in addition to determining potential transcription factor binding sites. Two short DNA sequences were identified at positions -1064 to -1117 and -971 to -1031 upstream the *and1* first exon that work in conjunction with the downstream minimal promoter to drive epithelial-specific expression in the developing fin. An *in silico* analysis was subsequently

conducted to determine potential DNA-protein interactions at these enhancer elements. Future studies can verify the activity of these 54 and 61bp *cis*-acting regulatory elements by cloning them to a minimal promoter such as human Beta-globin (deBoer et al., 1988) in replacement of the *and1* minimal promoter. Additionally, induced mutations at the predicted transcription factor binding sites can provide insight towards the necessity of these sequences for enhancing gene expression. The *1947and1* sequence is not highly conserved with the upstream regions of *and1* genes in other fish (data not shown). The difference in sequences does not rule out different functions however since previous work has shown that changes in enhancer sequences does not necessarily preclude enhancer function (Rabinowitz and Vokes, 2012; Taher et al., 2011).

3.3.1. Potential transcription factors involved in epithelial expression of and1 *in the fin*

a. Dlx

The -971 to -1031 region upstream the *and1* first exon contains a predicted binding site for Dlx3 (Table 3.2). Genes of the *dlx* (*distalless*-related homeobox) gene family are transcriptional activators that are expressed in the developing nervous system, neural crest derivatives, and in the appendages of vertebrates (Kraus and Lufkin, 2006). Predominantly an ectoderm-specific gene, *Dlx3* is expressed in the early ectoderm of the tetrapod limb bud and persists within the AER throughout development (Lee et al., 2010; Mullen et al., 1996). Similarly, *dlx3* is associated with the AER of the zebrafish pectoral fin bud and the cleft cells of the median fin fold (Akimenko et al., 1994). The *and* genes are excluded from the AER and cleft cells (Zhang et al., 2010), therefore *dlx3* may not be a good candidate for epithelial-specific expression of *and1*.

The homeodomain of the Dlx proteins have nearly identical amino acid sequences which suggests that the different members of the family can bind to the same nucleotide sequences (Liu et al., 1997). In contrast to *dlx3*, other members of the *dlx* gene family, including *dlx2* and *dlx5*, are expressed in non-AER ectoderm of the fins and may bind the epithelial enhancer sequence of *and1*. In zebrafish, *dlx2* is expressed throughout the pectoral fin fold ectoderm from 32-48 hpf (Neumann et al., 1999). The *dlx2* expression overlaps with *and1* which is expressed in the non-AER ectoderm from 36-40 hpf in the pectoral fin bud. Expression of *dlx2* has been observed in the AER as well as the non-AER ectoderm (Neumann et al., 1999). It is possible that while Dlx2 activates *and1* expression in the non-AER ectoderm, other factors may prevent the activation of *and1* by Dlx2 specifically in the AER.

Expression of *dlx5a* in the zebrafish pectoral fin buds is localized to the AER at 42hpf. Therefore in the pectoral fin folds, Dlx5a may not activate *and1* expression. In the median fin fold, *dlx5a* is expressed in the ectoderm along the midline starting at 16hpf (Abe et al., 2007). The expression of *dlx5a* is initiated in the cleft cells of the median fin fold, although at later stages it appears that *dlx5a* expression is distributed throughout the entire epithelium of the fin fold (Abe et al., 2007). Starting at 24hpf, both *dlx5a* and *and1* are expressed in the ectoderm of the median fin fold (Abe et al., 2007; Zhang et al., 2010). Similar to the expression pattern of *dlx2* in the pectoral fin fold, it is possible that in the median fin fold Dlx5a positively regulates *and1* in the non-cleft cell ectoderm, although this activation does not occur in the cleft cells.

b. Pmx1

The -1063 to -1117 and -971 to -1031 regions upstream the *and1* first exon each possess consensus binding sequences for Pmx1 (Paired related homeobox gene 1) (Tables 3.1, 3. 2) (De

Jong et al., 1993). Pmx1, also known as Prx1, Prrx1 (in zebrafish), Mhox, or K2, plays an essential role for the skeletal development in mice and chick limb buds, and has reported activities in the developing limb bud of *Xenopus* and the fin bud of zebrafish (Hernández-Vega and Minguillón, 2011; Opstelten et al., 1991; Suzuki et al., 2007). Expression of Pmx1 has been implied in the epithelial-mesenchymal interactions required for the ontogeny of many organs (Opstelten et al., 1991). While the *1947and1* element induces fin epithelium-specific gene expression, Pmx1 is largely associated with mesenchyme gene expression and it is not clear exactly how Pmx1 may be implicated in the epithelium-specific expression of *and1*. However, some possibilities described below suggest possible relationships between Pmx1 and *and1*.

Expression of *Pmx1* is largely observed in undifferentiated mesenchyme of the mice limb buds (Leussink et al., 1995). In zebrafish pectoral fin buds, *prrx1a* and *prrx1b* are expressed in the presumptive mesenchyme of the pectoral fin buds starting at 24hpf (Hernández-Vega and Minguillón, 2011). Enhancer elements for the zebrafish *prrx1* genes have also been shown to drive reporter expression specifically in the endochondral disk of the pectoral fin as well as the mesenchymal cells invading the fin fold (Hernández-Vega and Minguillón, 2011; Suzuki et al., 2007). These results indicate that Prrx1 factors are not regulating *and1* expression in the pectoral fin epithelium. However, the Prrx1 binding site in the *1947and1* element may function with other unidentified enhancers of the *and1* genes for expression in the pectoral fin fold mesenchyme cells.

As determined by *in situ* hybridization, *prrx1a* is expressed in the caudal domain of the zebrafish median fin fold at 24hpf (Thisse et al., 2001). While expression of *prrx1a* is likely present in the median fin fold mesenchyme (Hernández-Vega and Minguillón, 2011), the gene expression pattern shown by Thisse et al. (2001) shows that the morphology of the cells stained

for *prrx1a* have an ectoderm-like appearance – they are large and geometrical – as opposed to mesenchyme cells which are smaller with filopodia extensions (Thisse et al., 2001; Wood, 1988). It is possible that these results from Thisse et al. (2001) show *prrx1* expression in the median fin fold ectoderm, though future analyses, such as cross-sectioning following *in situ* hybridization, are required to verify this. It is possible that Prrx1 activates *and1* expression in the median fin fold ectoderm. In contrast, in the pectoral fin fold, Prrx1 may be involved in the activation of *and1* expression in the mesenchyme.

c. Cdx2

There is a potential binding site for Cdx2 (Caudal-related homeobox 2) in the -1063 to -1117 region (Table 3.1). Cdx1 binding sites were identified in the *1947and1* region, but not in the two epithelium specific segments (Fig 3.5, Appendix 3D). In early stages of mouse and zebrafish embryogenesis, *cdx* is expressed in the lateral mesoderm. Cdx function is implied in limb and fin specification and plays an indirect role for their development (Gurnett et al., 2007; Tabariès et al., 2005; van den Akker et al., 2002). Cdx factors also regulate tail bud formation. In zebrafish, *cdx1* genes are expressed in the tail bud paraxial mesoderm at the late gastrula stage, but are not well described in epithelium tissue (Hawkins et al., 2008). Regulation of *and1* by Cdx in the epithelium of the fin is therefore unlikely. In certain zebrafish mutants, *cdx2* and *cdx1b* gene expression is reduced in the tail bud mesenchyme by 72hpf (Flores et al., 2008), though it remains unclear whether wild-type embryos display overlapping expression of *cdx* and *and1* during these later stages of development.

d. Hoxa13

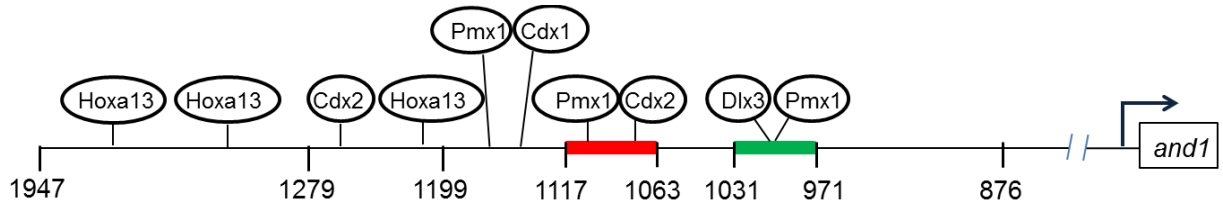


Fig 3.5. Schematic representation of predicted transcription factor binding sites in the upstream enhancer element of *and1*. The green region represents the -917 to -1031 fin epithelial enhancer and the red region represents the -1063 to -1117 fin epithelial enhancer. The region from -1117 to -1199 possesses additional sequences that may bind to transcription factors which may contribute to *and1* gene expression. Hoxa13 potential binding sites are not located in the two identified fin epithelium enhancer elements.

There is a number of predicted *hoxa13* binding sites on the *1947and1* element which are not located within the fin epithelium-specific enhancer regions -1063 to -1117 or -971 to -1031 (Fig 3.5, Appendix 3D). Freitas et al. (2012) showed that overexpression of *hoxd13a* in the zebrafish pectoral fin bud results in increased *hoxa13b* expression and decreased *and1* expression (Freitas et al., 2012). Therefore, *hoxa13* upregulation is associated with the inhibition of *and1* expression. While the focus of this study is on the fin epithelium, and *hox* genes are expressed in the fin mesenchyme, the possibility that *hoxa13* inhibits *and1* genes through direct binding of the *cis*-acting regulatory element may be an important aspect of fin-to-limb evolution.

Hoxa13 and *Hoxd13* are co-expressed in the distal limb bud and are required for autopod development (Sordino et al., 1995; Woltering and Duboule, 2010). Another *Hox* gene, *Hoxa11* is not expressed in the autopod-forming domain, and instead is expressed proximally where it specifies zeugopod fate (Nelson et al., 1996). In the ancestral tetrapod species, the evolution of the autopod involved the distal expansion of *Hoxd13* genes (Shubin et al., 2009; Sordino et al., 1995). Increased *Hoxd13* expression may have resulted in the concomitant upregulation of *Hoxa13* and downregulation of *and1* in the distal appendage, as was shown in the experiments by Freitas et al. (2012). As a result, *and1* was not strongly expressed in the distal appendage which may have contributed to the tetrapod-specific loss of a fin fold (Freitas et al., 2012; Shubin et al., 2009; Woltering and Duboule, 2010).

In contrast, *hoxa13, hoxd13* and *hoxa11* are co-expressed in the developing fin mesenchyme (Davis et al., 2007; Takamatsu et al., 2007). Firstly, this co-localization is suggested to contribute to the major endoskeleton differences observed between fins and limbs (Shubin et al., 2009; Woltering and Duboule, 2010). Secondly, during normal fin development, the co-localization of *hoxa13, hoxd13* and *hoxa11* does not inhibit *and1* expression and

actinotrichia develop normally. On the other hand, co-expression of only two genes, *hoxa13* and *hoxd13,* without the accompanying expression of *hoxa11* in the distal appendage, may inhibit *and1* expression as was the possible scenario during autopod evolution (Fig 3.6a). Overall this implies that Hoxa13 may directly inhibit *and1* in the presence of Hoxd13 but in the absence of Hoxa11.

It has been previously shown that Hox transcription factors, including members of the Hoxa and Hoxd families, are capable of inhibiting gene transcription either by direct DNA binding or indirectly through protein-protein interactions (Kataoka et al., 2001; Pinsonneault et al., 1997; Shen et al., 2001). Inhibition of *and1* gene expression by Hoxa13 may have led to a loss of and1 function and eventual deterioration of the gene from the ancestral tetrapod genome. Freitas et al. (2012) reported that the inhibition of *and1* by Hoxd13a is indirect due to different temporal expression patterns (Freitas et al., 2012). It is possible that Hoxd13a inhibits *and1* via Hoxa13 which may bind to the *and1 cis*-acting regulatory element (Fig 3.6b). Interestingly, *and1* and *and2* double-morpholino knockdown causes an anterior expansion of *hoxd13a* indicating that *and* genes and *hoxd13* genes may negatively impact each other, though this interaction is likely indirect (Zhang et al., 2010).

3.3.2. Unidentified enhancers and silencers may exist in the 1947and1 *element*

The two constructs, *1279and1:eGFP* and *1199and1:eGFP,* displayed higher eGFP expression profiles than the lengthier genomic sequence, *1947and1:eGFP.* Silencer elements may be embedded in the DNA lost when cloning either *1279and1* or *1199and1* from *1947and1* (Fig 3.5). In this scenario however, these possible silencers are also lost from the smaller subfragments analyzed, but the smaller subfragments do not display significantly increased eGFP

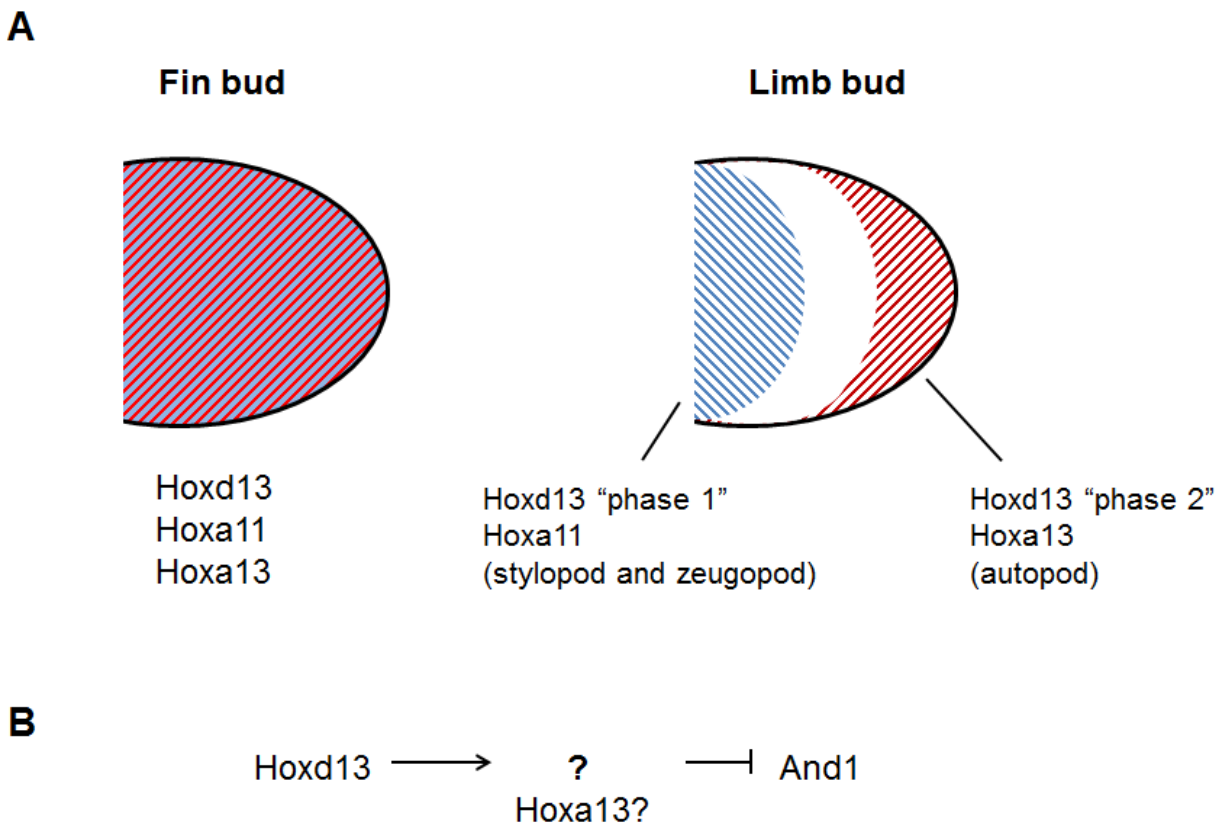


Fig 3.6. The impact of *hox* genes on *and1* expression. A, Schematic representation of different *Hox* gene expression patterns in fin and limb buds. The fin bud expresses overlapping domains of *hoxd13*, *hoxa11*, and *hoxa13* genes. The limb bud expresses Hoxd13 at an early phase (phase 1) which contributes to Hoxa11 localization to the proximal limb. As a result, proximal *hox* genes will give rise to the proximal bones of the limb (stylopod and zeugopod). A second phase of Hoxd13 expression (phase 2) is localized to the distal domain and contributes to Hoxa13 localization to the distal limb. Distally expressed *hox* genes will promote distal limb bone formation (autopod). Proximal is towards the left and distal is towards the right. **B,** Illustration of a possible regulatory pathway involving *hox* and *and1* genes. Hoxd13 inhibits *and1* expression indirectly by means of an unknown intermediate. Hoxa13 is a candidate for directly inhibiting

and1 expression upon the presence Hoxd13 while in the absence of Hoxa11. The absence of Hoxa11 correlates to the expression pattern of the distal limb bud shown in **A**.

expression. To elaborate on the possibility that silencers are lost, an additional enhancer component is likely present in *1279and1* and *1199and1* but deleted upon cloning the smaller DNA subfragments. Accordingly, an additional Pmx1 binding site is located in between the -1279 and -1117 locations (Fig 3.5, Appendix 3D). It has previously been reported that silencers and enhancers regulating the same gene can be in close proximity on the chromosome. Competition between neighboring silencer and enhancer elements can induce different cell fates or gene expression localization (Muroi et al., 2013; Prasad and Paulson, 2011). Future work using deletion or mutational analyses can be used to test the presence of potential *and1* silencer elements and their function in comparison with the downstream enhancers. Interestingly, the upstream potential silencer elements have consensus sequences for Hoxa13 binding.

It is also noteworthy that the gradual deletion of DNA from the 5' end of the *1279and1* element is correlated with a gradual decrease in eGFP expression (Fig 3.2). This can also be explained by the gradual deletion of successive enhancers, possibly by deletion of successive Pmx1 binding sites, that when present, individually add to the strength of gene expression. Additional weak enhancer elements may be present in the constructs that express significantly reduced, however not completely absent, eGFP expression.

3.3.3. Fin epithelium-specific expression versus fin-mesenchyme expression

Preliminary data from our lab has uncovered an approximately 2.5kb enhancer element which overlaps with the *1947and1* element studied above and also includes the first exon and intron of *and1*. When cloned to an *eGFP* reporter, the 2.5kb element can induce fin epithelium expression, fin mesenchyme expression, and expression in the epidermis of the whole embryo by 72hpf. The epithelial and mesenchymal expression in the fins recapitulated *and1* expression in

these tissues once mesenchymal cells have invaded the fin fold (Zhang et al., 2010). Like *1947and1*, the 2.5kb fragment can drive fin epithelium-specific expression as early as 30hpf. The 54 and 61 bp enhancers studied above are likely controlling the fin epithelial expression induced by the 2.5kb element, while other sequences within this genomic DNA are contributing to the mesenchymal expression in the fin. The difference in expression patterns between *1947and1* elements and the 2.5kb element suggests that different transcription factors specific to different cell lineages regulate *and1* expression. Furthermore, our lab has shown that the 2.5kb element can drive LacZ expression in mouse non-AER limb ectoderm only, implying conservation in transcription factors between the fin and limb bud ectoderm but not the mesenchyme.

3.3.4. Statistical analysis

Variability is a factor when analyzing transient transgenic animals. To account for variability, approximately 100 embryos were microinjected per subfragment. However, only 45 embryos were scored for the construct *1117-971Δ876and1:eGFP.* This was due to unexplained death prior to 24hpf of many embryos microinjected with this construct. This same construct was determined as significantly similar to the full *1947and1:eGFP* clone, thus it contains the important enhancer elements. The calculated significance of this construct may be due to the low score number of embryos. Future tests should verify the significance with additional microinjections to gain substantial numbers.

3.4. Conclusion and future directions

Fin epithelium expression of *and1* is likely regulated by transcription factors that are expressed in the fin fold ectoderm tissue, for example Dlx2 or Dlx5. Other transcription factors

expressed in the distal regions of the bud, such as Pmx1 or Hoxa13, may bind additional *cis*-acting regulatory elements to control *and1* expression in the mesenchyme. Additional mesenchyme-specific *cis*-acting regulatory elements are located beyond the *1947and1* element within the first exon-intron region of *and1*. These additional mesenchyme-specific elements may work in conjunction with other sequences situated in the *1947and1* fragment. It has been suggested that while genes may change over evolutionary time, their upstream and downstream effectors likely retain function across different species (Freitas et al., 2012). This is the case for the ectoderm expression observed for *and1 cis*-acting regulatory elements in the mouse ectoderm, although the mesenchyme expression control is lost between mice and zebrafish. The absence of mesenchyme conservation is possibly due to the different *Hox* gene backdrops in fin versus limb development.

Some suggestions for future work on the fin epithelial enhancer elements are mentioned in the discussion above. Additional future studies may include a chromatin immunoprecipitation assay to identify the specific proteins that interact with the DNA at the fin epithelial regulatory elements. For example, it can be determined whether Hoxa13 directly binds to the upstream sequences of *1947and1*. It would be interesting to test the binding capabilities of Dlx proteins with the *1947and1* elements, although previous studies have reported difficulties when using antibody detection of zebrafish Dlx proteins. Alternate experiments may use EMSA or DNAase footprinting to identify DNA-protein interactions in the fin epithelium enhancer elements *in vivo*. Additionally, future application of the fin epithelium enhancers can be used to drive the ectopic expression of genes in the fin bud ectoderm. For example, mutant genes can be ectopically expressed in the fin ectoderm. In this type of test, the function of an embryonic-lethal mutation can be analyzed specifically in the fin without perturbing the development of the whole embryo.

Future analyses may also parallel the tests performed above to study the regulatory mechanisms of other *and* genes.

Chapter 4 - Objective 2: Function of the Actinodin proteins and a comparative analysis to Ecrg4

4.1. Introduction

4.1.1. Actinodin proteins are comprised of an N-terminal and C-terminal domain

A protein encoded by an *and* gene can be divided into two distinct domains. Zhang et al. (2010) previously reported that each And protein has a collection of repeated 9-amino acid motifs with the sequence C(N/D)PXXDPXC, where X is any amino acid. The number of motifs varies in each gene, and in zebrafish the region containing the motifs is always encoded by the last exon (Zhang et al., 2010). In a variety of species, many structural proteins (included in fibers, cell matrices, and membranes) are stabilized by the many disulfide bonds that result from high cysteine content (Kodali et al., 2011). The high cysteine content of the repeated motifs in And proteins may contribute to the structure of the actinotrichia fibrils. The repeated motifs occupy a large fraction of each protein and are always situated in the C-terminus, and this region can therefore be considered a C-terminal domain. In contrast, the N-terminal region of *and* genes encodes a well conserved polypeptide sequence of approximately 60 amino acids that does not contain the repeated motifs, and this region can be considered the N-terminal domain (Zhang et al., 2010).

The And proteins contain at the N-terminus a leading hydrophobic peptide suggesting that it is targeted to the secretory pathway. Another characteristic of the N-terminal domain is a highly conserved di-basic amino acid sequence, RX(R/K)R↓, a canonical target for post-translational processing by convertase enzymes in the secretory pathway where X can be any amino acid and the arrow indicates the potential cleavage site (Seidah and Chrétien, 1999). In

zebrafish, And1 and And2 have an additional RXRR↓ sequence in the C-terminal domain, although And3 and And4 do not (Fig 1.3, General introduction). Convertase-mediated cleavage is significant because the biological function of proteins is typically activated after processing in the secretory pathway (Seidah and Chrétien, 1999; Steiner, 1998). While the And C-terminal domain likely plays a structural role, the potential bioactive function of the N-terminal domain is unclear. A preliminary genomic database search using the 60-amino acid conserved domain has shown that there is a partial sequence conservation between the And N-terminal domain and a relatively small protein called Esophageal cancer related gene 4 (Ecrg4) (Fig 4.1) (M. Andrade-Navarro, unpublished data). The And and Ecrg4 proteins have a similar organization of their predicted helical secondary structure and a putative convertase recognition motif, RX(R/K)R , that aligns in all observed sequences (Fig 4.1). Previous reports have shown that Ecrg4 has characteristics of a secreted protein molecule with cell signaling abilities. It is possible that the sequence similarities between And and Ecrg4 proteins, in addition to other similarities discussed below, indicate a functional similarity between these proteins.

4.1.2. Function of Ecrg4 at the cell membrane

In 1998, Su et al. provided the earliest report of a biological function associated with the gene they named *Ecrg4* because its transcription was downregulated in esophageal cancer tumor tissue (Su et al., 1998). Since its discovery, *Ecrg4,* also known as *Augurin*, has been described in several studies that have suggested different roles of *Ecrg4.* For example, *Ecrg4* has been titled as a peptide hormone (Mirabeau et al., 2007) and a candidate tumor suppressor gene (Li et al., 2009; Li et al., 2010) and it is evidently associated with the progression of various cancers (Camões et al., 2012), cell differentiation (Huh et al., 2009) and senescence (Kujuro et al., 2010).

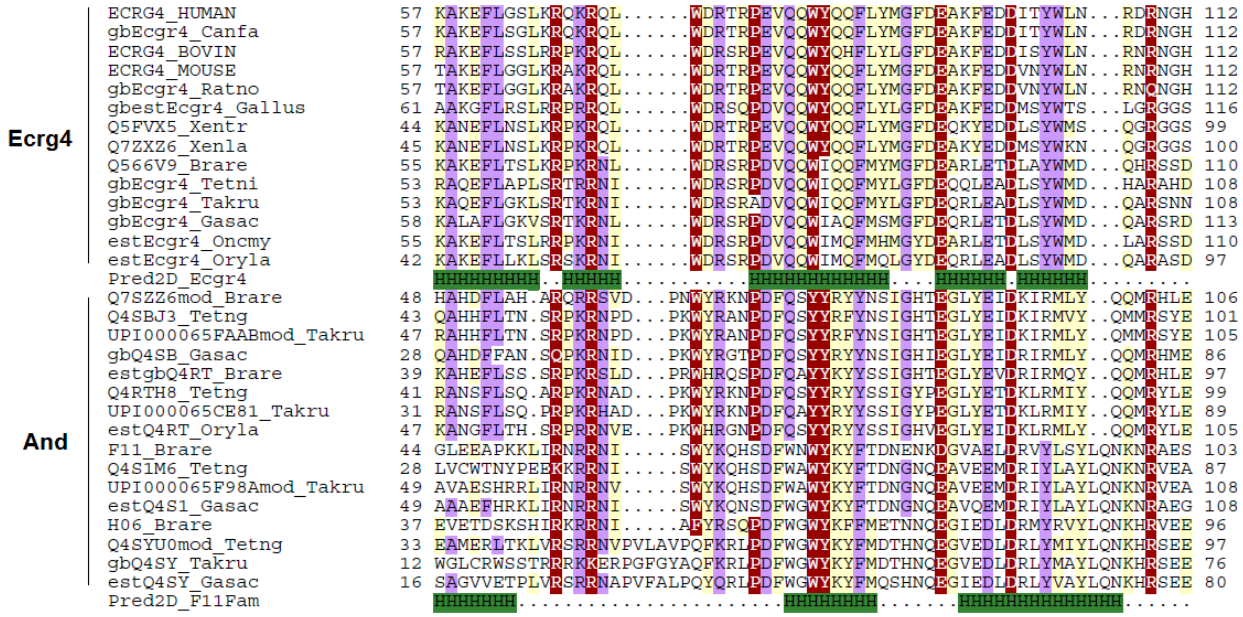


Fig 4.1. Actinodins and ECRG4 display significant similarity. Multiple sequence alignment (MSA) of fragments of proteins from human Ecrg4 and orthologs in selected organisms (top) and Actinodin from fish indicates common patterns of conserved residues. Conservation is shown as high (red), medium (violet), or low (yellow). Below each set of sequences displays predictions for helical secondary structures (H/green) which agrees with the detected homology. Sequences are identified by their UniProt IDs. Other identifiers were used if the sequences were obtained from EST collections. The species indicated are Human *(Homo sapiens)*, Canfa (*Canis lupus familiaris)*, BOVIN *(Bos taurus)*, Mouse *(Mus musculus)*, Ratno *(Rattus norvegicus*), Gallus *(Gallus gallus)*, Xentr *(Xenopus tropicalis)*, Xenla (*Xenopus laevis*), Brare (*Brachydanio rerio)*, Tetni *(Tetraodon nigroviridis)*, Takru *(Takifugu rubripes)*, Gasac *(Gasterosteus aculeatus)*, and Oncmy *(Oncorhynchus mykiss)*. The MSA was produced using TCoffee and the secondary structure was predicted with Jpred3.

Ecrg4 is expressed in a variety of tissues and organs including the blood, brain, internal organs and endocrine glands in developing and adult humans and mice. The exact molecular function of Ecrg4 is not completely understood, and it is not part of any known gene family further suggesting a unique biological role (Mirabeau et al., 2007).

The *ecrg4* gene encodes a leading N-terminus hydrophobic domain that is required for endoplasmic reticulum (ER)-Golgi-mediated secretion and cell membrane tethering (Dang et al., 2012). Ecrg4 has been identified at the cell surface of various cell and tissue types, and it has been suggested that retention of Ecrg4 at the cell membrane corresponds to a quiescent phenotype (Baird et al., 2012; Gonzalez et al., 2011; Mirabeau et al., 2007; Podvin et al., 2011). Injury to the rat hypothalamus causes the loss of membrane-bound Ecrg4 protein and lowered *ecrg4* gene expression in the choroid plexus epithelium (Gonzalez et al., 2011; Podvin et al., 2011). The loss of membrane-bound Ecrg4 is thought to trigger proliferation in neural cells for the repair of damaged tissue (Gonzalez et al., 2011; Podvin et al., 2011). Additionally, upon the initiation of an immune response *in vivo*, membrane-bound Ecrg4, *ecrg4* gene expression and protein levels are greatly reduced from human leukocytes (Baird et al., 2012). These results indicate that Ecrg4 is involved with tissue homeostasis; downregulation of Ecrg4 correlates to injury recognition and retrieval of cellular Ecrg4 occurs at the recovery stages (Podvin et al., 2011; Shaterian et al., 2013). Similarly, increased Ecrg4 expression also correlates to a differentiated phenotype, as observed in articular chondrocytes whereas their dividing precursor cells do not express *ecrg4* (Huh et al., 2009). As suggested by the cancer models, these studies show that the absence of *ecrg4* promotes a proliferative signal to the cell (Baird et al., 2012; Sabatier et al., 2011).

4.1.3. Function of Ecrg4 when released from the cell

Different molecular fragments of Ecrg4 have been observed in the conditioned media of cultured cell lines indicating that Ecrg4 is secreted and proteolytically cleaved (Mirabeau et al., 2007; Ozawa et al., 2011). Generally, 14 and 6-8 kiloDalton (kDa) Ecrg4 peptides have been observed in the conditioned media corresponding to the full length protein (in humans the full length protein is 148 amino acids) and cleaved variants, respectively. Different tissues and cell types process Ecrg4 differently, and less frequently other variable Ecrg4 peptides have been detected (Dang et al., 2012; Ozawa et al., 2011). While some studies imply that secreting Ecrg4 activates a proliferative cue (Baird et al., 2012; Podvin et al., 2011), the purpose of cleaving Ecrg4 into different fragments remains elusive. The highest level of sequence conservation between Ecrg4 orthologs is located from amino acids 71-148 of the human protein, which corresponds to the secreted 6-8 kDa peptide (Mirabeau et al., 2007). Cleavage of Ecrg4 was predicted at the level of the primary structure due to a single putative pro-hormone cleavage site characterized by di-basic amino acid residues. Studies have shown that Furin is likely the natural cellular convertase enzyme that processes pro-Ecrg4 at the di-basic residues to give rise to the Ecrg4(71-148) peptide (Ozawa et al., 2011)It is noteworthy that Furin is required for the anti-proliferative effect of Ecrg4 in cultured mouse pituitary cancer cells. It is therefore likely that after cleavage, Ecrg4(71-148) is involved with inhibiting cell proliferation.

The function of Ecrg4(71-148) has been analyzed in other studies. Upon treatment with the inflammatory stimulant, LPS, the loss of cell-surface Ecrg4 is coupled with the secretion of the Ecrg4(71-148) peptide from leukocytes (Baird et al., 2012). In this case, it remains unclear whether the production of Ecrg4(71-148) is positively or negatively linked to proliferation during inflammation *in vivo*. Additionally, injection of Ecrg4(71-148) into the rat hypothalamus

stimulates the release of neuro-hormones (Tadross et al., 2010). These data suggest that the Ecrg4(71-148) peptide may behave as a cytokine factor when secreted, though its downstream effects are not well defined. It is also possible that different cell types during different developmental stages may process and respond to Ecrg4 differently (Dang et al., 2012).

4.1.4. Gene expression of and *partially overlaps with* ecrg4 *in zebrafish*

The level of similarity between Actinodin and Ecrg4 broadens when observing their gene expression activity in zebrafish. The *and* genes are highly expressed in the epithelial cells of the median fin fold starting at 24hpf and the pectoral fin fold starting at 36hpf (Zhang et al., 2010). In the zebrafish genome, there are two paralogs of *ecrg4*, *ecrg4a* and *ecrg4b*. E*crg4a* is expressed in the median fin fold epithelium at 24hpf (unpublished result, JZ). However, *ecrg4a* and *ecrg4b* are expressed at the base of the pectoral fin bud, but not in the fin fold, at 72hpf.

At 72 hpf, overstaining *in situ* hybridization shows that *and1* is expressed in the dorsal surface of the brain (Fig 4.2a) (Zhang et al., 2010). Earlier, at 48 hpf *and2* is expressed in the dorsal surface of the brain in a location anterior to the *and1* domain (Fig 4.2b) (Zhang 2010). At 48hpf, *ecrg4a* and *b* are expressed in the choroid plexus, a ventricular structure at the dorsal surface of the brain (Fig 4.2c, d) (Gonzalez 2011, Podvin 2011, Garcia-Lecca 2008). The expression pattern of the *and* genes in the brain appears to coincide with the posterior and anterior domains of the choroid plexus.

In addition to disrupted fin fold morphogenesis, double-morpholino gene knockdown of *and1* and *and2* results in hindbrain defects and pericardial edema in zebrafish embryos (Zhang et al., 2010). It has been reported that *ecrg4a* knockdown in zebrafish causes similar edema and brain ventricular hydrocephalus-like phenotypes (Gonzalez et al., 2011). The enlarged tissue

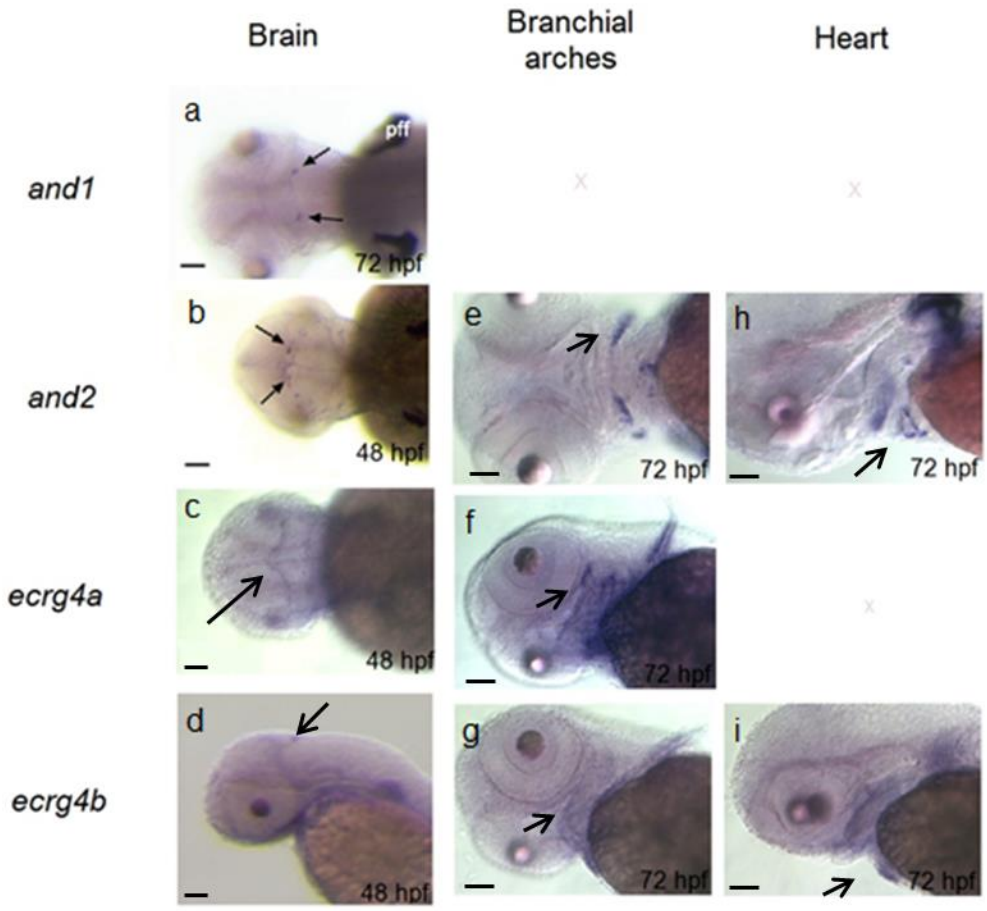


Fig 4.2. Expression of *and1* and *and2* correlates with expression of *ecrg4a* and *ecrg4*b. a-i, *In situ* hybridization of *and1* (a), *and2* (b,e,h), *ecrg4a* (c,f), *ecrg4b* (d,g,i) riboprobes in the brain (a-d), branchial arches (e-g) and heart (h,i) at 48 or 72 hpf (as indicated). Note the expression of *and1* at 72hpf (a) and *and2* at 48hpf (b) in the brain is located at the dorsal surface, correlating to the location of the choroid plexus (arrows in a and b). Images produced by JZ as preliminary data. Scale bars: 80μm.

phenotype of *ecrg4a*-morphant embryos was due to a heightened level of cell proliferation (Gonzalez et al., 2011). It was not determined what kinds of cellular processes caused the edema in the *and1/and2* double-morphant fish. However, based on the similarities between the Actinodin and Ecrg4 sequences and a potential similar function, it is possible that the edema following *and1/and2* knockdown is related to an over-proliferative effect. Other common gene expression domains occur at 72hpf for *and2, ecrg4a* and *ecrg4b* in the branchial arches, and in the heart for *and2* and *ecrg4b* (Fig 4.2e-i). The similarity to *ecrg4* and expression in other tissues suggests that *and* genes may contribute to biological processes other than fin development.

4.1.5. Potential functions of secreted and cleaved And proteins

The expression of *and* genes in other tissues emphasizes that they likely have additional functions beyond contributing to the structure of the actinotrichia fibrils. Similarities between Ecrg4 and And proteins suggests that they may share a functional relationship. Ecrg4 function is dependent on the different ways which it is cleaved namely by convertase enzymes, and whether it is retained at the cell surface or secreted. The function of And proteins may follow a similar pattern. Their biological activity may depend on whether they are secreted as a full protein, potentially cleaved by convertases and secreted as a fragmented protein, or retained at the cell membrane.

I propose that the presence of differentially cleaved Actinodin proteins, consisting of different molecular weights, is an indication that they have been converted into functional molecules. Furthermore, different molecular variants of And proteins implies the possibility that each variant has a different function. The following section will first provide an analysis of the sequence similarities between *and* and *ecrg4* genes and their protein products. Additionally,

within the And1 amino acid sequence, a novel motif that may be cleaved by convertase enzymes is identified. Using a western blot analysis, I show that a Flag-tagged Actinodin1 (And1(flag)) clone is cleaved into molecules of different molecular weights during the first 48hpf in zebrafish embryos. Based on the position of the novel putative convertase cleavage site within the And1 sequence, these analyses raise the possibility that the And1 N-terminal domain is cleaved from the C-terminal domain and may have an individual function as a peptide molecule.

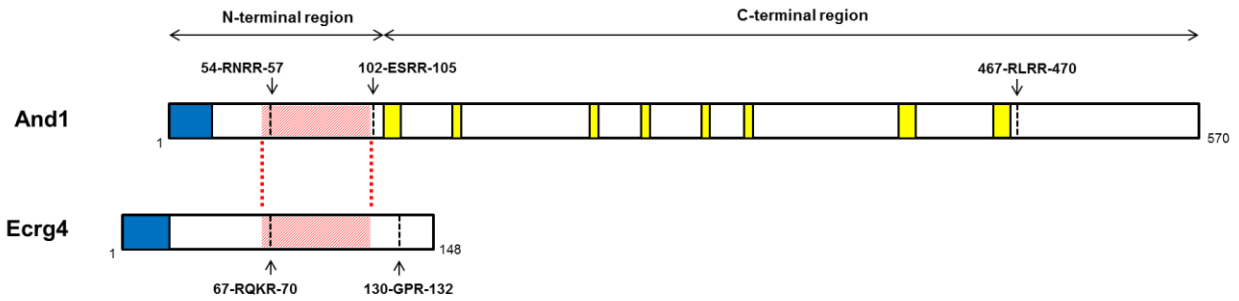
4.2. Results

4.2.1. The And protein N-terminus shows similarities with the Ecrg4(71-148) molecule

As shown in figure 4.1, the partially conserved sequences of And and Ecrg4 occupy only a fraction of the full-length proteins. The schema presented in figure 4.3a shows the locations of the aligning regions in the context of the full-length proteins (Fig 4.3a). The full And1 protein is 570 amino acids and the region which aligns to Ecrg4 correlates to 60 amino acids situated at positions 44 to 103. The human Ecrg4 full length protein is 148 amino acids, and the aligning region correlates to 56 amino acids situated at positions 57 to 112. A sequence alignment of the zebrafish And1 (44-103) and the Human Ecrg4 (57-112) (Fig 4.3b) has a calculated sequence identity of 13/60 (21.7%) and the sequence similarity was calculated as 24/60 (40.0%).

The aligning region of the And proteins occurs immediately downstream the putative convertase cleavage site in the N-terminal domain. The aligning 56-amino acid region of Ecrg4 is also situated immediately downstream a conserved convertase cleavage site with the amino acid sequence RXKR, where X is any amino acid (Fig 4.3b, Appendix 4A). This RXKR sequence of Ecrg4 in humans is located at position 67-70 and is reportedly cleaved by Furin to produce the bioactive Ecrg4(71-148) molecule (Ozawa et al., 2011). Thus the aligning region

a



b

```
D.rerAnd1          44 GLEEAPKKLIRNRRNISWYKQHSDFWNWYKYFTDNENKDGVAELDRVYLS     93
                      ..:|....|.|.:|.: |.:...:...||:.|......:...|.|   ::
H.sapEcrg4         57 KAKEFLGSLKRQKRQL-WDRTRPEVQQWYQQFLYMGFDEAKFEDD---IT    102

D.rerAnd1          94 YLQNKNRAES    103
                      |..|::|...
H.sapEcrg4        103 YWLNRDRNGH    112
```

Fig 4.3. Similar regions of Actinodin and Ecrg4 in the context of the full length, unprocessed proteins. a. Schematic representation of the zebrafish pro-And1 and human pro-Ecrg4 proteins. The red shaded area within each protein diagram indicates the relative location of the sequence alignments shown in Fig 4.1 which is shared among many Actinodin and Ecrg4 orthologs. Putative cleavage sites in And1 and well-documented cleavage sites in Ecrg4 (Baird 2012) are represented as black dotted lines. The amino acid sequences at the sites of the dotted lines are indicated, where the numbers represent their amino acid location in the proprotein and predicted cleavage occurs following the last indicated residue. Additional features of the proteins include the leading signal peptide (blue) and Actinodin C-terminal-specific motifs (yellow). **b.** Pairwise sequence alignment of amino acid sequences corresponding to the red-shaded area shown in a. Predicted convertase cleavage sites are highlighted in green. Alignment produced with Emboss.

between And and Ecrg4 proteins contains a putative convertase cleavage site at the N-terminal side and the remainder of the aligning sequence overlaps with the bioactive Ecrg4(71-148) molecule. Though processing events have only been documented for human, mouse, and rat Ecrg4, the known Ecrg4 cleavage sites are well conserved among orthologs including zebrafish *ecrg4a* and *ecrg4b* (Appendix 4A). Thus, the Actinodin N-terminal domain aligns to the putative bioactive fragment of zebrafish Ecrg4a and Ecrg4b (Appendix 4B).

Using Neuropred (http://neuroproteomics.scs.illinois.edu/cgi-bin/neuropred.py), an online tool for predicting proprotein convertase cleavage sites, the previously mentioned potential di-basic cleavage sites consisting of the RXRR sequence in the And proteins were verified. However, an additional di-basic site (two arginines) was identified at the amino acid position 105 of And1. This site is located at the end of the N-terminal domain and fits the criteria for cleaving by convertases in the secretory pathway (Seidah and Prat, 2012). The probability of convertase cleavage at And1 R-105 is slightly lower than the RXRR sequences (88% versus 92%, respectively). The paired arginines at position 105 are conserved in And1 proteins across many fish species (Appendix 4C), but are absent in the other Actinodin paralogs (Fig 4.4). This di-basic site is located just upstream the beginning of the C-terminal domain. Based on the location of this putative cleavage site, it is possible that the And1 N-terminal domain is cleaved and separated from the C-terminal domain.

4.2.2. And *and* ecrg4 *genes have similar exon sizes and boundaries*

An additional similarity between *and* and *ecrg4* genes is found in the organization of their exons. In a comparison of all zebrafish *and* genes, and zebrafish and human *ecrg4* genes, the sizes of their first three exons display similar sizes (Table 4.1). For example, the size of the

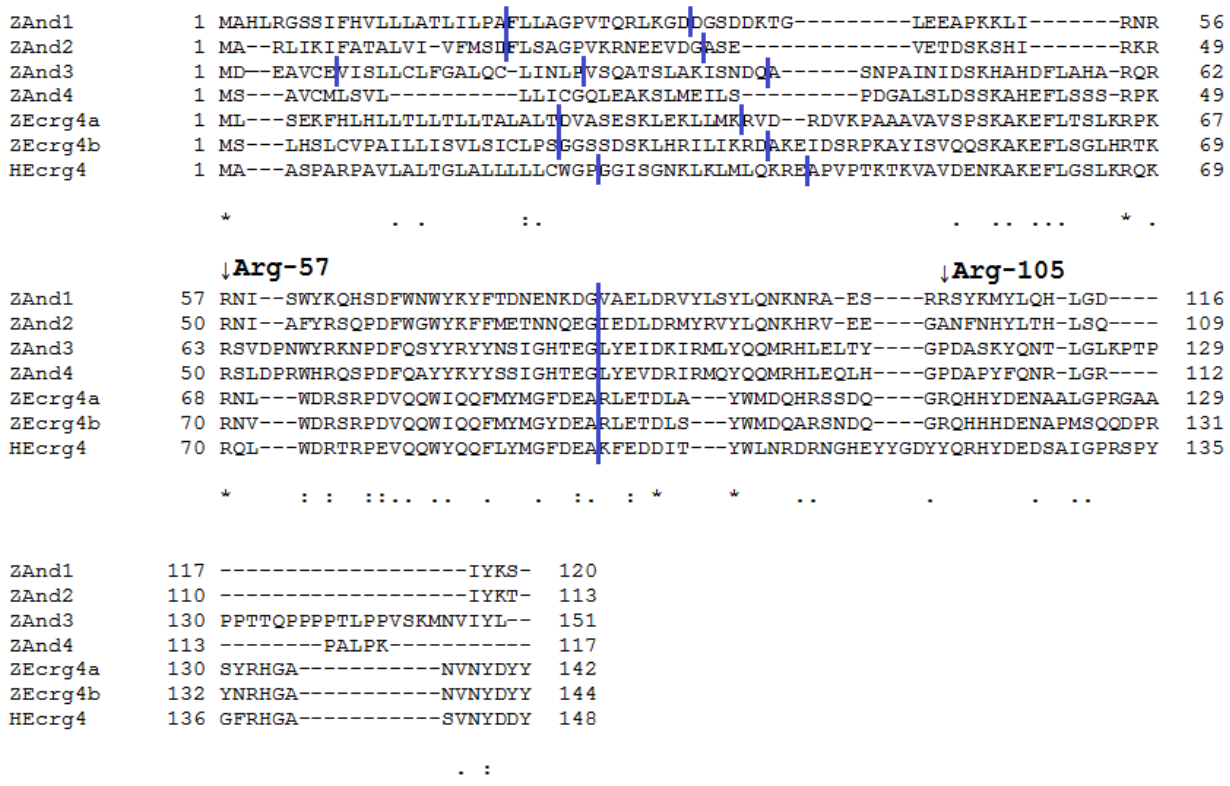


Fig 4.4. Putative cleavage site and exon boundaries for Actinodin and Ecrg4 proteins closely overlap. The putative convertase consensus site at And1 arginine 57 is conserved in all zebrafish And paralogs and aligns with a convertase consensus motif of the Ecrg4 peptides. The putative convertase site at arginine 105 is specific to And1. Note the that the region of highest sequence similarity is located immediately downstream And1 arginine 54, or the conserved RX(R/K)R convertase consensus site. The exon boundary (blue lines) in this aligning segment overlaps in all sequences shown. Identical amino acids (*), six out of seven identical amino acids (:), and four to five out of seven identical amino acids (.) are indicated on the bottom row of the alignment. Sequence and exon information obtained from Ensembl and Zfin (http://zfin.org/) databases. The multiple sequence alignment was produced with Tcoffee. After visual inspection, manual corrections were made to the alignment.

protein sequence encoded by the first exon of all these genes ranges from 19-26 amino acids, the second exon encodes a range of 14-17 amino acids, and the third exon encodes a range of 42-56 amino acids (Table 4.1). A pairwise sequence alignment of the amino acid sequences of each protein shows that the exon boundaries fall closely within one another (Fig 4.4). In this sequence alignment, the boundary between the third and fourth exons (fourth and fifth for Actinodin3 [And3]) directly overlaps (Fig 4.4). Notably, the overlapping exon boundary is situated within the aligning segments of the And N-terminal domain and the Ecrg4(71-148) regions. Unlike the exons however, the intron sizes for each gene vary. The similarities in amino acid sequence plus the overlapping exon boundary further supports a relationship, both in protein product and in the genomic arrangement, between the *and* and *ecrg4* genes.

Table 4.1. Exon sizes of *and* and *ecrg4* genes. Species indicated by letter (Zebrafish as Z., and Human as H.). Non available information indicated by n/a. Exon sizes are presented in number of amino acids encoded by the exon.

	Translated exons		
Gene	Exon1	Exon2	Exon3
Z. And1	22	14	47
Z. And2	19	15	42
Z. And4	n/a	n/a	n/a
Z. Ecrg4a	23	15	55
Z. Ecrg4b	22	16	56
H. Ecrg4	26	17	53
Z. And3	Exon1+2	Exon3	Exon3
	25	14	52
Average	23	15	51

4.2.3. A Flag-tagged And1 protein is cleaved at the N-terminal and C-terminal domains

To detect And1 proteins that have been proteolytically cleaved in the secretory pathway, I cloned a Flag-tag antigen sequence (DYKDDDDK) directly downstream of the putative cleavage site RXRR located within the And1 N-terminal domain (Fig 4.5a, and see methods). With this tool, staining with the Flag antibody can be used specifically for the immunological detection of the N-terminal domain of And1. I predict that following putative cleavage at the RXRR N-terminal site, the flag-tag will remain bound to the And1 N-terminal region that bears a resemblance to Ecrg4(71-148). Previous studies have used this method to detect cleavage variants of Ecrg4 and Insulin-like growth factor 1 (IGF1) proteins showing that cloning of the Flag antigen is compatible with convertase-mediated cleavage (Duguay et al., 1995; Mirabeau et al., 2007).

Zebrafish embryos were co-microinjected at the 1-cell stage with *and1(flag)* RNA and *eGFP* RNA as a control for successful RNA translation. No phenotypic effect was observed for co-microinjected embryos up to 7 days post fertilization (dpf). Co-microinjected embryos were screened for fluorescence, and total proteins were extracted at 48hpf. There is a high expression of *and1* at 48hpf in the median and pectoral fin folds as previously determined by *in situ* hybridization (Zhang et al., 2010), thus any cleavage variants observed for And1(flag) at this stage correlates to those that may appear during fin fold development. Western blotting was used to detect the protein content of extracts from the microinjected embryos.

The schema in figure 4.5a shows the potential cleavage products of And1(flag), and their expected molecular weights, following different combinations of cleavage reactions including cleavage at the RXRR sequences in the N-terminal and C-terminal domains, at the R-105 site and upon the removal of the leading signal peptide (Fig 4.5a). Extracts from *and1(flag)*-injected embryos show a 48kDa band corresponding to the expected size of And1(flag) when cleaved at

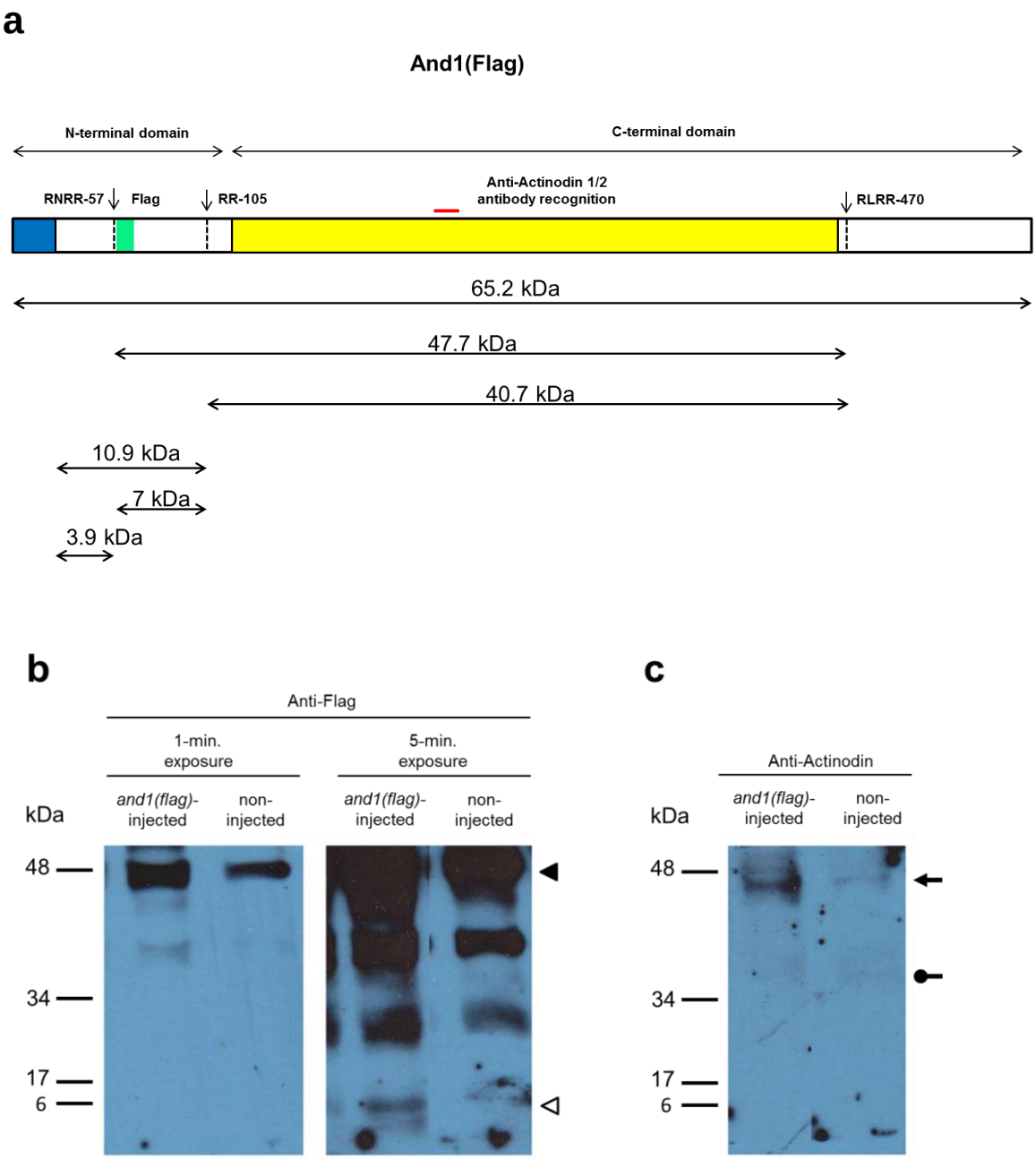


Fig 4.5. Cleavage variants of the And1 protein appear in 48 hpf zebrafish embryos. a, Schematic of full And1(Flag) protein with putative cleavage products and their predicted molecular weights indicated. Protein features shown are the signal peptide (blue), Flag antigen (green), and C-terminal repeated motifs (yellow). **b,** Total protein extracts from 48hpf embryos were stained on an immunoblot with Anti-Flag antibody after 1 minute exposure (left panel) and

after 5 minute exposure (right panel) to chemiluminescent ECL. An approximately 48 kDa band was observed in all *And1(Flag)*- and non-injected samples (black arrowhead). A 6 kDa band was observed at a high exposure for *And1(Flag)*-injected samples (white arrowhead). **c,** The same blot as b**,** stripped then re-probed with the And-antibody for the detection of endogenous And1 and And2 proteins. In *and1(flag)*- and non-injected samples, an approximately 48 kDa and a 36 kDa band are detected by the endogenous And antibody (pointed arrow and rounded arrow, respectively).

the two RXRR sites (Fig 4.5a,b) (4 replicates performed). It is expected that this 48kDa protein contains both the And1(flag) N-terminal and C-terminal domains intact. However, the uninjected samples displayed a similar 48kDa band indicating that the Flag antigen binds non-specifically to a protein of this size whether the Flag antigen was microinjected or not (Fig 4.5b).

The Flag antibody was expected to bind to a 7kDa band which is the product of two cleavage reactions located at the N-terminal RXRR (R-57) as well as the R-105 sequences of And1(flag) (Fig 4.5a). A longer chemiluminescent exposure time of the anti-Flag immunoblot revealed a small peptide of approximately 6kDa that was absent from the uninjected samples (Fig 4.5b). This small peptide may correlated to a small And1(flag) peptide. Many proteins involved in the secretory pathway also undergo post-translational modifications such as glycosylation, phosphorylation, or sulfation (Kögel and Gerdes, 2010; Ozawa et al., 2011). These modifications may change the expected molecular weights of proteins. It is possible that some unidentified proteins are the result of these post-translational modifications on And1, although possible modifications of And proteins currently have not been explored.

4.2.4. Endogenous And1 protein recapitulates the cleavage of Flag-tagged And1

The endogenous Actinodin antibody used in previous studies recognizes an antigen located at the conserved amino acid sequence, DVECLQYHLRAAYGYR, specific to And1 and And2 proteins (Zhang et al., 2010). This shared sequence is situated in the C-terminal domain of each protein (Fig 4.5a). As a control, the immunoblots described above were stripped and re-probed with the endogenous And antibody to confirm whether the Flag antibody and endogenous antibody were binding to the same And1 proteins. The endogenous And antibody recognizes the same 48kDa protein mentioned above, confirming that the C-terminal domain is present along

with the N-terminal domain (Fig 4.5c) (2 replicates performed). The presence of both domains intact implies that the protein was likely cleaved at the two RXRR locations (Fig 4.5a).

With the Flag-antibody discussed above, a small peptide of 6kDa was detected. It is thought that this peptide may be an isolated N-terminal domain following cleavage at the R-105 site. In the event that the N-terminal and C-terminal domains are separated, the endogenous And antibody would detect only the C-terminal domain. The expected band for the isolated C-terminal domain is 40.7kDa and was not observed by western blotting (Fig 4.5a,c). A band of approximately 36kDa was observed after endogenous And antibody staining which may correspond to another unknown C-terminal fragment of And1 or And2 (Fig 4.5c).

4.2.5. And1(flag) is cleaved between one and two days post fertilization

As mentioned above, the Flag antibody bound non-specifically to proteins and interfered with the detection of And1(flag) from 48hpf embryo protein extracts. To determine whether different And1(flag) cleavage variants are present during earlier stages of development, and whether they can be detected without overlapping non-specific antibody binding, protein extracts were carried out at different developmental time points and the protein content was characterized by western blotting. From 5 to 20hpf, the anti-Flag antibody detects high molecular weight proteins ranging from 65-100 kDa (Fig 4.6) (no replicates performed). The 65kDa band corresponds to the predicted size of the unprocessed And1(flag) precursor protein (Fig 4.5a). At 48hpf, there is a faint detection of large molecular weight proteins, and two bands of approximately 48 and 39 kDa are prevalent. The 48 kDa band represents the And1(flag) N-terminal and C-terminal domains intact, while the identity of the 39kDa protein is unknown. These results show that in between the first and second day of embryogenesis some of the high

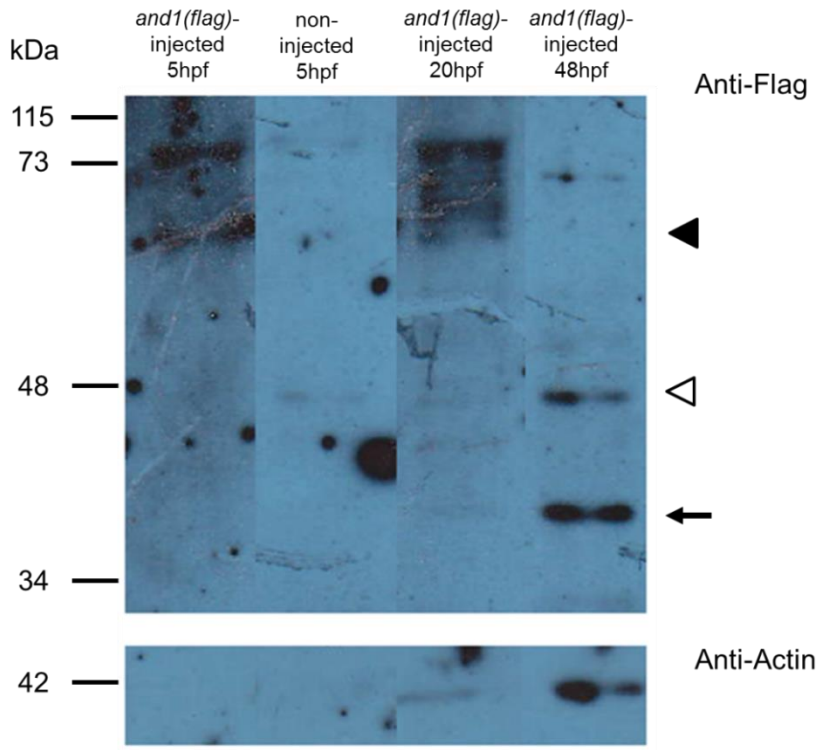


Fig 4.6. Variable protein content is present in zebrafish embryos between 5 and 48 hpf. Protein extraction was carried out on *and1(flag)* RNA-injected and non-injected embryos at time intervals from 5, 20, and 48 hpf and immunoblotted with the Flag antibody (upper panel). At 5hpf and 20hpf, injected embryos display the presence of an approximately 65 kDa band, the predicted size for an unprocessed And1(Flag) protein (black arrowhead). The 65 kDa band is not detected in non-injected samples at 5 hpf. At 48 hpf, the 65 kDa band is lost and a 48 kDa band (white arrowhead) and an approximately 39 kDa band (arrow) are detected. Note that no protein was detected in anti-Flag or Anti-Actin blots on non-injected controls at 20 hpf (not shown). The lanes presented are from the same western blot. The same blot was stripped and re-probed with the Actin antibody as a control (lower panel).

molecular weight proteins are being reduced into low molecular weights which may suggest protein processing activity.

The Flag antibody recognized a non-specific protein of 48kDa (faint band) in uninjected samples from 5hpf in the blot presented in figure 4.6 (Fig 4.6). However at other time points, uninjected samples did not show the non-specific 48kDa band (results not shown). The inconsistent appearance of this 48kDa protein in uninjected samples further supports that this is a non-specific binding of the Flag antibody that occurs randomly.

4.3. Discussion

The main objective of this section was to uncover the activity and potential functions of the And proteins. Firstly, an analysis of the protein sequence was used to predict possible cellular functions of the And proteins. Secondly, an *in vivo* analysis was carried out to test these predictions. In the previous report by Zhang et al. (2010), *and* genes from different species were identified based on the conservation of an N-terminal region as well as the conservation of repeated motifs towards the C-terminal domain (Zhang et al., 2010). Preliminary data also revealed that the polypeptide sequence of the N-terminal domain show some similarities to Ecrg4 proteins. Here I show that the similar region in Ecrg4 corresponds to one of the cleaved Ecrg4 molecules, Ecrg4(71-148). The similarities in polypeptide sequence is an indication that the And N-terminal domain may have a distinct function which may resemble the function of Ecrg4. Moreover, the Ecrg4(71-148) molecule is implied in various cellular processes including proliferation, differentiation, and possibly functions as a signaling molecule (Dang et al., 2012; Li et al., 2010; Tadross et al., 2010). The N-terminal region of And proteins may be involved in related cellular activities.

4.3.1. Actinodin proteins are likely processed by kexin-like convertase enzymes at the RXRR motifs

The N-terminal-most amino acids of the aligning sequences between And and Ecrg4 exhibit the sequences RX(R/K)R. Previous reports have identified that this sequence in Ecrg4 is well conserved and cleaved *in vitro* by the convertase enzyme, Furin (Ozawa et al., 2011). Furin is part of the kexin-like group of convertase enzymes which specifically cleave secretory protein precursors at basic amino acids within the motif $(R/K)X_n(R/K)\downarrow$, where X is any amino acid, n is 0,2,4 or 6, and the arrow indicates the cleavage site (Seidah and Prat, 2012). The And protein N- and C-terminal RXRR motifs display characteristic sequences for recognition by kexin-like convertases. Due to the sequence similarities with Ecrg4, it is likely that the N-terminal RXRR sequence of And proteins is processed by Furin. Furin is ubiquitously expressed in mammalian tissues, and in zebrafish displays prominent expression in the anterior embryo as well as the median fin fold at 24hpf (Carney et al., 2010; Seidah et al., 1994). Zebrafish *sturgeon* mutants deficient for FurinA exhibit a ruffled appearance and blistering of the fin folds (Carney et al., 2010; Walker et al., 2006). Carney et al. (2010) showed that two proteins that maintain the basement membrane, Frem2 and Fras1, are not properly activated in the absence of FurinA resulting in abnormalities of the fin fold tissue (Carney et al., 2010). It is possible that in the absence of Furin, And proteins are not properly activated, which may cause actinotrichia defects and may also contribute to the fin fold abnormalities observed in the *sturgeon* mutants. Additionally, the kexin-like convertase enzyme PC5 often functions redundantly with Furin, and their expression overlaps in regions of the regenerating fin (Khatib et al., 2010) where *and* genes

are also expressed (Zhang et al., 2010). The proteins encoded by the *and* genes are possibly targets of both Furin and PC5.

Secretion of proteinous material into the subepithelial space of the early fin fold was previously reported by Dane and Tucker (1985). This subepithelial granular material likely consists of Collagen and And proteins that will synthesize the actinotrichia (Dane and Tucker, 1985; Durán et al., 2011). Duran et al. showed that two collagen proteins required for actinotrichia synthesis, Col1a1a and Col2a1b, are likely processed and secreted through the activity of Lysyl hydroxylase 1 in the fin fold tissue (Durán et al., 2011). Following convertase enzyme-mediated cleavage, the cleaved substrate is often secreted or localized to the cell membrane (Seidah and Prat, 2012). Additionally, the substrate is converted to an active state following cleavage. It is likely that following convertase-mediated cleavage, And proteins are secreted and become biologically active. Once the And and Collagen molecules have been secreted into the subepithelial space (Dane and Tucker, 1985), they are likely in an activated state which may be a requirement for their involvement with actinotrichia formation.

A 48kDa protein was detected by the Flag antibody and by the endogenous And antibody against protein extracts from 48hpf zebrafish embryos microinjected with *and1(flag)* RNA. The 48kDa protein corresponds to the size of the And1(flag) or the endogenous And1 proteins when cleaved at the N- and C-terminal RXRR putative convertase cleavage sites. Processing of the And1 protein specifically at these locations was previously predicted by Zhang et al. (2010) who also detected a protein of similar molecular weight in 3 dpf wild type zebrafish embryos (Zhang et al., 2010). Following cleavage at both RXRR motifs, And1 and And2 are likely secreted from the cells of the fin fold with the N- and C-terminal domains intact.

An And protein with intact N- and C-terminal domains provides valuable information towards the function of the protein. While the C-terminal domain may contribute to actinotrichia structure, the attached N-terminal domain is present in the actinotrichia fibrils as well. The N-terminal domain may not directly contribute to actinotrichia structure; instead it is possibly carrying out another molecular function within the fin fold. The aligning region between And and Ecrg4 proteins continues downstream from the N-terminal RX(R/K)R sequence. There are several highly conserved amino acids in the N-terminal domain of And proteins that are also conserved in the Ecrg4 proteins. The conservation of these residues suggests that they are retained under selective pressure and may be crucial for the functions of these domains (Conant and Wolfe, 2008; Wilson et al., 1977). As it is similar to Ecrg4, it is possible that the N-terminal domain is performing a similar function as Ecrg4. For example, Ecrg4 is implied in the induction of cell differentiation as well as the repression of cell proliferation (Baird et al., 2012; Huh et al., 2009; Podvin et al., 2011). Similarly, it is possible that the N-terminal domain of And proteins may signal to the tissues adjacent to the actinotrichia for the induction of a differentiated or an anti-proliferative effect. It has been shown that the epithelial cells of the growing fin fold do not multiply. Rather they elongate to contribute to growth (Carney et al., 2010; Yano et al., 2012). Furthermore, once the mesenchyme cells infiltrate the fin fold they are specified for various types of mature tissue, including dermal bone (lepidotrichia) and connective tissue (Hall, 2007). The fate of cells within the fin fold could be partially attributed to their exposure to the N-terminal domain of And proteins within the actinotrichia.

Many other examples of proteins which are processed by convertases are simultaneously secreted, converted into biologically active molecules, and affect their surrounding cellular environments (Lissitzky et al., 2000). As mentioned above, Frem2 and Fras1 are processed by

Furin as a requirement for their function towards the maintenance of the basement membrane (Carney et al., 2010). Additionally, the α-subunit of Integrin and the Repulsive guidance molecule a (RGMa) are processed by convertases in order to function both as signaling molecules as well as contribute to the formation of extracellular structures (Kuninger et al., 2008; Lissitzky et al., 2000). The And proteins may be a novel example of proteins that are regulated by convertase enzymes and function in both cell signaling as well as tissue structure.

4.3.2. The And1 N-terminal domain is possibly cleaved at the RR amino acid motif at position 105

Towards the end of the N-terminal domain in And1 there is a di-basic sequence, RR, located at amino acid position 105 and this is a predicted cleavage site for convertase enzymes. This di-basic site has a different amino acid sequence than the other predicted convertase cleavage sites in the And proteins, the RXRR motifs, which have three arginine residues. The di-basic site at R-105 fits the criteria for processing by kexin-like enzymes. However, due to the difference in sequence, the R-105 site may not be cleaved by the same enzymes as the RXRR sites. In the event that the R-105 site is cleaved, the And1 N-terminal domain would be separated from the C-terminal domain. This R-105 residue does not occur in the polypeptide sequence of other And paralogs, therefore And1 is the only member of the And protein family which may undergo a separation of N-terminal and C-terminal domains.

A 6kDa protein was identified after extended chemiluminescent exposure times of the immunoblot containing protein extracts from *and1(flag)*-microinjected embryos. The detection of this protein upon a long exposure time indicates that it is present at a low concentration. In the event that the And1(flag) protein was processed at the RXRR sequence in addition to the R-105

sequence, a small flag-containing peptide of 7kDa is expected as the product (Fig 4.5a). The 7kDa peptide would consist of an isolated And1(flag) N-terminal domain. I interpreted that the 6kDa protein detected here is a small form of the isolated N-terminal peptide.

The separation of the N-terminal region from the C-terminal region would result from cleavage at the R-105 site. Cleavage at R-105 also implies that the C-terminal fragment would be present as an isolated molecule with a predicted molecular weight of 40.7kDa (Fig 4.5a). Only the endogenous And antibody would recognize the 40.7kDa molecule because the Flag epitope-containing N-terminal domain has been cleaved off. The presence of the 40.7kDa band would also confirm the identity of the small 6kDa protein as the isolated N-terminal domain, further suggesting that cleavage occurred at the And1(flag) R-105 residue. However, the 40.7kDa protein was not detected in short or lengthier chemiluminescent exposures with the endogenous And antibody (Fig 4.5c, and data not shown). Therefore the identity of the 6kDa protein was not confirmed as an isolated N-terminal domain. Future work, discussed below, is required to confirm the identity of the 6kDa peptide.

A requisite for convertase functioning is that the convertase enzyme must be co-expressed with its targeted precursor protein (Seidah and Chrétien, 1999). The putative convertase enzymes that cleave R-105 may be expressed in the fin and may cleave And1 at the R-105 residue in the fin tissue. The scope of the results in this study was limited to the proteolysis of Actinodin proteins in zebrafish during the first 48hpf. Transcripts for *and1* are expressed in the fins during the first 48hpf, and the peptides identified in the results above may correlate to those produced in the fins. In addition, transcripts for *and1* have been identified in the brain at 72hpf (Zhang et al., 2010). It is also possible that the R-105 residue is cleaved by convertases that are co-expressed with And1 in the brain. Future analyses may focus on the

protein processing of And1 at 72hpf to include an analysis of the And1 proteins in the brain tissue.

4.3.3. Ubiquitous RNA expression in microinjected embryos

Microinjection of *and1(flag)* RNA will result in ubiquitous expression of the transcript in the zebrafish embryo (Nüsslein-Volhard and Dahm, 2002). Some of the And1(Flag) transcripts may be expressed by cells of the developing fin, in which case the proteolysis pattern of the fin by 48hpf would be represented in the anti-Flag western blots. Convertase enzymes in other tissues may ectopically process the And1(flag) protein. Ectopic proteolysis would result in potential cleavage variants that do not reflect the fin-specific post translational processing of And1. Due to the high probability of RXRR motif cleavage, it remains possible that regardless of tissue, And1 is cleaved at these specific sequences. Unidentified proteins were detected by western blotting, and it is possible that this results from cleavage at unpredicted residues or ectopic proteolysis of And1(flag). Additionally, *and1* gene expression has not been reported at high levels elsewhere in the embryo by 48hpf. Processing of the And1(flag) protein occurred between 20hpf and 48hpf (Fig 4.6). While convertase enzymes are active in other tissues at this stage (Chitramuthu and Bennett, 2011), the timing of And1(flag) cleavage correlates with the initiation of *and1* gene expression in the fin.

4.3.4. Similar exon structures imply an evolutionary link between and *and* ecrg4 *genes*

The relationship between *and* and *ecrg4* genes was first elucidated by the remote but significant amino acid sequence similarities. This relationship is strengthened by the similar exon boundaries between *and* and *ecrg4.* In the region where the amino acid sequences are most

similar, the exon boundaries closely overlap suggesting that the *and* and *ecrg4* genes may be derived from a common ancestral gene. Additionally, a strong evolutionary constraint implies important structural features or mechanisms are associated with the conserved protein-coding region of the genes (Conant and Wolfe, 2008; Wilson et al., 1977). A possible scenario for the evolutionary origin of the *and* gene is that the ancestral gene for *ecrg4* was fused to the repeated motifs characteristic of the And C-terminal domain. Consistently, *ecrg4* is common to chordates, whereas *and* genes may have appeared later during evolution at the origin of the vertebrates (discussed in Objective 3).

This scenario would also imply that *ecrg4* was duplicated prior to fusion since the genomes of fish consist of an isolated *ecrg4* gene (not fused) as well as the *and* genes. Novel genes may originate from gene duplications (Ohno, 1970), exon shuffling (Gilbert et al., 1997), or genomic insertions of RNA (Brosius, 1999). Other studies have documented the origin of novel genes through the fusion of two separate genes by conjoining RNA transcripts, or by the deletion of genomic DNA in between two adjacent genes (Nurminsky et al., 1998; Thomson et al., 2000). At times, fused genes act as linked independent functional units, or a new function may arise, such as a different cellular localization (Long, 2000). There is a wide variation in the number of C-terminal motifs of And proteins across different species. The number of repetitions is possibly the product of DNA duplications within the chromosomal locus. The fusion of *ecrg4* to these repeated motifs may have been the result of *ecrg4* translocation to a genomic region consisting of highly repetitive DNA. The resultant protein would contain the *ecrg4* exon structures at the N-terminus, including transcriptional start signals, and an additional exon at the C-terminal end consisting of repeated motifs. The novel *and* gene may now fuse the Ecrg4-like secretion and signaling function with the structural properties of the C-terminal repeats.

4.4. Conclusion and future directions

The similarities between the And N-terminal domain and Ecrg4(71-148) peptide provides the suggestion that the And proteins may exhibit some functions that are similar to the functions of Ecrg4. Whether this function is related to cell differentiation, inhibition of cell proliferation, or another molecular process remains unclear. Santamaria et al. (1996) suggested that actinotrichia play a role in the morphogenesis of the lepidotrichia (Santamaria et al., 1996; Santamaría and Becerra, 1991). It is possible that the N-terminal domain of And proteins are specifically involved with this process. For example, the N-terminal domain may induce differentiation of the lepidotrichia-forming scleroblast cells (Santamaria et al., 1996). Furthermore, the expression of *and* genes in tissues other than the fin implies that these genes have functions that do not involve actinotrichia formation. The N-terminal domain of the And proteins may play a molecular role that is required for the development or function of these other tissues.

The Western blots performed in this study showed that And1 and And2 are cleaved into 48kDa proteins with the N-terminal and C-terminal domains intact. The hypothesis that the And1 protein was cleaved at the arginine residue at position 105 was not verified. Future studies may be directed towards repeating the procedure described above, possibly using a higher protein concentrations for Western blotting for the detection of the putative 7kDa peptide. In contrast, it is possible that the R-105 residue is not cleaved in zebrafish embryos during the first 48hpf. Future experiments may explore this possibility by replicating the above experiments or through other approaches that involve the analysis of protein molecules (described below). Additionally,

more replicates are required to verify the change in protein content between 5 and 48hpf zebrafish embryos.

Another approach can be taken for observing the cleavage of the And proteins. The *and1(flag)* fusion gene can be transfected into cultured cells, followed by Western blotting on protein extracts from these cells. Previous studies have tested Ecrg4 cleavage in culture by transfecting the protein into Furin-expressing and Furin-deficient cell lines (Ozawa 2011). The same type of test can be applied to And proteins to determine whether specific convertase enzymes may cleave And proteins. Furthermore, Furin-deficient cells were not subjected to decreased cell proliferation, indicating that Ecrg4 was unable to induce this effect in the absence of Furin (Ozawa, 2011). Similar tests can be applied to the And proteins; differences in cell behavior, including differentiation or proliferation, can be assessed when And is co-expressed or not co-expressed with a convertase enzyme.

The *and1* and *and2* double-morphant zebrafish embryos exhibit edema of the heart and hindbrain abnormalities (Zhang et al., 2010). Similar results were observed for *ecrg4a*-knockdown zebrafish (Gonzalez et al., 2011). Gonzalez et al. used a phosphorylated histone-3 antibody (H3P) that is a marker for proliferating cells in developing zebrafish to determine that the *ecrg4a*-knockdown phenotype was caused by increased proliferation (Gonzalez et al., 2011). To determine whether the *and* genes may have some similar function in the control of cell proliferation, H3P staining can be used to compare wild-type versus *and1* and *and2* double-morphant embryos.

Chapter 5 – Objective 3: Coelacanth *and* genes and the conservation of synteny of the *and* gene loci in vertebrates

5.1. Introduction

Actinotrichia have been documented in the fins of the coelacanth, *Latimeria chalumnae* (Geraudie and Meunier, 1980), which implies the presence of *and* genes in the coelacanth genome. A recent study by Amemiya et al. (2013) identified a single *and* gene in the coelacanth genome (Amemiya et al., 2013). The presence of an *and* gene in the coelacanth provides insight towards the genetic mechanisms of actinotrichia synthesis in coelacanth as well as confirms the original hypothesis that *and* genes were lost specifically from tetrapods (Zhang et al., 2010). This loss of *and* genes during the evolution of an aquatic sarcopterygian into a tetrapod is correlated with the transition from a lobed fin into a limb. Amemiya et al. (2013) showed that while *and* genes are specifically absent from tetrapod genomes, the synteny of this *and* gene is otherwise well-conserved between coelacanth, tetrapods and teleosts (Amemiya et al., 2013). In this study I report the same findings as Amemiya et al. (2013), although I identify an additional novel *and* gene in the coelacanth genome. The results below present novel information in regards to the characterization of both coelacanth *and* genes, termed 'large' *and* and 'small' *and* based on their relative sizes. Also, this study reports the conserved synteny of the novel coelacanth 'small' *and* gene. Like the 'large' *and*, the synteny of the 'small' *and* gene is conserved in tetrapods albeit the specific loss of the 'small' *and* gene. I propose that the conserved *and* synteny dates back to the common ancestor of the teleosts and coelacanth. Based on the observation that synteny of the *and* loci are conserved in tetrapods, possible biological mechanisms that caused the loss of the *and* genes are discussed.

5.2. Results

5.2.1. Identification of the coelacanth 'large' actinodin

In teleost fish, the genes *and1* and *and2* make a paralogous pair and encode larger proteins than *and3* and *and4*, which make another paralogous pair (Zhang et al., 2010). Performing a Blast (tblastn) search of either zebrafish *and1* or *and2* against the coelacanth, *Latimeria chalumnae,* genome identifies a single 'large' *actinodin* ortholog. The corresponding transcript in coelacanth is listed as ENSLACT00000001605, and is situated on the scaffold JH126651.1 (as listed on Ensembl.org, by Wellcome Trust Sanger Institute). The total gene length spans 2586 bp, and the transcript length is 1380 nucleotides. Upon the analysis of the polypeptide sequence of the coelacanth 'large' *and,* many typical features of the Actinodin (And) proteins can be observed. Within this sequence, at amino acid position Arg-2 there is the sequence RQKR which is a consensus sequence for the recognition and cleavage by convertase enzymes (Seidah and Chrétien, 1999)(Fig 5.1). Downstream from the putative convertase site is a series of twelve motifs bearing a close resemblance to the repeated sequence, C(N/D)PXXDPXC, characteristic of the And protein C-terminal domains. There is an additional putative convertase cleavage sequence, RLRR, immediately downstream the C-terminal domain repeats (Fig 5.1). The organization and appearance of the coelacanth 'large' And protein is highly similar to the 'large' And proteins, And1 and And2, of zebrafish.

By definition, an ortholog is one member of a group of genes in different species that are derived by phylogenetic descent from a single common ancestral gene (Koonin, 2005; Postlethwait, 2007). The identification of orthologs is important for tracing the ancestry of a specific gene as well as drawing conclusions about shared gene functions (Postlethwait, 2007).

Fig 5.1. Polypeptide sequence alignment of coelacanth (*L. chalumnae*) 'large' And versus zebrafish (*D. rerio*) And1 reveals similar proteins. Pairwise sequence alignment produced with Emboss. The length of the full sequence alignment is 637 amino acids. Green highlights indicate predicted convertase recognition sequences. Yellow highlights indicate conserved motifs characteristic to And protein C-terminal domains. Red highlighted amino acids indicate the

identical sequence spanning 14 units. The zebrafish sequence was obtained from zfin.org and the coelacanth sequence from ensembl.org.

The identification of orthologs is typically accomplished by measuring sequence alignments. I next performed a 'reciprocal best hit' (RBH) sequence alignment (Postlethwait, 2007; Wall et al., 2003) to determine whether the sequence of zebrafish And1 or And2 more closely resembles the translated sequence of the coelacanth 'large' *and* gene. A closer sequence may be an indication of a similar function. The first step of RBH is to blast the protein product of the gene of interest (in this case zebrafish *and1* or *and2*) to the genome of interest (in this case coelacanth), which are the results mentioned above. The second step is to blast the protein product of the resultant gene (in the case from coelacanth) back to the initial genome from which the gene of interest was derived (the zebrafish genome). When blasting the coelacanth 'large' And back to the zebrafish genome, the result is that And2 is the more similar sequence with a 35.4% identity across a 286 amino acid segment of the coelacanth 'large' And protein.

I further analyzed the sequence similarities by conducting pairwise alignments of the polypeptide sequences of zebrafish And1 and And2 against the coelacanth 'large' And (Figs 5.1, 5.2). The zebrafish And1 and the coelacanth 'large' And protein sequence identity was calculated as 166/637 (26.1%), and the sequence similarity was calculated as 225/637 (35.3%). The zebrafish And2 and the coelacanth 'large' And protein sequence identity was calculated as 158/564 (28.0%), and the sequence similarity calculated as 231/564 (41.0%). The alignments indicate that And2 has slightly higher similarity and identity to the coelacanth ortholog. There is not a large difference in similarity and identity in both alignments and it appears that both zebrafish protein sequences have diverged from the coelacanth sequence (Figs 5.1, 5.2). However, upon a closer look at the sequence alignments, the longest stretch of sequence identity occurs between zebrafish And1 and the coelacanth 'large' And where 14 identical amino acids

Fig 5.2. Polypeptide sequence alignment of coelacanth (*L. chalumnae*) 'large' And versus zebrafish (*D. rerio*) And2 reveals similar proteins. Pairwise sequence alignment produced with Emboss. The length of the full sequence alignment is 564 amino acids. Green highlights indicate predicted convertase recognition sequences. Yellow highlights indicate conserved motifs characteristic to And protein C-terminal domains. The zebrafish sequence was obtained from zfin.org and the coelacanth sequence from ensembl.org.

spans from positions 451-464 of the zebrafish protein (Fig 5.1). In contrast, the longest span of identical amino acids is 6 when comparing zebrafish And2 with the coelacanth protein.

It is noteworthy that when aligning the And proteins, many of the aligning amino acids occur in the repeated motifs of the C-terminal domains (Figs 5.1, 5.2). To test whether the relationship between proteins can be determined solely by the conservation of C-terminal motifs, I performed a sequence alignment of only the C-terminal motif sequences in their naturally occurring order (Fig 5.3). The coelacanth 'large' And C-terminal motifs and the zebrafish And1 C-terminal motifs sequence identity was calculated as 43/106 (40.6%) and the sequence similarity calculated as 51/106 (48.1%). When compared to the zebrafish And2 C-terminal motifs, the sequence identity was calculated as 59/106 (55.7%) and the sequence similarity was calculated as 66/106 (62.3%). This alignment reveals that the C-terminal repeats of And2 and not And1 are more similar to those of the coelacanth 'large' And (Fig 5.3). The overall sequence analyses support that zebrafish And2 is more closely related to the coelacanth protein, however the span of 14 identical amino acids between zebrafish And1 and the coelacanth 'large' And supports that these two proteins share an identical region that has been well preserved in each species.

5.2.2. Synteny of the 'large' and *locus in teleosts and coelacanth*

Synteny is defined as the preservation of the arrangement of genes on a chromosome in different species. Conserved synteny can help resolve the identification of orthologous genes based on having retained the same neighboring genes on their chromosomal locus, and can also reveal the ancestral genomic organization of the particular locus (Amores et al., 1998; Cañestro et al., 2009; Jaillon et al., 2004; Lv et al., 2011; Putnam et al., 2007). Many past studies analyze

a

Coelacanth 'large' And (12 motifs)	Zebrafish And1 (8 motifs)	Zebrafish And2 (10 motifs)
CDPLKDTHC	CAESDDPNC	CANSDDPDC
CDPRKDLYC	CDPYRDPYC	CNPYLDPYC
CDSSDTAC	CDAEDVEC	CNPSDTEC
CDPQKDPSC	CDPKTDPQC	CDPTKDPYC
CDPETDPRC	CDPRVDPYC	CNPATDPLC
CNPYYDSGC	CNPLYEDNC	CNPLFDEGC
CDTTYDPYC	CFVGYDKEC	CDPRYDPYC
CDPRYHPYC	CDPEYDRNC	CYMFYDKDC
CEDYYDLRC		CHPFDPNC
CYIYYDKDC		CDPEIDYNC
CNPRDPNC		
CDPEYDPNC		

b

```
D.rerio            1 CAESDDPNCCDPYRDPYCCDAEDVECCDPKTDPQCCDPRVDPYCCNPLYE     50
                     |....|.:||||.:|.||||:.|..||||:.||.||||..||.||||.|:
L.chalumnae        1 CDPLKDTHCCDPRKDLYCCDSSDTACCDPQKDPSCCDPETDPRCCNPYYD     50

D.rerio           51 D--------------------------NCCFVGYDKE--------CCDP     65
                     .                          .||::.|||:        ||||
L.chalumnae       51 SGCCDTTYDPYCCDPRYHPYCCEDYYDLRCCYIYYDKDCCNPRDPNCCDP    100

D.rerio           66 EYDRNC     71
                     |||.||
L.chalumnae      101 EYDPNC    106
```

c

```
D.rerio            1 CANSDDPDCCNPYLDPYCCNPSDTECCDPTKDPYCCNPATDPLCCNPLFD     50
                     |....|..||:|..|.|||:.|||.||||.|||.||:|.||||.||||.:|
L.chalumnae        1 CDPLKDTHCCDPRKDLYCCDSSDTACCDPQKDPSCCDPETDPRCCNPYYD     50

D.rerio           51 EGCCDPRYDPY------------------CCYMFYDKDCCHPFDPNCCDP     82
                     .||||..||||                  |||::||||||:|.|||||||
L.chalumnae       51 SGCCDTTYDPYCCDPRYHPYCCEDYYDLRCCYIYYDKDCCNPRDPNCCDP    100

D.rerio           83 EIDYNC     88
                     |.|.||
L.chalumnae      101 EYDPNC    106
```

Fig 5.3. Sequence comparison of 'large' And protein C-terminal motifs reveals similarities between zebrafish And2 and the coelacanth 'large' And. a, C-terminal motifs of the And proteins are characterized by having a close resemblance to the sequence C(N/D)PXXDPXC (Zhang et al., 2010). This motif is repeated 12 times in coelacanth 'large' And (left column), 8 times in zebrafish And1 (middle column), and 10 times in And2 (right column). The order of appearance of the motifs from N-terminus to C-terminus in the endogenous protein is presented in the columns from top to bottom, respectively. **b,** Pairwise sequence alignment of the C-

terminal motifs in their endogenously occurring order of zebrafish (*D.rerio*) And1 versus the coelacanth (*L.chalumnae*) 'large' And. The length of the full sequence alignment is 106 amino acids. **c,** Pairwise sequence alignment of the C-terminal motifs in their endogenously occurring order of zebrafish (D.rerio) And2 versus the coelacanth (L.chalumnae) 'large' And. The length of the full sequence alignment is 106 amino acids. Sequence alignments were produced with Emboss.

the conservation of synteny on a large chromosomal scale in the order of megabasepairs (Mb) (Albalat et al., 2012; Cañestro et al., 2009; Catchen et al., 2009; Lv et al., 2011; Makino and McLysaght, 2010; Putnam et al., 2007) whereas other patterns of conserved synteny have been studied on a smaller subset of genes within the order of tens or hundreds of kilobasepairs (kb) (Force et al., 1999; Postlethwait, 2007; Santagati et al., 2003; Zhu et al., 2007). Large scale synteny is useful to determine the large scale evolution of genomes and allows the observation patterns such as whole genome duplications (Meyer and Van de Peer, 2005). On the other hand, a small scale analysis of synteny can reveal patterns of specific gene translocations or deletions. The report below will focus on a smaller-scale synteny of the *and* gene loci, and identifies genes which are linked by a maximum of 500 kb. In this study, the analysis of small scale synteny facilitates a detailed observation of how genes are kept in the same order on the chromosome or how a specific gene is lost from a syntenic locus while the other genes are maintained in a similar order.

I investigated the conserved synteny of teleost *and1* and *and2* to determine whether these teleost genes share an ancestral syntenic locus with the coelacanth 'large' *and* gene. For a variety of species, a graphical representation of their chromosomes and the positions of their resident genes can be visualized on the Ensembl.org genome browser (www.ensembl.org). On this website, the "region comparison" tool allows the user to perform sequence alignments of large chromosomal regions, scaffolds, and contigs of a given species versus others, and the results are presented in a graphical display. With this tool, I have identified that the locus of the coelacanth 'large' *and* is syntenic to the loci of *and1* across many teleost species, including zebrafish on chromosome (Ch) 24, as well as fugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*), tilapia (*Oreochromis niloticus*) and medaka (*Oryzias latipes*) (Fig 5.4a, Appendix 5A). The

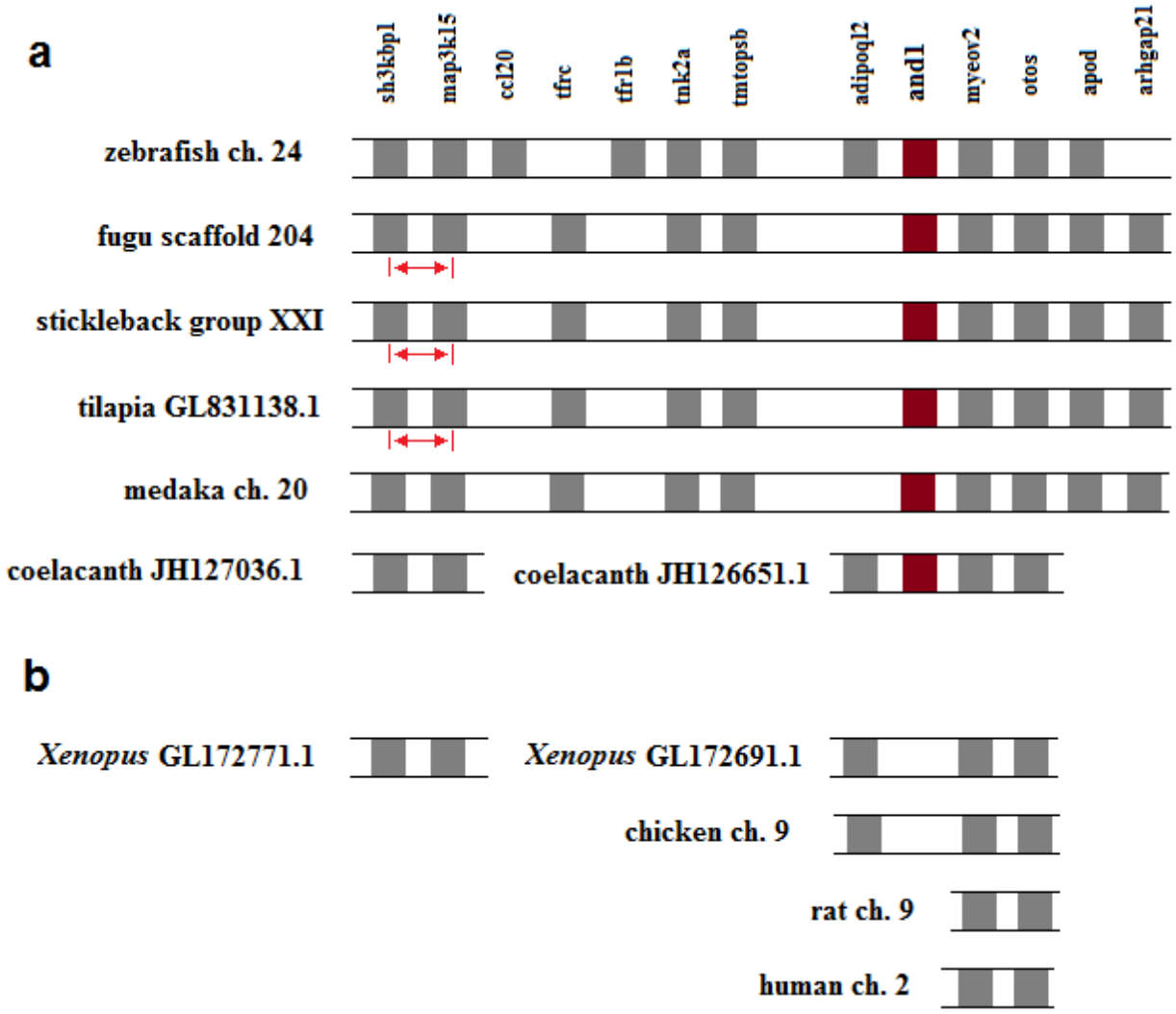


Fig 5.4. Schematic representation of various chromosomal loci of fish species (a) and tetrapods (b) showing synteny to the zebrafish *and1* locus. Common name of species and chromosome (Ch.) or contig names are shown on the left of each strand. The presence of a gene (top) is indicated by a grey square in the locus of a given species. The presence of the *and1* gene is indicated by a red square. Gene sizes and intergenic distances were redrawn without scale for this schema to illustrate the shared genes and collinearity within different genomes. Red arrows indicate gene inversions. Non-syntenic genes within each locus were excluded from the schema.

See Appendix for full gene content of each locus. Chromosome and contig names are represented by their Ensembl IDs.

chromosomal locations for the loci of each species are indicated in figure 5.4a and the appendices.

In a sequential order, the coelacanth scaffold JH126651.1 contains the genes *adipoql2,* the 'large' *and, myeov2,* and *otos.* The collinearity, or the linear arrangement, of the genes *and1, myeov2,* and *otos* is common to the teleost species mentioned above. Conversely, the arrangement of the genes immediately surrounding *and2* on the zebrafish Ch 2 consists of *lrrcc1, rally, and2, Fam134b* and *ensdarg00000061049* (Ensembl.org). There is a low degree of synteny of these genes which are adjacent to zebrafish *and2* versus other teleosts (Appendix 5B). Additionally, the genes immediately adjacent to *and2* in all observed teleosts display no synteny to the coelacanth 'large' *and* locus. The synteny of the *and1* locus is common to teleosts and the coelacanth, therefore this arrangement of genes was likely inherited from their shared common ancestor that existed approximately 450 million years ago (mya) (Xiaobo and Min, 2010). Orthologous genes theoretically possess identical syntenies immediately after a whole-genome duplication event (Ohno, 1970; Wolfe, 2000). In the early teleosts, the synteny of *and1* and *and2* was likely the same immediately following the teleost-specific whole genome duplication approximately 400-300 mya (Taylor et al., 2003; Vandepoele et al., 2004). With the evidence presented here I propose that *and1* retained this ancestral synteny, while *and2* of teleosts was either translocated or its synteny has greatly diverged.

5.2.3. Synteny of the 'large' and *locus in tetrapods*

As shown above, the synteny of the genes *adipoq2, and1*, *myeov2,* and *otos* is retained in teleost and the coelacanth genomes. Using the same methods described above, the syntenic genes *adipoq2, myeov2,* and *otos* have been identified as collinear genes in the genomes of *Xenopus*

tropicalis and chicken (*Gallus gallus*). Notably, no gene is situated in between *adipoq2* and *myeov2* in these tetrapod loci (Fig 5.4b, Appendix 5C), whereas in fish species this is the usual location of *and1* (Fig 5.4a). Rat (*Rattus norvegicus*) and human (*Homo sapiens*) display partial synteny to this region such that their genes for *myeov2* and *otos* are adjacent (Fig 5.4b, Appendix 5C). This data shows that the 'large' *and* gene was lost specifically from the tetrapod genomes, while the surrounding synteny has been retained. This pattern of conserved synteny of the *and1* locus in fish and tetrapods was previously observed by Amemiya et al. (2013) (Amemiya et al., 2013). All other results presented in this chapter have not been previously documented.

While the "region comparison" tool is useful for the assessment of genes that are conserved in different loci, this tool does not provide specific information about conserved non-coding elements (CNEs) in the DNA. Using another online tool, mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml, created by the DOE Joint Genome Institute), I compared the locus of the coelacanth 'large' *and* gene to syntenic regions in zebrafish and tetrapods. An alignment of the coelacanth and the *Xenopus* loci show that the sequences *adipoq2*, *myeov2*, and *otos* are conserved (Fig 5.5). There is no sequence alignment for *and1* in *Xenopus*, although there are some short sequences with low identity in the region in between *adipoq2* and *myeov2* (Fig 5.5). During evolution, genes may mutate into pseudogenes and are gradually lost (Li et al., 1981; Zhang et al., 2010b), and this may have been the case for *and* genes. The short sequences identified with mVISTA may indicate remnants of either coding or non-coding sequences that once accompanied the *and1* gene in an ancestral version of the tetrapod locus. However, no prominent CNEs were identified across all species.

5.2.4. Identification of the coelacanth 'small' actinodin

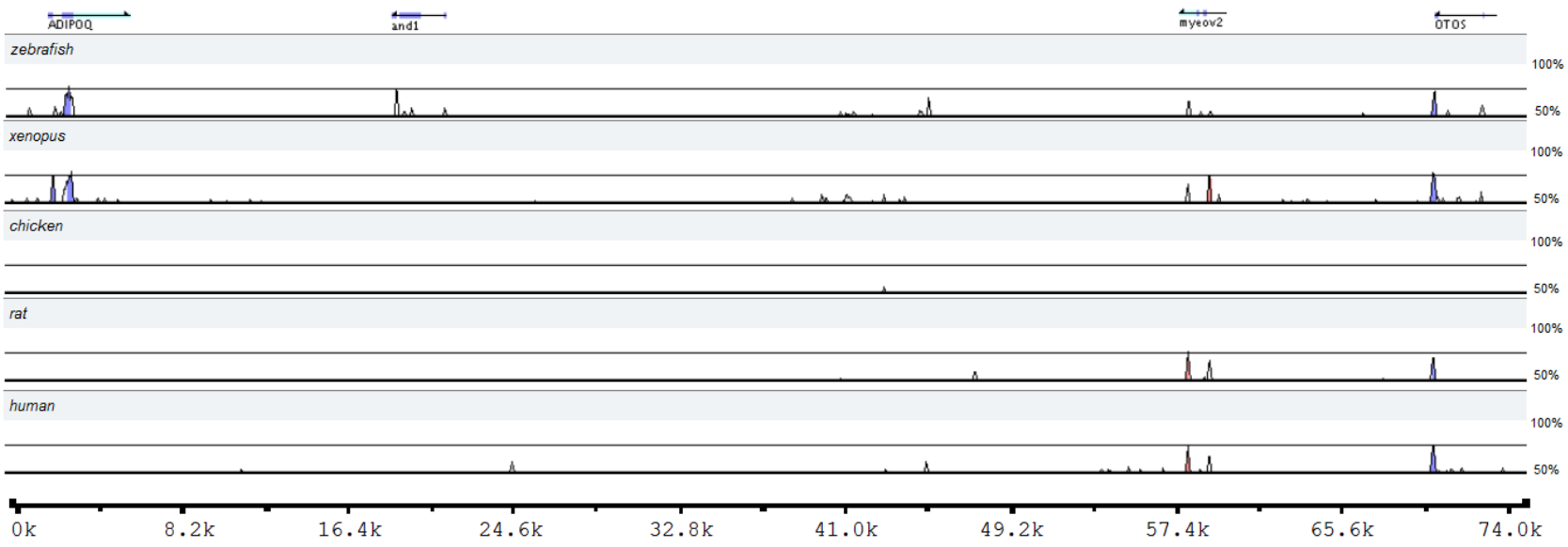

ADIPOQ
and1
myeov2
OTOS
zebrafish
xenopus
chicken
rat
human
100%
50%
0k
8.2k
16.4k
24.6k
32.8k
41.0k
49.2k
57.4k
65.6k
74.0k

Fig 5.5. Sequence alignment of the *and1* syntenic region across vertebrates shows conserved gene sequences and a low conservation of non-coding sequences. All alignments were produced in reference to the coelacanth 'large' *and* gene locus on scaffold JH126651.1. The *and1* gene is not present in the tetrapod species, although the other syntenic genes display sequence conservation. There are small regions of conserved non-coding sequences in between the positions of *and1* and *myeov2.* The largest of these non-coding elements in the *Xenopus* locus is 60bp. All genomic sequences were obtained from ensemble, and IDs are shown in Fig 5.4. Sequence alignments were produced with mVISTA. The scale bar is measured in kb.

Blasting (tblastn) either zebrafish And3 or And4 results in the detection of a single coelacanth scaffold AFYH01283041.1 (as listed on Ensemble.org). There is no gene annotation for this scaffold. This scaffold is a total of 3.4kb, although the aligning 'small' *and* gene occupies a sequence of 1124bp within this scaffold. One possible translated product of this 1124bp sequence results in a polypeptide that has typical characteristics of an And protein. In this coelacanth 'small' And protein, there are three conserved motifs resembling the C(N/D)PXXDPXC sequences of the And C-terminal domains, and there is a predicted convertase cleavage site with the sequence SKKR downstream from the C-terminal motifs (Fig 5.6). Using the RBH approach, it was shown that zebrafish And3 is most similar to the coelacanth 'small' And protein with a 59.1% identity over a span of 40 amino acids of the coelacanth protein. In contrast, a pairwise sequence alignment indicates that zebrafish And4, and not And3, has a slightly higher percentage of similarity and identity to the coelacanth ortholog (Figs 5.6, 5.7). The zebrafish And3 and the coelacanth 'small' And protein sequence identity was calculated as 85/357 (23.8%), and the sequence similarity calculated as 112/357 (31.4%). The zebrafish And4 and the coelacanth 'small' And protein sequence identity was calculated as 84/342 (24.6%), and the sequence similarity calculated as 111/342 (32.5%). There is not a large difference in similarity and identity in both alignments. As with the 'large' And proteins, both zebrafish 'small' *and* proteins sequences have diverged from the coelacanth ortholog.

Many aligning amino acids of the 'small' *and* proteins occur at the C-terminal motifs (Figs 5.6, 5.7). To determine whether the relationship of the proteins can be better defined solely on the C-terminal motif sequences, an alignment of only the C-terminal motif sequences, in their naturally occurring order, was carried out. It is revealed that the C-terminal repeats of zebrafish And3, and not And4, are more similar to those of the coelacanth 'small' And (Fig 5.8). The

Fig 5.6. Polypeptide sequence alignment of coelacanth (*L. chalumnae*) 'small' And versus zebrafish (*D. rerio*) And3 reveals similar proteins. Pairwise sequence alignment produced with Emboss. The length of the full sequence alignment is 357 amino acids. Green highlights indicate predicted convertase recognition sequences. Yellow highlights indicate conserved motifs characteristic to And protein C-terminal domains. The zebrafish sequence was obtained from zfin.org and the coelacanth sequence from ensembl.org.

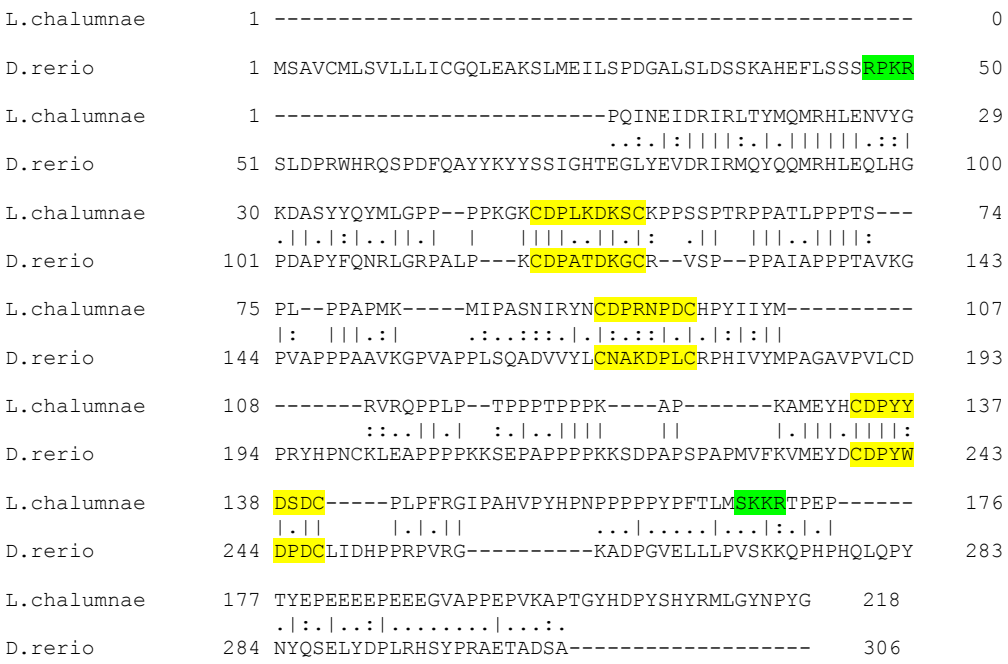


Fig 5.7. Polypeptide sequence alignment of coelacanth (*L. chalumnae*) 'small' And versus zebrafish (*D. rerio*) And4 reveals similar proteins. Pairwise sequence alignment produced with Emboss. The length of the full sequence alignment is 342 amino acids. Green highlights indicate predicted convertase recognition sequences. Yellow highlights indicate conserved motifs characteristic to And protein C-terminal domains. The zebrafish sequence was obtained from zfin.org and the coelacanth sequence from ensembl.org.

a

Coelacanth 'small' And (3 motifs)	Zebrafish And3 (3 motifs)	Zebrafish And4 (4 motifs)
CDPLKDKSC	CNSKDPQC	CDPATDKGC
CDPRNPDC	CDPRYHPNC	CNAKDPLC
CDPYYDSDC	CDPYWDPDC	CDPRYHPNC
		CDPYWDPDC

b

```
L.chalumnae        1 CDPLKDKSCCDPR-NPDCCDPYYDSDC     26
                     |:. ||..||||| :|:|||||:|.||
D.rerio            1 CNS-KDPQCCDPRYHPNCCDPYWDPDC     26
```

c

```
L.chalumnae        1 CDPLKDKS--------CCDPR-NPDCCDPYYDSDC     26
                     |||..||.        ||||| :|:|||||:|.||
D.rerio            1 CDPATDKGCCNAKDPLCCDPRYHPNCCDPYWDPDC     35
```

Fig 5.8. Sequence comparison of 'small' And protein C-terminal motifs reveals similarities between zebrafish And3 and the coelacanth 'small' And. a, The typical And protein C-terminal motif (Zhang et al., 2010) is repeated 3 times in coelacanth 'small' And (left column), 3 times in zebrafish And3 (middle column), and 4 times in And4 (right column). The order of appearance of the motifs from N-terminus to C-terminus in the endogenous protein is presented in the columns from top to bottom, respectively. **b,** Pairwise sequence alignment of the C-terminal motifs in their endogenously occurring order of zebrafish (*D.rerio*) And3 versus the coelacanth (*L.chalumnae*) 'small' And. The length of the full sequence alignment is 27 amino acids. **c,** Pairwise sequence alignment of the C-terminal motifs in their endogenously occurring order of zebrafish (*D.rerio*) And4 versus the coelacanth (*L.chalumnae*) 'large' And. The length of the full sequence alignment is 35 amino acids. Sequence alignments were produced with Emboss.

coelacanth ‘small’ And C-terminal motifs and the zebrafish And3 C-terminal motifs sequence identity was calculated as 17/27 (63.0%) and the sequence similarity calculated as 21/27 (77.8%). When compared to the zebrafish And4 C-terminal motifs, the sequence identity was calculated as 19/35 (54.3%) and the sequence similarity was calculated as 22/35 (62.9%). It must be considered that due to the incomplete sequencing and annotation of the coelacanth genome, it is possible that the sequence information of the ‘small’ And protein is not represented in it’s entirely in this study. However, with the data available it appears that the sequence of zebrafish And3 is overall more similar to the coelacanth ‘small’ And protein sequence.

5.2.5. Synteny of the ‘small’ actinodin *locus in teleosts and coelacanth*

Determining the synteny of the coelacanth ‘small’ *and* gene is not directly possible because the contig on which it is situated has no other identified genes. Using the patterns of synteny in other well-characterized regions of teleost, coelacanth, and tetrapod genomes, the conservation of synteny of the ‘small’ *and* gene can be inferred. Using the “region comparison” tool, it was found that *and4* of zebrafish Ch 14 is situated in a highly syntenic region to seven other teleost species analyzed (Appendix 5D). The schematic representation in figure 5.9a shows a group of closely linked genes in the zebrafish *and4* locus and that these genes have the same collinearity in all the other analyzed teleosts (Fig 5.9a). Within these loci, three genes, *stag3l3, mcm7,* and *cops6* are closely linked. Similarly, on a single scaffold in the coelacanth genome, JH128754.1, these same genes, *stag3l3, mcm7,* and *cops6* , are linked (Fig 5.9a, Appendix 5D). The sequence of this coelacanth scaffold is incomplete, and it is currently unknown whether the coelacanth ‘small’ *and* gene is embedded within this scaffold. In all teleost species analyzed,

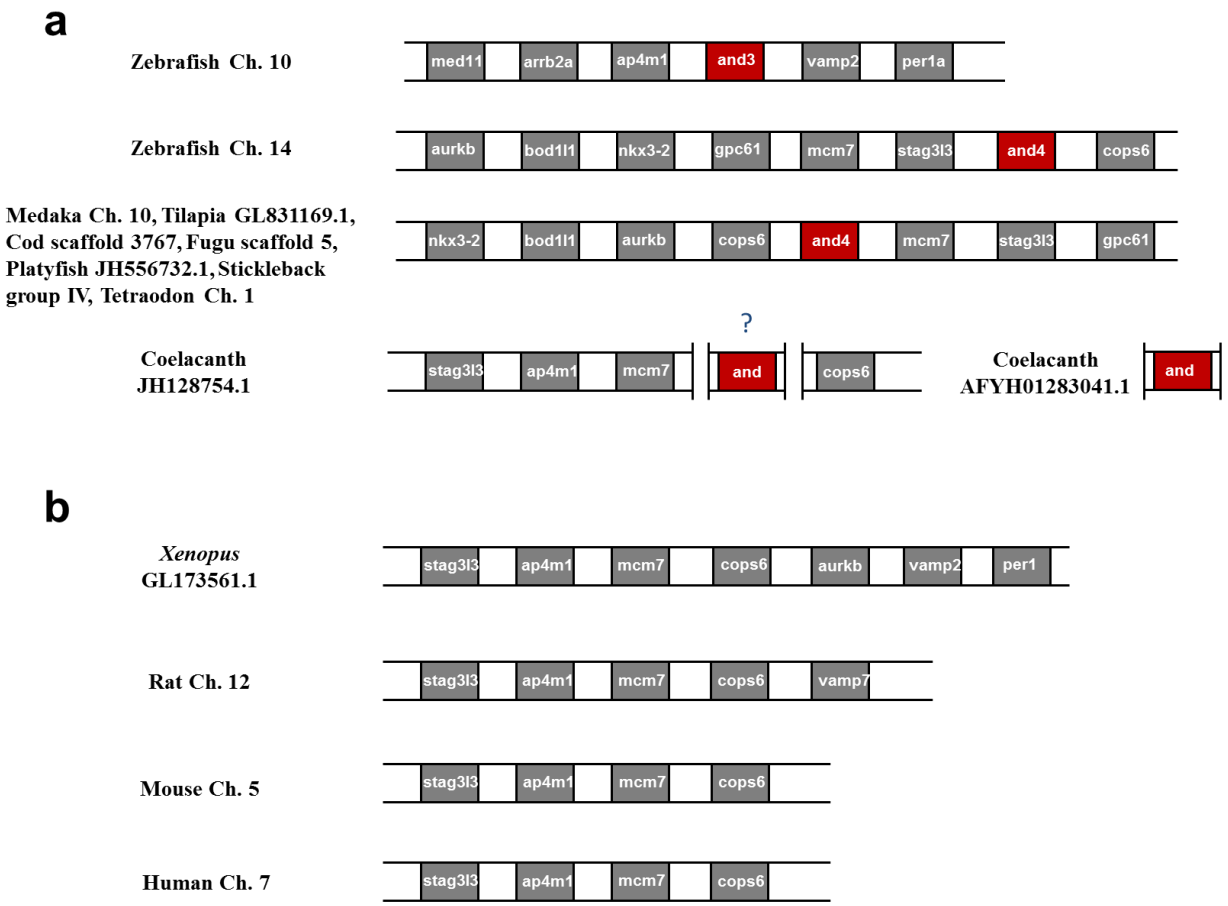


Fig 5.9. Schematic representation of various chromosomal loci of fish species (a) and tetrapods (b) showing synteny to the zebrafish *and4* and *and3* loci. Common name of species and locus or chromosome name are indicated on the left. The presence of a gene is indicated within a grey box in the locus of a given species. The presence of the *and* genes in each locus is indicated by a red box. The coelacanth scaffold JH128754.1 is syntenic to the *and4* locus of other fish species. The coelacanth 'small' *and* gene is located on scaffold AFYH01283041.1 which is 3.4 kb, and has no other known resident genes. Due to the conserved synteny of fish *and4* loci, it is proposed that the coelacanth scaffold AFYH01283041.1 is located within the larger scaffold JH128754.1. Genes are shown in their naturally occurring order on the chromosome (Ensembl.org). Non-syntenic genes within each locus were excluded from the

schema. See Appendix for full gene content of each locus. Gene sizes, and intergenic distances are not to scale. Loci and gene names are presented by their Ensembl IDs.

and4 is located in between the genes *cops6* and *mcm7* (Fig 5.9a, Appendix 5D). It is noteworthy that there is a gap of un-sequenced content in between the coelacanth orthologs of *cops6* and *mcm7* (as observed on Ensebl.org genome browser, April 29, 2013).

With the "region comparison" tool, I observed that there is a lesser degree of conserved synteny of the zebrafish *and3* locus, located on Ch 10, as compared to the synteny of *and4.* Two genes that are closely adjacent to *and3, ap4m1* and *arrb2a* are syntenic to one locus of other teleosts, whereas other closely adjacent genes to zebrafish *and3* are dispersed on other loci of different teleosts (Appendix 5E). This data reveal that the organization of the genes immediately surrounding *and3* is not conserved between zebrafish and other teleosts.

The zebrafish gene *ap4m1* is closely adjacent to *and3,* yet the coelacanth ortholog of *ap4m1* is linked to the coelacanth genes *stag3l3, mcm7,* and *cops6* on the scaffold JH128754.1. This shows that this coelacanth scaffold possesses genes which are linked to both zebrafish *and3* and *and4.* Conversely, other closely adjacent genes to zebrafish *and3,* for example *arrb2a* and *med11*, are syntenic to the coelacanth locus JH129683.1, whereas two other closely neighboring genes to zebrafish *and3, per1a* and *vamp2,* are syntenic to the coelacanth locus JH129489.1 (Appendix 5E). Note that these latter two coelacanth loci are not syntenic to *and4.* This data shows that there is no highly conserved synteny for the immediate genomic surroundings of zebrafish *and3* versus the coelacanth genome. Based on the consistency that *and4* is located in between the genes *cops6* and *mcm7,* and the well-conserved synteny of the *and4* locus, I predict that the coelacanth 'small' *and* gene is located in between the genes *cops6* and *mcm7* on the coelacanth scaffold JH128754.1 (Fig 5.9a).

As mentioned above, the coelacanth locus JH128754.1 is mostly syntenic to teleost *and4,* but has one gene, *ap4m1,* which is syntenic to zebrafish *and3*. Theoretically, the teleost *and3* and

and4 loci were derived from the same ancestral chromosome that was replicated during the teleost-specific whole genome duplication (Vandepoele et al., 2004). The coelacanth locus may be reflective of the ancestral locus which harbors synteny to either *and3* or *and4.* The *and4* locus of teleosts bears a closer resemblance to the coelacanth locus, and can therefore be considered to have retained much of the ancestral synteny. In this case, the synteny of the zebrafish *and3* locus has greatly diverged with the exception of *ap4m1*.

The non-zebrafish teleost species mentioned in this study belong to the subgroup Acanthomorpha whereas zebrafish belong to the subgroup Ostariophysi (Azuma et al., 2008; Steinke et al., 2006; Wittbrodt et al., 2002). The *and3* locus of the Acanthomorpha species (5 species shown in Appendix 5F) are highly syntenic to each other, and as mentioned above are not syntenic to zebrafish *and3.* The exclusive similarity of the Acanthomorpha synteny is due to their close phylogenetic relationship to each other and more distant phylogenetic relationship to zebrafish (Steinke et al., 2006; Wittbrodt et al., 2002). Furthermore, the neighboring genes to Acanthomorpha *and3* are syntenic to a locus in coelacanth, although this coelacanth locus has no known association to any *and* paralog (Appendix 5F). Additionally, the neighboring genes to Acanthomorpha *and3* are syntenic to different zebrafish loci, although these zebrafish loci have no known relationship to any *and* paralog (see Methods, Synteny Database). Overall this data suggests that the Acanthomorpha *and3* locus evolved independently from other fish species, and this locus has no clear relationship to the ancestral *and* loci.

5.2.6. Synteny of the 'small' actinodin *locus in tetrapods*

As shown above, the coelacanth scaffold JH128754.1 contains the genes *stag3l3, mcm7,* and *cops6* which are syntenic to the teleost *and4* locus. This coelacanth locus also possesses the

gene *ap4m1* which is syntenic to the zebrafish *and3* locus. Upon comparing the zebrafish *and4* locus to the genomes of various tetrapods, it is observed that these four genes *stag3l3, mcm7, cops6,* and *ap4m1* remain syntenic (Fig 5.9b, Appendix 5G). It was previously stated that the coelacanth synteny reflects the ancestral chromosome prior to the teleost-specific genome duplication because one coelacanth locus bears synteny to two duplicated genes in the teleosts. This concept is strengthened upon analyzing the synteny of the *Xenopus* locus GL173561.1. This *Xenopus* locus is syntenic to zebrafish *and4* such that it contains the genes *stag3l3, mcm7, cops6* and *aurkb* (Fig 5.9b). The same *Xenopus* locus is syntenic to the zebrafish *and3* locus such that it contains the genes *ap4m1, vamp2* and *per1* (Fig 5.9b). This *Xenopus* locus can be interpreted as having a resemblance to the ancestral 'small' *and* gene locus prior to the split of the sarcopterygians and the actinopterygians – however the *and* gene has been lost from the *Xenopus*. It was also previously mentioned that in the fish genomes analyzed, *and4* is consistently located between the genes *cops6* and *mcm7*. These two genes remain adjacent to one another in all tetrapod genomes analyzed (Fig 5.9b, Appendix 5G). It appears that the 'small' *and* gene from these loci was lost from the region directly in between *cops6* and *mcm7.*

The zebrafish *and3* locus was observed to retain one gene, *ap4m1*, which was likely present in the ancestral 'small' *and* gene locus. Additionally, *and3* does not have a strong synteny across all teleost species. I analyzed the synteny between *and3* and the available tetrapod species genomes to verify that the *and4* locus has an overall higher degree of conserved synteny. With the exception of *ap4m1*, the remaining genes in the *and3* locus are not syntenic to one particular region in the tetrapod genomes. The neighboring genes of *and3* are dispersed across numerous genomic locations of coelacanth and tetrapods (Fig 5.10, Appendix 5H). Additionally,

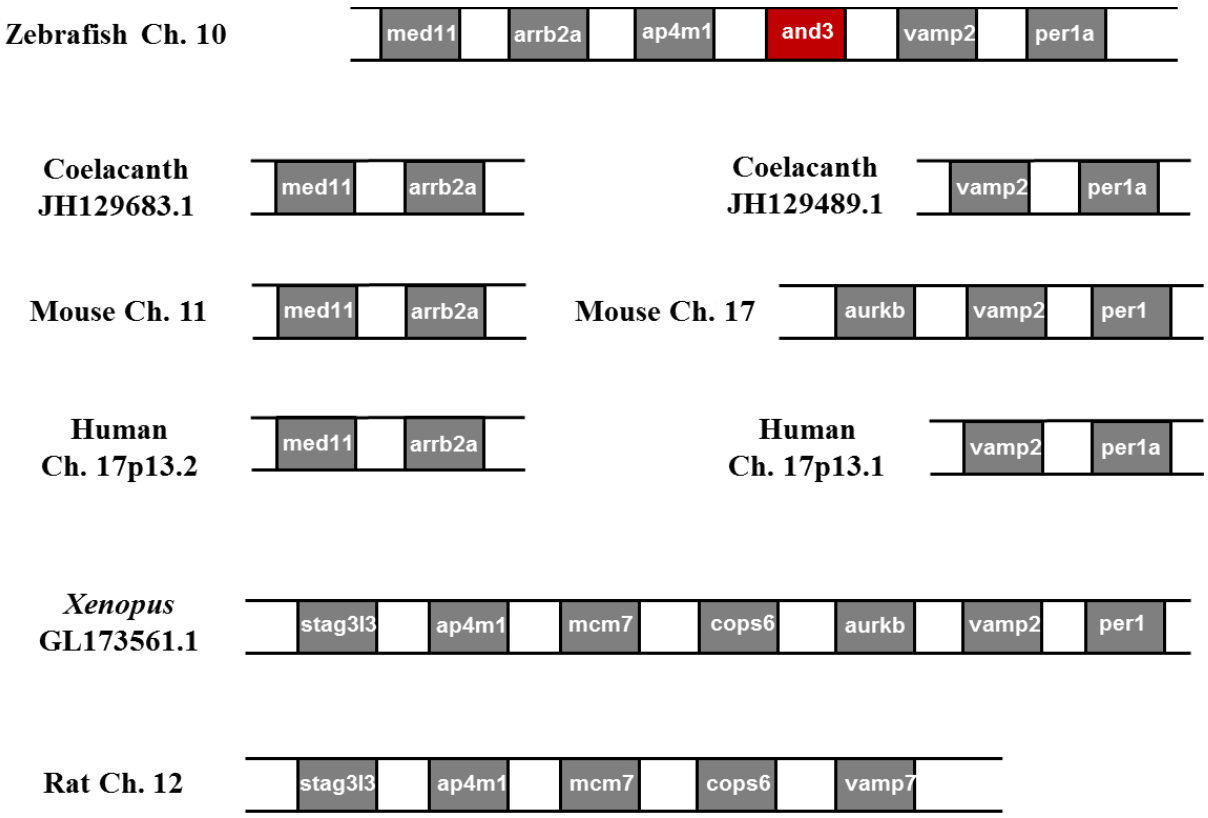


Fig 5.10. Schematic representation the zebrafish *and3* locus and syntenic genomic loci in the sarcopterygians. Common name of species and chromosome or locus name are indicated on the left. The presence of a gene is indicated within a grey box in the locus of a given species. The presence of the *and3* gene is indicated by a red box. Note *Xenopus* and rat loci presented are also syntenic to the zebrafish *and4* locus. Non-syntenic genes within each locus were excluded from the schema. See Appendix for full gene content of each locus. Gene sizes, and intergenic distances are not to scale. Loci and gene names are presented by their Ensembl IDs.

the Acanthomorpha *and3* loci are syntenic to dispersed locations in the tetrapod loci (see Methods, Synteny Database). In contrast, the synteny of the genes adjacent to *and4* is well maintained across many vertebrate species, including teleosts, coelacanth and tetrapods (Fig 5.9). I propose the *and4* synteny represents a locus that predates the divergence of the actinopterygians and sarcopterygians.

5.3. Discussion

5.3.1. Identification of and *genes in the coelacanth genome and implications for fin development and evolution*

The availability of the coelacanth genome sequence provides a useful tool for the analysis of the genetic changes associated with the evolution of aquatic to terrestrial vertebrates. Of particular interest on this topic is the implication that the loss of the *and* genes from tetrapods contributed to the fin-to-limb transition (Zhang et al., 2010). Here it is shown that the coelacanth, a sarcopterygian sister species to the tetrapods, possesses in their genomes two paralogs of *and* genes. The protein products of the coelacanth 'large' and 'small' *and* genes have characteristics of a typical And protein, including repeated C-terminal motifs and putative convertase cleavage sequences. It is generally useful to identify orthologs for the accurate assessment of the function of a gene in one species and compare this function to its ortholog in another species (Postlethwait, 2007). Both teleost *and1* and *and2* are orthologous to the coelacanth 'large' *and* gene and both teleost *and3* and *and4* are orthologous to the coelacanth 'small' *and* gene. The sequences of these orthologs have similarities and distinctions, and future studies may aim to identify how some of their functions have been preserved or diverged.

Previous reports have shown that the proteins encoded by the *and* genes are associated with the structures of the actinotrichia (Zhang et al., 2010). The presence of *and* in the coelacanth genome is consistent with previous observations that coelacanth fins develop actinotrichia fibrils (Geraudie and Meunier, 1980). Formation of the actinotrichia is commonly associated with fin fold development (Dane and Tucker, 1985; Géraudie, 1977). The presence of *and* genes as well as actinotrichia in coelacanth suggests that these lobed fins develop a fin fold. Based on Thorogood's clock model, the predicted coelacanth fin fold develops after a large fin endoskeleton has been specified (Thorogood, 1991). In concordance with Thorogood's model, it is possible that the expression of *and* genes are initiated at the stage when the fin fold begins to replace the AER after the specification of the large endoskeleton. Thorogood's concept also predicts that the complete loss of the fin fold was permissive for the development of a limb with digits instead of rays (Thorogood, 1991). Similarly, the loss of the *and* genes are associated with the evolutionary loss of the fin fold and the evolutionary gain of limb morphology (Zhang et al., 2010). It was previously shown that actinopterygians, including teleosts, and chondrichthyes, including sharks and rays, possessed *and* genes in their genomes. Here it is revealed that *and* genes are in the genome of the coelacanth. From these results, it is now evident that an aquatic sarcopterygian ancestor of the tetrapods possessed *and* genes. Based on Thorogood's model (Thorogood, 1991) and the evidence that the loss of *and* genes was involved in the fin-to-limb transition (Zhang et al., 2010), these results suggest that the loss of *and* genes from the sarcopterygian genome may have been involved with the evolution of a lobed-fin into a limb.

5.3.2. Synteny and putative ancestry of the and *gene loci*

Here I report that the collection of genes within a 500kb distance of teleost *and1* or *and4* are syntenic to chromosomal regions found in coelacanth and tetrapod genomes. The synteny of the *and1* locus is well conserved across the genomes of teleost species and the coelacanth. Despite the loss of *and* genes from the tetrapods, the neighboring genes of *and1* are syntenic in tetrapods. There is a low level of conserved synteny for *and2* among teleosts, and less of a syntenic pattern can be resolved between teleost *and2* and sarcopterygian genomes. This pattern suggests that the synteny of the *and1* locus has been conserved since the common ancestor of the actinopterygians and the sarcopterygians. The genes *and1* and *and2* are likely "ohnologs", meaning their single ancestral gene was duplicated during a whole genome duplication event (Ohno, 1970; Wolfe, 2000) (further discussed below). This implies that their original synteny was also once identical, though over the course of evolution they may have diverged. While the synteny of the genes immediately surrounding *and1* has been conserved, it appears that the synteny of the genes immediately surrounding *and2* diverged from this ancestral locus.

The synteny for the *and4* locus is well conserved in teleosts. The neighboring genes to *and4,* including *cops6, mcm7* and *stag3l3,* are syntenic to loci in the genomes of the coelacanth and a variety of tetrapod species. Overall, the synteny of the *and4* locus is well preserved throughout the vertebrates.

The evolution of the synteny of the *and3* locus is less clear. The zebrafish *and3* locus is partially syntenic to those sarcopterygian loci which are also syntenic to *and4*. Specifically, the gene *ap4m1* neighbors zebrafish *and3,* it does not neighbor zebrafish *and4,* although it is found in proximity to *cops6, mcm7,* and *stag3l3* in sarcopterygian genomes (Fig 5.9). As mentioned for *and1* and *and2,* a similarity in synteny for *and3* and *and4* is expected because they are likely "ohnologs", and therefore were derived from a single ancestral locus. While teleost *and4* and

and3 do not display synteny to each other, they are both syntenic to a single locus of the sarcopterygian genomes. This implies that the sarcopterygian locus resembles the ancestral locus of both teleost *and3* and *and4.* The teleost *and3* and *and4* loci have diverged, although they both differentially retained some synteny to their common ancestral locus.

The *and3* loci of the Acanthomorpha teleosts are syntenic, and nearly identical, to each other. This is likely due to their close phylogenetic relationship. However, the neighboring genes to the Acanthomorpha *and3* gene are not consistently syntenic to a single locus in other species. This is unlike the scenario of *and4* for which the synteny is well conserved in zebrafish, Acanthomorpha, coelacanth, and tetrapod genomes. The genomic coordinates of the coelacanth 'small' *and* gene is not yet defined. Based on the high degree of synteny for the *and4* locus across vertebrates, and the low degree of synteny for the *and3* locus, I predict that the coelacanth 'small' *and* gene is found in the region that is syntenic to *and4* – between the genes *cops6* and *mcm7.* Further sequencing of the coelacanth genome would verify the precise location of the coelacanth 'small' *and* gene. Furthermore, it appears that the synteny of the *and4* locus predates the split of the sarcopterygians and the actinopterygians, therefore I propose that the locus, which contains *cops6, mcm7, stag3l3* and the 'small' *and* gene, represents the ancestral locus of the 'small' *and* gene.

It has been reported that the coelacanth morphology has not changed much over the last 150-65 million years, and its genome has not been altered much since (Amemiya et al., 2010; Noonan et al., 2004). This implies an ancestral quality to the coelacanth genome. The *and* loci in coelacanth may closely resemble the synteny of the ancestral osteichthyes, whereas the synteny of *and* genes in other extant species may not be as closely related to the ancestral state. This is likely the case for the synteny of the *and1* region which has an ancestral appearance. However,

the data supports that *Xenopus* has retained a stronger degree of the ancestral state for the 'small' *and* locus. Within a close vicinity on the chromosome, the *Xenopus* locus GL173561.1 has three genes which are syntenic to zebrafish *and3,* including *per1, vamp2, ap4m1,* and four genes which are syntenic to zebrafish *and4,* including *cops6, mcm1, stag3l3,* and *aurkb* (Fig 5.9). The ancestral 'small' *and* locus may have possessed these seven genes which were differentially deleted or translocated in the genomes of extant species.

The retention of conserved synteny has been argued as a means to preserve functional regulatory networks within the chromosome (Kikuta et al., 2007; Lang et al., 2010). The conserved synteny of the genes *adipoq2, myeov2,* and *otos* of the *and1* locus, or the conserved synteny of the genes *cops6, stag3l3, mcm7* from the *and4* locus, in a wide range of species suggests that a given function, for example regulatory elements, within these loci have been positively selected. Using mVISTA I showed that there were no highly detectable CNEs in the syntenic region of *and1*. Despite the lack of CNEs, it has been documented that regulatory elements may conserve their function regardless of undergoing sequence changes throughout evolution (Taher et al., 2011). In the syntenic loci observed, it is possible that the regulatory functions were maintained, though their sequences may have been altered. In tetrapods, these regulatory elements may have been retained, though the *and* genes were excluded.

5.3.3. Whole genome duplications during vertebrate evolution

Two rounds of whole genome duplications occurred at the base of the vertebrate lineage prior to the split of the osteichthyes and chondrichthyes (Dehal and Boore, 2005; McLysaght et al., 2002; Ohta, 1989; Panopoulou et al., 2003; Sidow, 1996). The first round is referred to as 1R and the second round is referred to as 2R and occurred as late as 525 mya (Vandepoele et al.,

2004). One additional round of a whole genome duplication, referred to as 3R, occurred at the base of the teleost lineage approximately 400-300 mya (Meyer and Schartl, 1999; Postlethwait et al., 1998; Taylor et al., 2003; Vandepoele et al., 2004). As a result, the genomes of the sarcopterygians, including the lobe-finned fish and the tetrapods, possess only half as many genes compared to the teleost fishes (Meyer and Van de Peer, 2005). Following 1R or 2R, the ancestral osteichthyes had two copies of *and* genes. These two ancestral copies were inherited by the sarcopterygians, whereas teleosts duplicated these two copies resulting in four *and* paralogs. Here I identified a total of two coelacanth *and* genes. This indicates that at the teleost-specific whole genome duplication, *and1* and *and2* were duplicated from the ancestral 'large' *and* gene, while *and3* and *and4* were duplicated from the ancestral 'small' *and* gene. In contrast, Amemiya et al. (2013) reported that there is only one *and* gene in the coelacanth genome and suggested that the zebrafish *and 1-4* genes originated from a single gene at the teleost-specific genome duplication (Amemiya et al., 2013).

The analysis by Amemiya et al. (2013) suggests that all *and* genes were derived from one single ancestral locus just prior to the teleost radiation. The notion of a single ancestral *and* gene is likely true, however I propose that a species with a single *and* gene existed much earlier in history, around the time of the 1R or 2R genome duplication events. There remains the possibility that synteny exists in all *and1-4* paralogs; a synteny which dates back to the single copy gene at the time of the 1R or 2R events. For example the genes *stag3l3* and *tmem128* are syntenic to *and4* while paralogs, *stag1* or *tmem22a*, are syntenic to *and1*. Genes which are syntenic to all four *and* paralogs is a rare occurrence and it appears that over the course of evolution very few have remained.

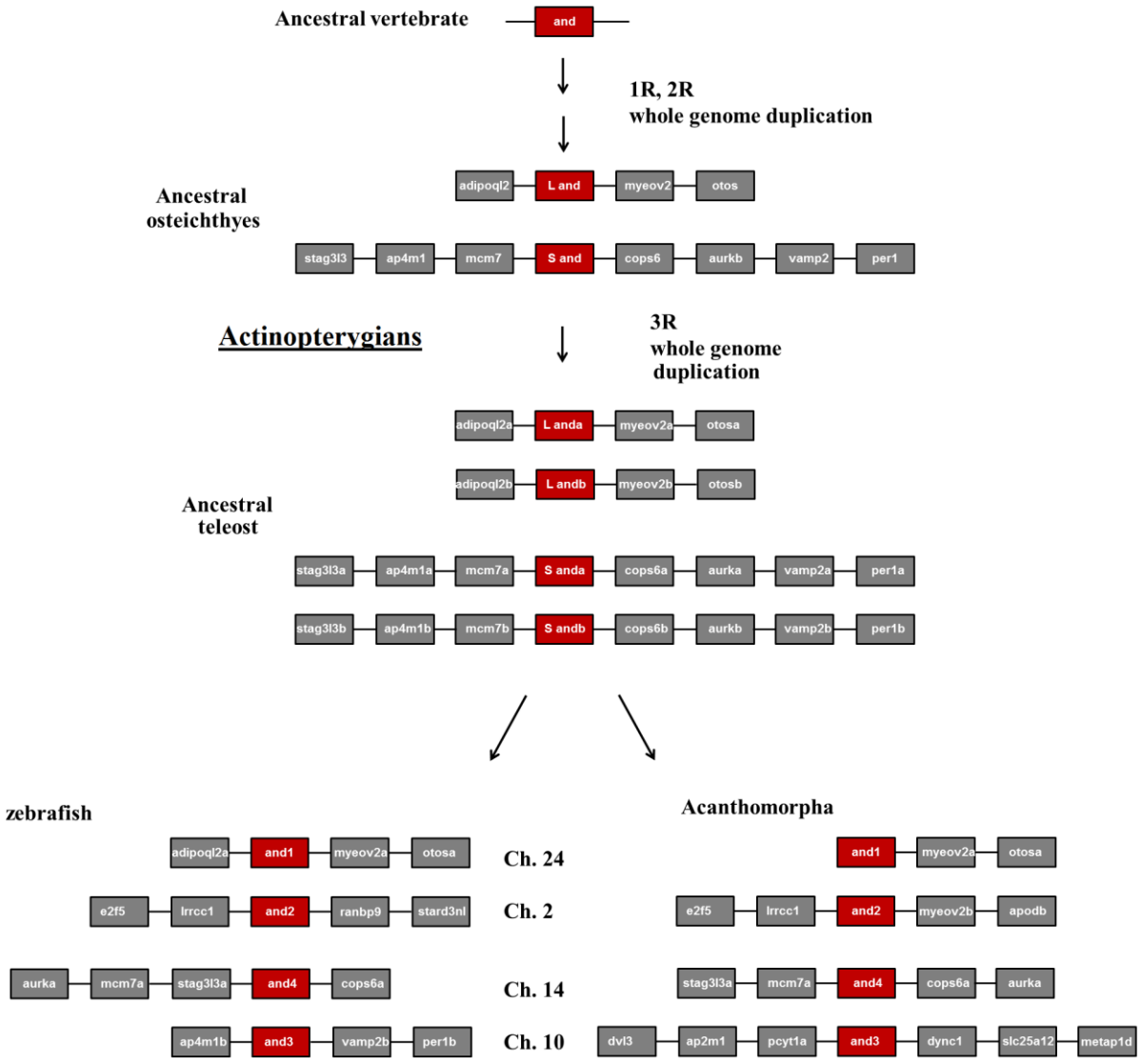


Fig 5.11. Conservation of syntenic genes to the *and* loci of the ancestral vertebrates and their direct ancestry to the extant teleosts. Represented at the top is the theoretical ancestral vertebrate genome that had one *actinodin* (and) gene. Two *and* genes resulted from the early vertebrate whole genome duplications (1R and 2R). In the ancestral osteichthyes, the *and* genes had distinct syntenies. The ancestral synteny was hypothesized based on the similarities in synteny of extant sarcopterygians and teleost species. Four *and* genes subsequently resulted from

the teleost-specific whole genome duplication (3R). The theoretical ancestral teleost had two identical copies of the synteny of each ancestral osteichthyes *and* gene. Over the course of teleost evolution, zebrafish and Acanthomorpha show the conservation of one copy of each ancestral *and* synteny, presently identified as the *and1* and *and4* loci. The synteny of the remaining teleost *and* genes, *and2* and *and3* diverged from the ancestral state. 'Large' *and* (L and) and 'small' *and* (S and) are indicated on hypothetical loci of extinct species.

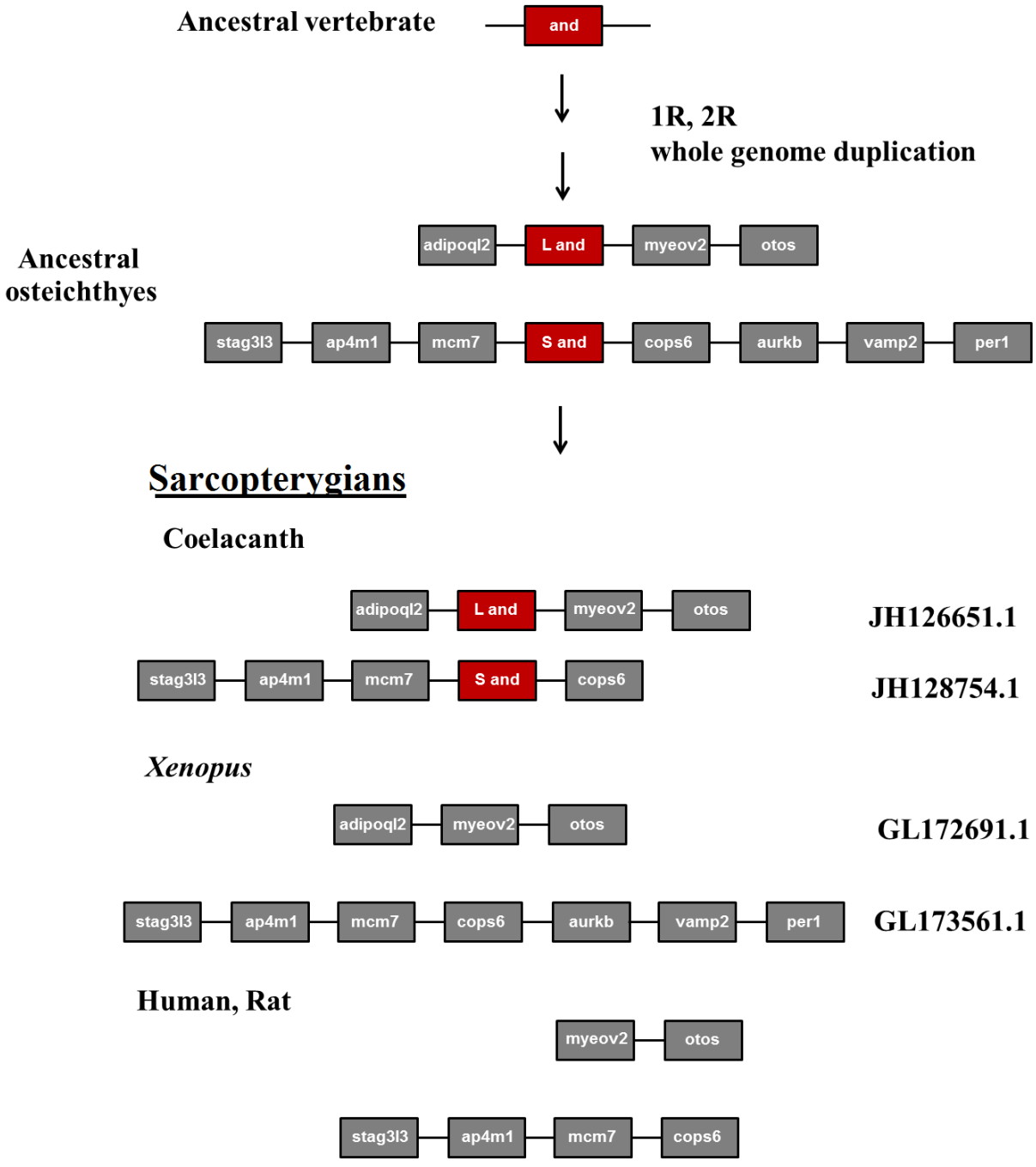


Fig 5.12. Conservation of syntenic genes to the *and* loci of the ancestral vertebrates and their direct ancestry to the extant sarcopterygians. The coelacanth is currently the only known sarcopterygian with *and* genes in its genome. The synteny of the two coelacanth *and*

genes resembles the theoretical ancestral synteny of these genes. The ancestral synteny was hypothesized based on the similarities in synteny of extant sarcopterygians and teleost species. The tetrapods, or limbed species, retain fragments of the conserved synteny, however the *and* genes were aberrantly lost from the tetrapod loci. ‘Large’ *and* (L and) and ‘small’ *and* (S and) are indicated on loci with no previously characterized *and* genes.

Figures 5.11 and 5.12 show schematic models that describe the progression of the *and* gene synteny throughout the evolution of the vertebrates (Fig 5.11, 5.12). This progression begins with the early vertebrates at the time of the 1R and 2R events, and the origin of the 'large' and 'small' *and* gene syntenies. The 'large' *and* gene synteny remained largely unchanged throughout vertebrate evolution. In tetrapods the *and* gene was lost from this locus. Following the teleost-specific whole genome duplication, theoretically two identical copies of each loci was present in the ancestral teleost genome. In teleosts the *and1* locus resembles the ancestral locus and the *and2* locus diverged from the ancestral synteny. Figure 5.11 also shows the most parsimonious scenario that explains the origins of the teleost *and3* and *and4* loci, and figure 5.12 shows the origin of the tetrapod synteny to the *and4* locus. The ancestral 'small' *and* locus is portrayed as being similar to the *Xenopus* locus GL173561.1 with the addition of the *and* gene (Figs 5.11, 5.12). Both the zebrafish and the Acanthomorpha species retained the ancestral synteny at their *and4* loci. In zebrafish, *and3* is syntenic to the ancestral neighboring gene, *ap4m1*. The Acanthomorpha *and3* locus is greatly diverged. The coelacanth 'small' *and* locus presumably retained synteny to the ancestral locus, as did the other tetrapods although with the specific loss of the 'small' *and* gene (Fig 5.12).

5.3.4. Tetrapod-specific gene loss

The earliest known aquatic tetrapod originated over 350 mya in the Devonian (Clack, 2012; Jarvik, 1955). The loss of the *and* genes from the tetrapod ancestor would have occurred just prior to or during this time. It seems improbable that the *and* genes were individually randomly deleted during the same historical period. It seems more likely that a specific biological function which affected both *and* paralogs caused their specific losses from otherwise

well conserved syntenic loci. One possibility is that the ancestral tetrapod species experienced the loss of a transcription factor that positively regulates *and* gene expression in the fins of fish. Another possibility is that *and* genes may have been downregulated by an inhibitory transcription factor that gained a new expression pattern in the ancestral tetrapod.

Pseudogenization may occur when a gene is no longer expressed and becomes subject to mutations and eventual complete loss (Li et al., 1981; Zhang et al., 2010b). I propose that the loss of *and* genes may have occurred by pseudogenization which was caused not by the loss of a promoter signal (Wang et al., 2006) rather it was through the gain of an inhibitory signal. Candidate transcription factors that downregulate *and* genes are Hoxa13 or Hoxd13 protein family members which played important roles in the developmental evolution of the endoskeleton during the fin to limb transition (discussed in Objective 1). Hox proteins can act as both transcriptional activators and repressors (Mann et al., 2009), and inhibitory functions have been reported for members of the Hoxa and Hoxd protein families (Kataoka et al., 2001; Shen et al., 2001). In the coelacanth, it is likely that these transcription factors are active for a long duration, resulting in the expansion of the endoskeleton as well as the temporal inhibition of fin fold formation. Based on Thorogood's model, it is likely that the endoskeleton-specific transcription factors are not active throughout lobed-fin development, and eventually their inhibition of *and* genes are de-activated and the fin fold develops at a later stage. In the tetrapods, it has been shown that novel regulatory elements account for a distal expansion of factors such as Hoxd13 (Schneider et al., 2011). Increased Hoxd13 expression in the ancestral tetrapod may have caused a permanent inhibition of *and* expression which lead to the loss of the fin fold and the pseudogenization of *and* genes. The fossil record shows polydactyly in the earliest known tetrapod species (Coates and Clack, 1990) and consistently, double-knockdown of

zebrafish *and1* and *and2* results in a fin with a gene expression pattern reminiscent of limbs in polydactylous mouse mutants (Zhang et al., 2010). In these morphant embryos there is an expansion of the Hoxd13a expression domain. Ultimately, the factors that lead to digit evolution may have acted synonymously with the downregulation of *and* gene expression.

It should also be considered that the opposite scenario is possible; that the loss of *and* expression is a cause rather than an effect for the expansion of transcription factor expression in the developing limb. The loss of *and* genes and the loss of the fin fold may have provided the opportunity for transcription factors in the limb to take on new functions. This scenario would imply that the loss of both *and* genes was not the result of a transcription factor that simultaneously inhibited both genes. Instead, another mechanism caused the loss of both *and* genes during the same evolutionary period. A possible explanation of this alternate scenario can be provided by considering the theories circulating ohnolog evolution. Ohno (1970) theorized that deleting an ohnolog should not be detrimental to the organism because there is a backup copy of the gene (Ohno, 1970). It has also been theorized that a copied gene should be able to acquire a new function without compromising the function of the other copy (Taylor and Raes, 2004). Additionally, duplicated genes may be preserved by subfunctionalization (Force et al., 1999), for example duplicated genes are expressed in different tissues at different times (Dover, 1882; Baltscheffshy and Ornval, 1986; Ohta, 1989). Conversely, Makino and McLysaght (2010) and Papp (2003) disagree with these earlier theories (Makino and McLysaght, 2010; Papp et al., 2003). These more recent studies propose that if the duplicated gene is part of a signaling pathway, or any polygenic process, then all gene components of the pathway were duplicated too. Deleting only one ohnolog member of the pathway would cause an imbalance in the dosage of the whole pathway (Makino and McLysaght, 2010). This concept is supported by the finding

that a significantly large percentage of ohnologs do not undergo single gene duplications or single gene deletions over the course of evolution (Makino and McLysaght, 2010; Papp et al., 2003).

A dosage effect related to losing the *and* gene function has already been observed. Zhang et al. (2010) performed a double-knockdown in zebrafish embryos of *and1* and *and2*, corresponding to both copies of the 'large' *and* gene. The phenotype of this knockdown provided insight toward the loss of *and* genes and the fin-to-limb transition. In these experiments, the 'small' *and* genes, *and3* and *and4* remained intact. In other words, loss of only the 'large' *and* genes is sufficient for the observed phenotype. In the ancestor to the tetrapod, a random deletion of the 'single' large *and* gene may have caused alterations in appendage development similar to the observations by Zhang et al. (2010). Subsequently, the remaining *and* gene in this early tetrapod genome experienced a dosage effect, such that it alone could not contribute to fin fold development. This would result in the loss of function for the 'small' *and* gene, and over time it may have accumulated many mutations without any additional phenotypic effects. The idea presented by Makino and McLysaght (2010) and Papp (2003) would support that the loss of one *and* gene is unlikely, although the loss of both during the same evolutionary period is more consistent with the observed patterns of ohnolog evolution. Future work can aim to knockdown both *and3* and *and4* in zebrafish embryos to determine whether the 'small' *and* genes can induce the same dosage effect observed for the 'large' *and* genes.

5.4. Conclusion and future directions

The identification of *and* genes in the coelacanth genome supports the original conclusion that *and* genes were lost specifically from the tetrapods. In this study I show that conserved

synteny can be used to trace the ancestry of the *and* genes. The conservation of this synteny dates approximately 450 mya (Lecointre and Guyader, 2006; Xiaobo and Min, 2010) to the common ancestor of the actinopterygians and sarcopterygians. I also propose that both ancestral *and* paralogs were derived from a single *and* gene at the origin of the vertebrates. Fins with rays have been reported in some of the earliest known fossils of vertebrate species (Zhang and Hou, 2004), and it is possible that these fin rays were the result of *and* gene expression and fin fold development in these ancient species. Future research can be directed towards uncovering *and* genes and their synteny in basal vertebrate species such as the lamprey (Smith et al., 2013), or in other extant aquatic sarcopterygians such as the lungfish. The presence of *and* genes in these genomes may strengthen the correlation between fin ray development and *and* gene expression.

In the results above, it is mentioned that the full sequence of the coelacanth 'small' *and* gene is likely not represented in its entirety. Future genomic analyses can aim to uncover additional exons of the coelacanth 'small' *and* gene which are possibly located in other poorly characterized scaffolds of the coelacanth genome. I proposed that the coelacanth 'small' *and* gene is located in between the genes *mcm7* and *cops6* on the scaffold JH128754.1. Analyses of the DNA sequences flanking coelacanth *mcm7* and *cops6* genes did not reveal any similar sequences to the *and* genes (results not shown). Future work can aim to complete the sequencing of the coelacanth genomic DNA located in between the genes *mcm7* and *cops6* to confirm whether the 'small' *and* gene is situated in this region.

Knowledge of *and* genes in novel species expands our current understanding of the phylogenetic distribution of these genes, and their correlations with fin development. The concomitant loss of fin rays and the *and* genes from tetrapods provides an interesting story in the field of evolutionary developmental biology. Many of the direct developmental outcomes caused

by the loss of actinotrichia or fin rays are yet to be elucidated. This specific area of study will prove to be a valuable component in the narrative of fin to limb evolution. Authors have proposed that during evolution, sometimes "gene loss can be adaptive" (Olson, 1999; Olson and Varki, 2003). The implication that the loss of *and* genes from tetrapods contributed to the evolution of fins into limbs certainly attests to the idea that "less is more".

References

Abe, G., Ide, H. and Tamura, K. (2007). Function of FGF signaling in the developmental process of the median fin fold in zebrafish. *Developmental biology* **304**, 355–66.

Agarwal, P., Wylie, J. N., Galceran, J., Arkhitko, O., Li, C., Deng, C., Grosschedl, R. and Bruneau, B. G. (2003). Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development (Cambridge, England)* **130**, 623–33.

Ahn, D. and Ho, R. K. (2008). Tri-phasic expression of posterior Hox genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Developmental biology* **322**, 220–33.

Ahn, K., Mishina, Y., Hanks, M. C., Behringer, R. R. and Crenshaw, E. B. (2001). BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. *Development (Cambridge, England)* **128**, 4449–61.

Akimenko, M. A., Ekker, M., Wegner, J., Lin, W. and Westerfield, M. (1994). Combinatorial expression of three zebrafish genes related to distal-less: part of a homeobox gene code for the head. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **14**, 3475–86.

Akimenko, M.-A., Marí-Beffa, M., Becerra, J. and Géraudie, J. (2003). Old questions, new tools, and some answers to the mystery of fin regeneration. *Developmental dynamics : an official publication of the American Association of Anatomists* **226**, 190–201.

Albalat, R., Martí-Solans, J. and Cañestro, C. (2012). DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. *Briefings in functional genomics* **11**, 142–55.

Altabef, M., Clarke, J. D. and Tickle, C. (1997). Dorso-ventral ectodermal compartments and origin of apical ectodermal ridge in developing chick limb. *Development (Cambridge, England)* **124**, 4547–56.

Amemiya, C. T., Powers, T. P., Prohaska, S. J., Grimwood, J., Schmutz, J., Dickson, M., Miyake, T., Schoenborn, M. A., Myers, R. M., Ruddle, F. H., et al. (2010). Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 3622–7.

Amemiya, C. T., Alföldi, J., Lee, A. P., Fan, S., Philippe, H., Maccallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–6.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science (New York, N.Y.)* **282**, 1711–4.

Andrade, M. A., Perez-Iratxeta, C. and Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *Journal of structural biology* **134**, 117–31.

Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K. and Nishida, M. (2008). Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC evolutionary biology* **8**, 215.

B Thisse, S Pflumio, M Fürthauer, B Loppin, V Heyer, A Degrave, R Woehl, A Lux, T Steffan, XQ Charbonnier, and C. T. (2001). Expression of the zebrafish genome during embryogenesis. *ZFIN Direct Data Submission*.

Baird, A., Coimbra, R., Dang, X., Lopez, N., Lee, J., Krzyzaniak, M., Winfield, R., Potenza, B. and Eliceiri, B. P. (2012). Cell surface localization and release of the candidate tumor suppressor Ecrg4 from polymorphonuclear cells and monocytes activate macrophages. *Journal of leukocyte biology* **91**, 773–81.

Barrow, J. Wnt/planar cell polarity signaling: an important mechanism to coordinate growth and patterning in the limb. *Organogenesis* **7**, 260–6.

Becerra, J., Montes, G. S., Bexiga, S. R. and Junqueira, L. C. (1983). Structure of the tail fin in teleosts. *Cell and tissue research* **230**, 127–37.

Begemann, G., Schilling, T. F., Rauch, G. J., Geisler, R. and Ingham, P. W. (2001). The zebrafish neckless mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. *Development (Cambridge, England)* **128**, 3081–94.

Bénazet, J.-D. and Zeller, R. (2009). Vertebrate limb development: moving from classical morphogen gradients to an integrated 4-dimensional patterning system. *Cold Spring Harbor perspectives in biology* **1**, a001339.

Bouvet, J. (1974). Différenciation et ultrastructure du squelette distal de la nageoire pectorale chez la truite indigène (salmo trutta fario L.) I. différenciation et ultrastructure des actinotriches. *Arch. d'Anat. Micr.* **63**, 79–96.

Bouvet, J. (1978). Cell proliferation and morphogenesis of the apical ectodermal ridge in the pectoral fin bud of the trout embryo (Salmo trutta fario L.). *Wilhelm Roux's Archives of Developmental Biology* **185**, 137–154.

Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D.G., Knight, J., Mani, P., Martin, R., Moxon, S.A., Paddock, H., Pich, C., Ramachandran, S., Ruef, B.J., Ruzicka, L., Bauer Schaper, H., Schaper, K., Shao, X., Singer, A., Sprague, J., Sp, M.

(2011). ZFIN: enhancements and updates to the zebrafish model organism database. *Nucleic Acids Res.* **39**,.

Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238**, 115–34.

Büscher, D., Bosse, B., Heymer, J. and Rüther, U. (1997). Evidence for genetic control of Sonic hedgehog by Gli3 in mouse limb development. *Mechanisms of development* **62**, 175–82.

C Nüsslein-Volhard and R Dahm ed. (2002). *Zebrafish: a practical approach*. New York: Oxford University Press.

Callier, V., Clack, J. A. and Ahlberg, P. E. (2009). Contrasting developmental trajectories in the earliest known tetrapod forelimbs. *Science (New York, N.Y.)* **324**, 364–7.

Camões, M. J., Paulo, P., Ribeiro, F. R., Barros-Silva, J. D., Almeida, M., Costa, V. L., Cerveira, N., Skotheim, R. I., Lothe, R. A., Henrique, R., et al. (2012). Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma. *PloS one* **7**, e49819.

Cañestro, C., Catchen, J. M., Rodríguez-Marí, A., Yokoi, H. and Postlethwait, J. H. (2009). Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS genetics* **5**, e1000496.

Capdevila, J. and Belmonte, J. C. (2001). Patterning mechanisms controlling vertebrate limb development. *Annual review of cell and developmental biology* **17**, 87–132.

Carney, T. J., Feitosa, N. M., Sonntag, C., Slanchev, K., Kluger, J., Kiyozumi, D., Gebauer, J. M., Coffin Talbot, J., Kimmel, C. B., Sekiguchi, K., et al. (2010). Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as potential novel fraser syndrome disease genes. *PLoS genetics* **6**, e1000907.

Catchen, J. M., Conery, J. S. and Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome research* **19**, 1497–505.

Chandross, R. J. (1982). Structure and packing of dry elastoidin: a collagen phase change. *Collagen and related research* **2**, 331–48.

Chandross, R. J. and Bear, R. S. (1979). Comparison of mammalian collagen and elasmobranch elastoidin fiber structures, based on electron density profiles. *Journal of molecular biology* **130**, 215–9.

Chitramuthu, B. P. and Bennett, H. P. J. (2011). Use of zebrafish and knockdown technology to define proprotein convertase activity. *Methods in molecular biology (Clifton, N.J.)* **768**, 273–96.

Clack, J. A. (2012). *Gaining Ground: The Origin and Evolution of Tetrapods*. 2nd ed. Bloomington, Indiana: Indiana University Press.

Coates, M. I. and Clack, J. A. (1990). Polydactyly in the earliest known tetrapod limbs. *Nature* **347**, 66–69.

Collar, D. C., Wainwright, P. C. and Alfaro, M. E. (2008). Integrated diversification of locomotion and feeding in labrid fishes. *Biology letters* **4**, 84–6.

Conant, G. C. and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature reviews. Genetics* **9**, 938–50.

Crotwell, P. L., Sommervold, A. R. and Mabee, P. M. (2004). Expression of bmp2a and bmp2b in late-stage zebrafish median fin development. *Gene expression patterns : GEP* **5**, 291–6.

Dahn, R. D., Davis, M. C., Pappano, W. N. and Shubin, N. H. (2007). Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature* **445**, 311–4.

Damodaran, M., Sivaraman, C. and Dhavalikar, R. S. (1956). Amino acid composition of elastoidin. *The Biochemical journal* **62**, 621–5.

Dane, P. J. and Tucker, J. B. (1985). Modulation of epidermal cell shaping and extracellular matrix during caudal fin morphogenesis in the zebra fish Brachydanio rerio. *Journal of embryology and experimental morphology* **87**, 145–61.

Dang, X., Podvin, S., Coimbra, R., Eliceiri, B. and Baird, A. (2012). Cell-specific processing and release of the hormone-like precursor and candidate tumor suppressor gene product, Ecrg4. *Cell and tissue research* **348**, 505–14.

Davis, M. C., Dahn, R. D. and Shubin, N. H. (2007). An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. *Nature* **447**, 473–6.

De Iuliis and Pulera (2011). *The Dissection of Vertebrates, Second Edition*. Burlington, MA: Elsevier Inc.

De Jong, R., Van der Heijden, J. and Meijlink, F. (1993). DNA-binding specificity of the S8 homeodomain. *Nucleic acids research* **21**, 4711–20.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988). The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *The EMBO journal* **7**, 4203–12.

Dehal, P. and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS biology* **3**, e314.

Doherty, P., Wassersug, R. and Lee, J. (1998). Mechanical properties of the tadpole tail fin. *The Journal of experimental biology* **201 (Pt 19**, 2691–9.

Duboc, V. and Logan, M. P. O. (2011). Regulation of limb bud initiation and limb-type morphology. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 1017–27.

Duguay, S. J., Lai-Zhang, J. and Steiner, D. F. (1995). Mutational analysis of the insulin-like growth factor I prohormone processing site. *The Journal of biological chemistry* **270**, 17566–74.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Durán, I., Marí-Beffa, M., Santamaría, J. A., Becerra, J. and Santos-Ruiz, L. (2011). Actinotrichia collagens and their role in fin formation. *Developmental biology* **354**, 160–72.

Fayez F. Safadi, Mary F. Barbe, Samir M. Abdelmagid, Mario C. Rico, R. A. A. and Judith Litvin, and S. N. P. (2009). Chapter 1 Bone Structure, Development and Bone Biology. In *Bone Pathology*, pp. 1–50.

Fernandez-Teran, M. and Ros, M. A. (2008). The Apical Ectodermal Ridge: morphological aspects and signaling pathways. *The International journal of developmental biology* **52**, 857–71.

Flores, M. V. C., Hall, C. J., Davidson, A. J., Singh, P. P., Mahagaonkar, A. A., Zon, L. I., Crosier, K. E. and Crosier, P. S. (2008). Intestinal differentiation in zebrafish requires Cdx1b, a functional equivalent of mammalian Cdx2. *Gastroenterology* **135**, 1665–75.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–45.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic acids research* **32**, W273–9.

Freitas, R., Gómez-Marín, C., Wilson, J. M., Casares, F. and Gómez-Skarmeta, J. L. (2012). Hoxd13 contribution to the evolution of vertebrate appendages. *Developmental cell* **23**, 1219–29.

G Dover, R. F. ed. (1882). *Genome evolution*. London: Academic Press.

Galceran, J., Fariñas, I., Depew, M. J., Clevers, H. and Grosschedl, R. (1999). Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. *Genes & development* **13**, 709–17.

Geraudie, J. and Meunier, F. J. (1980). Elastoidin actinotrichia in Coelacanth fins: a comparison with teleosts. *Tissue & cell* **12**, 637–45.

Géraudie, J. (1977). Initiation of the actinotrichial development in the early fin bud of the fish, Salmo. *Journal of morphology* **151**, 353–61.

Géraudie, J. (1978). Scanning electron microscope study of the developing trout pelvic fin bud. *The Anatomical record* **191**, 391–4.

Géraudie, J. (1988). Fine structural peculiarities of the pectoral fin dermoskeleton of two brachiopterygii, Polypterus senegalus and Calamoichthys calabaricus (Pisces, Osteichthyes). *The Anatomical record* **221**, 455–68.

Géraudie, J. and Landis, W. J. (1982). The fine structure of the developing pelvic fin dermal skeleton in the trout Salmo gairdneri. *The American journal of anatomy* **163**, 141–56.

Gilbert, W., De Souza, S. J. and Long, M. (1997). Origin of genes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7698–703.

Gonzalez, A. M., Podvin, S., Lin, S.-Y., Miller, M. C., Botfield, H., Leadbeater, W. E., Roberton, A., Dang, X., Knowling, S. E., Cardenas-Galindo, E., et al. (2011). Ecrg4 expression and its product augurin in the choroid plexus: impact on fetal brain development, cerebrospinal fluid homeostasis and neuroprogenitor cell response to CNS injury. *Fluids and barriers of the CNS* **8**, 6.

Goodrich, E. S. (1904). On the dermal fin-rays of fishes-living and extinc. *Quart. J. Microscop. Sci. N.S.* **47**, 465–522.

Grandel, H. and Schulte-Merker, S. (1998). The development of the paired fins in the zebrafish (Danio rerio). *Mechanisms of development* **79**, 99–120.

Grandel, H., Lun, K., Rauch, G.-J., Rhinn, M., Piotrowski, T., Houart, C., Sordino, P., Küchler, A. M., Schulte-Merker, S., Geisler, R., et al. (2002). Retinoic acid signalling in the zebrafish embryo is necessary during pre-segmentation stages to pattern the anterior-posterior axis of the CNS and to induce a pectoral fin bud. *Development (Cambridge, England)* **129**, 2851–65.

Gross, J. and Dumsha, B. (1958). Elastoidin: a twocomponent member of the collagen class. *Biochimica et biophysica acta* **28**, 268–70.

Gurnett, C. A., Bowcock, A. M., Dietz, F. R., Morcuende, J. A., Murray, J. C. and Dobbs, M. B. (2007). Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *American journal of medical genetics. Part A* **143**, 27–32.

Hall, B. K. ed. (2007). *Fins into limbs: evolution, development, and transformation*. Chicago, Il: The University of Chicago Press.

Harfe, B. D. (2011). Keeping up with the zone of polarizing activity: New roles for an old signaling center. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 915–9.

Hatta, K., Bremiller, R., Westerfield, M. and Kimmel, C. B. (1991). Diversity of expression of engrailed-like antigens in zebrafish. *Development (Cambridge, England)* **112**, 821–32.

Hawkins, T. A., Cavodeassi, F., Erdélyi, F., Szabó, G. and Lele, Z. (2008). The small molecule Mek1/2 inhibitor U0126 disrupts the chordamesoderm to notochord transition in zebrafish. *BMC developmental biology* **8**, 42.

Hernández-Vega, A. and Minguillón, C. (2011). The Prx1 limb enhancers: targeted gene expression in developing zebrafish pectoral fins. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 1977–88.

Huh, Y. H., Ryu, J.-H., Shin, S., Lee, D.-U., Yang, S., Oh, K.-S., Chun, C.-H., Choi, J.-K., Song, W. K. and Chun, J.-S. (2009). Esophageal cancer related gene 4 (ECRG4) is a marker of articular chondrocyte differentiation and cartilage destruction. *Gene* **448**, 7–15.

Hui, C. C. and Joyner, A. L. (1993). A mouse model of greig cephalopolysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. *Nature genetics* **3**, 241–6.

Hukins, D. W., Woodhead-Galloway, J. and Knight, D. P. (1976). Molecular tilting in dried elastoidin and its implications for the structures of other collagen fibrils. *Biochemical and biophysical research communications* **73**, 1049–55.

J Baltscheffshy, HL Ornval, and R. R. ed. (1986). *Molecular evolution of life*. Cambridge: Cambridge University Press.

Jackman, W. R. and Stock, D. W. (2006). Transgenic analysis of Dlx regulation in fish tooth development reveals evolutionary retention of enhancer function despite organ loss. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19390–5.

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., et al. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–57.

Jarvik, E. (1955). The Oldest Tetrapods and Their Forerunners. *The Scientific Monthly* **80**, 141–154.

John J. Sambrook, D. D. W. R. (2001). *Molecular cloning: a laboratory manual*. Cold Spring Harbor laboratory Press, 2001.

JR Hinchliffe, D. J. (1980). *The Development of the Vertebrate Limb: An approach through experiment, genetics and evolution*. Oxford: Clarendon Press.

Kataoka, K., Yoshitomo-Nakagawa, K., Shioda, S. and Nishizawa, M. (2001). A set of Hox proteins interact with the Maf oncoprotein to inhibit its DNA binding, transactivation, and transforming activities. *The Journal of biological chemistry* **276**, 819–26.

Kawakami, Y., Capdevila, J., Büscher, D., Itoh, T., Rodríguez Esteban, C. and Izpisúa Belmonte, J. C. (2001). WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo. *Cell* **104**, 891–900.

Khatib, A.-M., Lahlil, R., Scamuffa, N., Akimenko, M.-A., Ernest, S., Lomri, A., Lalou, C., Seidah, N. G., Villoutreix, B. O., Calvo, F., et al. (2010). Zebrafish ProVEGF-C expression, proteolytic processing and inhibitory effect of unprocessed ProVEGF-C during fin regeneration. *PloS one* **5**, e11438.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engström, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome research* **17**, 545–55.

Kimura, S. and Kubota, M. (1966). Studies on elastoidin. I. Some chemical and physical properties of elastoidin and its components. *Journal of biochemistry* **60**, 615–21.

Kimura, S. and Kubota, M. (1969). Tyrosine derivatives in a structural protein, elastoidin. *Journal of biochemistry* **65**, 141–3.

Kodali, V. K., Gannon, S. A., Paramasivam, S., Raje, S., Polenova, T. and Thorpe, C. (2011). A novel disulfide-rich protein motif from avian eggshell membranes. *PloS one* **6**, e18187.

Kögel, T. and Gerdes, H.-H. (2010). Maturation of secretory granules. *Results and problems in cell differentiation* **50**, 1–20.

Koonin, E. V (2005). Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics* **39**, 309–38.

Kraus, P. and Lufkin, T. (2006). Dlx homeobox gene control of mammalian limb and craniofacial development. *American journal of medical genetics. Part A* **140**, 1366–74.

Krauss, S., Concordet, J. P. and Ingham, P. W. (1993). A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* **75**, 1431–44.

Krukenberg, C. F. W. (1885). Über die chemische Beschaffenheit der sog. hornfäden von Mustelus und über die Zusammensetzung der keratinösen Hüllen um die Eier von Scyllium stellate. *Mittheilungen Zool Stat Neapel* **6**, 286–296.

Kujuro, Y., Suzuki, N. and Kondo, T. (2010). Esophageal cancer-related gene 4 is a secreted inducer of cell senescence expressed by aged CNS precursor cells. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8259–64.

Kuninger, D., Kuns-Hashimoto, R., Nili, M. and Rotwein, P. (2008). Pro-protein convertases control the maturation and processing of the iron-regulatory protein, RGMc/hemojuvelin. *BMC biochemistry* **9**, 9.

Lang, M., Hadzhiev, Y., Siegel, N., Amemiya, C. T., Parada, C., Strähle, U., Becker, M.-B., Müller, F. and Meyer, A. (2010). Conservation of shh cis-regulatory architecture of the coelacanth is consistent with its ancestral phylogenetic position. *EvoDevo* **1**, 11.

Lecointre, G. and Guyader, H. Le (2006). *The Tree of Life: A phylogenetic Classification*. The President and Fellows of Harvard University.

Lee, G. S., Liao, X., Shimizu, H. and Collins, M. D. (2010). Genetic and pathologic aspects of retinoic acid-induced limb malformations in the mouse. *Birth defects research. Part A, Clinical and molecular teratology* **88**, 863–82.

Leussink, B., Brouwer, A., El Khattabi, M., Poelmann, R. E., Gittenberger-de Groot, A. C. and Meijlink, F. (1995). Expression patterns of the paired-related homeobox genes MHox/Prx1 and S8/Prx2 suggest roles in development of the heart and the forebrain. *Mechanisms of development* **52**, 51–64.

Li, W. H., Gojobori, T. and Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237–9.

Li, L.-W., Yu, X.-Y., Yang, Y., Zhang, C.-P., Guo, L.-P. and Lu, S.-H. (2009). Expression of esophageal cancer related gene 4 (ECRG4), a novel tumor suppressor gene, in esophageal cancer and its inhibitory effect on the tumor growth in vitro and in vivo. *International journal of cancer. Journal international du cancer* **125**, 1505–13.

Li, W., Liu, X., Zhang, B., Qi, D., Zhang, L., Jin, Y. and Yang, H. (2010). Overexpression of candidate tumor suppressor ECRG4 inhibits glioma proliferation and invasion. *Journal of experimental & clinical cancer research : CR* **29**, 89.

Lingham-Soliar, T. (2005a). Caudal fin allometry in the white shark Carcharodon carcharias: implications for locomotory performance and ecology. *Die Naturwissenschaften* **92**, 231–6.

Lingham-Soliar, T. (2005b). Caudal fin in the white shark, Carcharodon carcharias (Lamnidae): a dynamic propeller for fast, efficient swimming. *Journal of morphology* **264**, 233–52.

Lissitzky, J. C., Luis, J., Munzer, J. S., Benjannet, S., Parat, F., Chrétien, M., Marvaldi, J. and Seidah, N. G. (2000). Endoproteolytic processing of integrin pro-alpha subunits involves the redundant function of furin and proprotein convertase (PC) 5A, but not paired basic amino acid converting enzyme (PACE) 4, PC5B or PC7. *The Biochemical journal* **346 Pt 1**, 133–8.

Liu, J. K., Ghattas, I., Liu, S., Chen, S. and Rubenstein, J. L. (1997). Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation. *Developmental dynamics : an official publication of the American Association of Anatomists* **210**, 498–512.

Logan, C., Hornbruch, A., Campbell, I. and Lumsden, A. (1997). The role of Engrailed in establishing the dorsoventral axis of the chick limb. *Development (Cambridge, England)* **124**, 2317–24.

Long, M. (2000). A new function evolved from gene fusion. *Genome research* **10**, 1655–7.

Long, J. A. and Gordon, M. S. (2004). The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition. *Physiological and biochemical zoology : PBZ* **77**, 700–19.

Lv, J., Havlak, P. and Putnam, N. H. (2011). Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC bioinformatics* **12 Suppl 9**, S11.

Makino, T. and McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9270–4.

Mann, R. S., Lelli, K. M. and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Current topics in developmental biology* **88**, 63–101.

Marí-Beffa, M., Carmona, M. C. and Becerra, J. (1989). Elastoidin turn-over during tail fin regeneration in teleosts. A morphometric and radioautographic study. *Anatomy and embryology* **180**, 465–70.

McLysaght, A., Hokamp, K. and Wolfe, K. H. (2002). Extensive genomic duplication during early chordate evolution. *Nature genetics* **31**, 200–4.

Mercader, N. (2007). Early steps of paired fin development in zebrafish compared with tetrapod limb development. *Development, Growth & Differentiation* **49**, 421–437.

Mercader, N., Fischer, S. and Neumann, C. J. (2006). Prdm1 acts downstream of a sequential RA, Wnt and Fgf signaling cascade during zebrafish forelimb induction. *Development (Cambridge, England)* **133**, 2805–15.

Meyer, A. and Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current opinion in cell biology* **11**, 699–704.

Meyer, A. and Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**, 937–45.

Mirabeau, O., Perlas, E., Severini, C., Audero, E., Gascuel, O., Possenti, R., Birney, E., Rosenthal, N. and Gross, C. (2007). Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome research* **17**, 320–7.

Montes, G. S., Becerra, J., Toledo, O. M., Gordilho, M. A. and Junqueira, L. C. (1982). Fine structure and histochemistry of the tail fin ray in teleosts. *Histochemistry* **75**, 363–76.

Mullen, L. M., Bryant, S. V, Torok, M. A., Blumberg, B. and Gardiner, D. M. (1996). Nerve dependency of regeneration: the role of Distal-less and FGF signaling in amphibian limb regeneration. *Development (Cambridge, England)* **122**, 3487–97.

Muroi, S., Tanaka, H., Miyamoto, C. and Taniuchi, I. (2013). Cutting edge: fine-tuning of Thpok gene activation by an enhancer in close proximity to its own silencer. *Journal of immunology (Baltimore, Md. : 1950)* **190**, 1397–401.

Nelson, J. S. (1994). *Fishes of the World*. 3rd ed. John Wiley and Sons, Inc.

Nelson, C. E., Morgan, B. A., Burke, A. C., Laufer, E., DiMambro, E., Murtaugh, L. C., Gonzales, E., Tessarollo, L., Parada, L. F. and Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. *Development (Cambridge, England)* **122**, 1449–66.

Neumann, C. J., Grandel, H., Gaffield, W., Schulte-Merker, S. and Nüsslein-Volhard, C. (1999). Transient establishment of anteroposterior polarity in the zebrafish pectoral fin bud in the absence of sonic hedgehog activity. *Development (Cambridge, England)* **126**, 4817–26.

Ng, J. K., Kawakami, Y., Büscher, D., Raya, A., Itoh, T., Koth, C. M., Rodríguez Esteban, C., Rodríguez-León, J., Garrity, D. M., Fishman, M. C., et al. (2002). The limb identity gene Tbx5 promotes limb initiation by interacting with Wnt2b and Fgf10. *Development (Cambridge, England)* **129**, 5161–70.

Niederreither, K., Subbarayan, V., Dollé, P. and Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature genetics* **21**, 444–8.

Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P. and Dollé, P. (2000). Retinoic acid synthesis and hindbrain patterning in the mouse embryo. *Development (Cambridge, England)* **127**, 75–85.

Noonan, J. P., Grimwood, J., Danke, J., Schmutz, J., Dickson, M., Amemiya, C. T. and Myers, R. M. (2004). Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. *Genome research* **14**, 2397–405.

Norton, W. H. J., Ledin, J., Grandel, H. and Neumann, C. J. (2005). HSPG synthesis by zebrafish Ext2 and Extl3 is required for Fgf10 signalling during limb development. *Development (Cambridge, England)* **132**, 4963–73.

Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**, 205–17.

Nurminsky, D. I., Nurminskaya, M. V, De Aguiar, D. and Hartl, D. L. (1998). Selective sweep of a newly evolved sperm-specific gene in Drosophila. *Nature* **396**, 572–5.

Ohno, S. (1970). *Evolution by genome duplication*. New York, NY: Springer-Verlag.

Ohta, T. (1989). Role of gene duplication in evolution. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada* **31**, 304–10.

Olson, M. V (1999). When less is more: gene loss as an engine of evolutionary change. *American journal of human genetics* **64**, 18–23.

Olson, M. V and Varki, A. (2003). Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature reviews. Genetics* **4**, 20–8.

Opstelten, D. J., Vogels, R., Robert, B., Kalkhoven, E., Zwartkruis, F., De Laaf, L., Destrée, O. H., Deschamps, J., Lawson, K. A. and Meijlink, F. (1991). The mouse homeobox gene, S8, is expressed during embryogenesis predominantly in mesenchyme. *Mechanisms of development* **34**, 29–41.

Ozawa, A., Lick, A. N. and Lindberg, I. (2011). Processing of proaugurin is required to suppress proliferation of tumor cell lines. *Molecular endocrinology (Baltimore, Md.)* **25**, 776–84.

Padhi, B. K., Joly, L., Tellis, P., Smith, A., Nanjappa, P., Chevrette, M., Ekker, M. and Akimenko, M.-A. (2004). Screen for genes differentially expressed during regeneration of the zebrafish caudal fin. *Developmental dynamics : an official publication of the American Association of Anatomists* **231**, 527–41.

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehrach, H. (2003). New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome research* **13**, 1056–66.

Papp, B., Pál, C. and Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–7.

Parr, B. A. and McMahon, A. P. (1995). Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. *Nature* **374**, 350–3.

Pinsonneault, J., Florence, B., Vaessin, H. and McGinnis, W. (1997). A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. *The EMBO journal* **16**, 2032–42.

Pizette, S., Abate-Shen, C. and Niswander, L. (2001). BMP controls proximodistal outgrowth, via induction of the apical ectodermal ridge, and dorsoventral patterning in the vertebrate limb. *Development (Cambridge, England)* **128**, 4463–74.

Podvin, S., Gonzalez, A.-M., Miller, M. C., Dang, X., Botfield, H., Donahue, J. E., Kurabi, A., Boissaud-Cooke, M., Rossi, R., Leadbeater, W. E., et al. (2011). Esophageal cancer related gene-4 is a choroid plexus-derived injury response gene: evidence for a biphasic response in early and late brain injury. *PloS one* **6**, e24609.

Postlethwait, J. H. (2007). The zebrafish genome in context: ohnologs gone missing. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **308**, 563–77.

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**, 345–9.

Prasad, M. S. and Paulson, A. F. (2011). A combination of enhancer/silencer modules regulates spatially restricted expression of cadherin-7 in neural epithelium. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 1756–68.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V, et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (New York, N.Y.)* **317**, 86–94.

Rabinowitz, A. H. and Vokes, S. A. (2012). Integration of the transcriptional networks regulating limb morphogenesis. *Developmental biology* **368**, 165–80.

Riddle, R. D., Johnson, R. L., Laufer, E. and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–16.

Robert, B. (2007). Bone morphogenetic protein signaling in limb outgrowth and patterning. *Development, growth & differentiation* **49**, 455–68.

Rodriguez-Esteban, C., Schwabe, J. W., Peña, J. D., Rincon-Limas, D. E., Magallón, J., Botas, J. and Izpisúa Belmonte, J. C. (1998). Lhx2, a vertebrate homologue of apterous, regulates vertebrate limb outgrowth. *Development (Cambridge, England)* **125**, 3925–34.

Rubin, L. and Saunders, J. W. (1972). Ectodermal-mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of the ectodermal induction. *Developmental biology* **28**, 94–112.

Sabatier, R., Finetti, P., Adelaide, J., Guille, A., Borg, J.-P., Chaffanet, M., Lane, L., Birnbaum, D. and Bertucci, F. (2011). Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PloS one* **6**, e27656.

Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K.-I. and Imai, K. (2003). Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny. *Genetics* **165**, 235–42.

Santamaria, J. A., Santos Ruiz, L. and Becerra, J. (1996). An antiserum against ceratotrichia (selachian) recognizes actinotrichia in teleost regenerating fins. *The International journal of developmental biology* **Suppl 1**, 175S–176S.

Santamaría, J. A. and Becerra, J. (1991). Tail fin regeneration in teleosts: cell-extracellular matrix interaction in blastemal differentiation. *Journal of anatomy* **176**, 9–21.

Saunders, J. W. (1948). The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *Journal of Experimental Zoology* **108**, 363–403.

Saunders, Gasseling, Fleischmajer, B. (1968). Epithelial-mesenchymal interactions. 78–97.

Schneider, I., Aneas, I., Gehrke, A. R., Dahn, R. D., Nobrega, M. A. and Shubin, N. H. (2011). Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12782–6.

Seidah, N. G. and Chrétien, M. (1999). Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain research* **848**, 45–62.

Seidah, N. G. and Prat, A. (2012). The biology and therapeutic targeting of the proprotein convertases. *Nature reviews. Drug discovery* **11**, 367–83.

Seidah, N. G., Chrétien, M. and Day, R. (1994). The family of subtilisin/kexin like pro-protein and pro-hormone convertases: divergent or shared functions. *Biochimie* **76**, 197–209.

Shaterian, A., Kao, S., Chen, L., DiPietro, L. A., Coimbra, R., Eliceiri, B. P. and Baird, A. (2013). The candidate tumor suppressor gene Ecrg4 as a wound terminating factor in cutaneous injury. *Archives of dermatological research* **305**, 141–9.

Shen, W. F., Krishnan, K., Lawrence, H. J. and Largman, C. (2001). The HOX homeodomain proteins block CBP histone acetyltransferase activity. *Molecular and cellular biology* **21**, 7509–22.

Sheth, R., Grégoire, D., Dumouchel, A., Scotti, M., Pham, J. M. T., Nemec, S., Bastida, M. F., Ros, M. A. and Kmita, M. (2013). Decoupling the function of Hox and Shh in developing limb reveals multiple inputs of Hox genes on limb growth. *Development (Cambridge, England)* **140**, 2130–8.

Shubin, N., Tabin, C. and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–23.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Current opinion in genetics & development* **6**, 715–22.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., et al. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nature genetics* **45**, 415–21, 421e1–2.

Sordino, P., Van der Hoeven, F. and Duboule, D. (1995). Hox gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–81.

Southey, B. R., Amare, A., Zimmerman, T. A., Rodriguez-Zas, S. L. and Sweedler, J. V (2006). NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. *Nucleic acids research* **34**, W267–72.

Steiner, D. F. (1998). The proprotein convertases. *Current opinion in chemical biology* **2**, 31–9.

Steinke, D., Salzburger, W. and Meyer, A. (2006). Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *Journal of molecular evolution* **62**, 772–84.

Su, T., Liu, H. and Lu, S. (1998). [Cloning and identification of cDNA fragments related to human esophageal cancer]. *Zhonghua zhong liu za zhi [Chinese journal of oncology]* **20**, 254–7.

Suzuki, M., Satoh, A., Ide, H. and Tamura, K. (2007). Transgenic Xenopus with prx1 limb enhancer reveals crucial contribution of MEK/ERK and PI3K/AKT pathways in blastema formation during limb regeneration. *Developmental biology* **304**, 675–86.

Tabariès, S., Lapointe, J., Besch, T., Carter, M., Woollard, J., Tuggle, C. K. and Jeannotte, L. (2005). Cdx protein interaction with Hoxa5 regulatory sequences contributes to Hoxa5 regional expression along the axial skeleton. *Molecular and cellular biology* **25**, 1389–401.

Tabin, C. and Wolpert, L. (2007). Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes & development* **21**, 1433–42.

Tadross, J. A., Patterson, M., Suzuki, K., Beale, K. E., Boughton, C. K., Smith, K. L., Moore, S., Ghatei, M. A. and Bloom, S. R. (2010). Augurin stimulates the hypothalamo-

pituitary-adrenal axis via the release of corticotrophin-releasing factor in rats. *British journal of pharmacology* **159**, 1663–71.

Taher, L., McGaughey, D. M., Maragh, S., Aneas, I., Bessling, S. L., Miller, W., Nobrega, M. A., McCallion, A. S. and Ovcharenko, I. (2011). Genome-wide identification of conserved regulatory function in diverged sequences. *Genome research* **21**, 1139–49.

Takamatsu, N., Kurosawa, G., Takahashi, M., Inokuma, R., Tanaka, M., Kanamori, A. and Hori, H. (2007). Duplicated Abd-B class genes in medaka hoxAa and hoxAb clusters exhibit differential expression patterns in pectoral fin buds. *Development genes and evolution* **217**, 263–73.

Taylor, J. S. and Raes, J. (2004). DUPLICATION AND DIVERGENCE: The Evolution of New Genes and Old Ideas. *Annu. Rev. Genet.* **38**,.

Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome research* **13**, 382–90.

Thomson, T. M., Lozano, J. J., Loukili, N., Carrió, R., Serras, F., Cormand, B., Valeri, M., Díaz, V. M., Abril, J., Burset, M., et al. (2000). Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome research* **10**, 1743–56.

Thorogood, P. (1991). The development of the teleost fin and implications for our understanding of tetrapod limb evolution. In *Developmental patterning of the vertebrate limb* (ed. J. Richard Hinchcliffe, Juan M. Hurle, D. S.), pp. 347–354.

Thorogood, P. and Wood, A. (1987). Analysis of in vivo cell movement using transparent tissue systems. *Journal of cell science. Supplement* **8**, 395–413.

Todt, W. L. and Fallon, J. F. (1987). Posterior apical ectodermal ridge removal in the chick wing bud triggers a series of events resulting in defective anterior pattern formation. *Development (Cambridge, England)* **101**, 501–15.

Van den Akker, E., Forlani, S., Chawengsaksophak, K., De Graaff, W., Beck, F., Meyer, B. I. and Deschamps, J. (2002). Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development (Cambridge, England)* **129**, 2181–93.

Van Eeden, F. J., Granato, M., Schach, U., Brand, M., Furutani-Seiki, M., Haffter, P., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A., et al. (1996). Genetic analysis of fin formation in the zebrafish, Danio rerio. *Development (Cambridge, England)* **123**, 255–62.

Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A. and Van de Peer, Y. (2004). Major events in the genome evolution of vertebrates: paranome age and size differ considerably

between ray-finned fishes and land vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1638–43.

Walker, M. B., Miller, C. T., Coffin Talbot, J., Stock, D. W. and Kimmel, C. B. (2006). Zebrafish furin mutants reveal intricacies in regulating Endothelin1 signaling in craniofacial patterning. *Developmental biology* **295**, 194–205.

Wall, D. P., Fraser, H. B. and Hirsh, A. E. (2003). Detecting putative orthologs. *Bioinformatics (Oxford, England)* **19**, 1710–1.

Wang, X., Grus, W. E. and Zhang, J. (2006). Gene losses during human origins. *PLoS biology* **4**, e52.

Westerfield, M. (2000). *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th ed. Eugene: University of Oregon Press.

Wilson, A. C., Carlson, S. S. and White, T. J. (1977). Biochemical evolution. *Annual review of biochemistry* **46**, 573–639.

Wittbrodt, J., Shima, A. and Schartl, M. (2002). Medaka--a model organism from the far East. *Nature reviews. Genetics* **3**, 53–64.

Wolfe, K. (2000). Robustness--it's not where you think it is. *Nature genetics* **25**, 3–4.

Woltering, J. M. and Duboule, D. (2010). The origin of digits: expression patterns versus regulatory mechanisms. *Developmental cell* **18**, 526–32.

Wood, A. (1982). Early pectoral fin development and morphogenesis of the apical ectodermal ridge in the killifish, Aphyosemion scheeli. *The Anatomical record* **204**, 349–56.

Wood, A. (1988). Contact guidance on microfabricated substrata: the response of teleost fin mesenchyme cells to repeating topographical patterns. *Journal of cell science* **90 (Pt 4)**, 667–81.

Wood, A. and Thorogood, P. (1984). An analysis of in vivo cell migration during teleost fin morphogenesis. *Journal of cell science* **66**, 205–22.

Woodhead-Galloway, J. and Knight, D. P. (1977). Some observations on the fine structure of elastoidin. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)* **195**, 355–64.

Yano, T. and Tamura, K. (2013). The making of differences between fins and limbs. *Journal of anatomy* **222**, 100–13.

Yano, T., Abe, G., Yokoyama, H., Kawakami, K. and Tamura, K. (2012). Mechanism of pectoral fin outgrowth in zebrafish development. *Development (Cambridge, England)* **139**, 2916–25.

Yokouchi, Y., Sasaki, H. and Kuroiwa, A. (1991). Homeobox gene expression correlated with the bifurcation process of limb cartilage development. *Nature* **353**, 443–5.

Yonei-Tamura, S., Abe, G., Tanaka, Y., Anno, H., Noro, M., Ide, H., Aono, H., Kuraishi, R., Osumi, N., Kuratani, S., et al. Competent stripes for diverse positions of limbs/fins in gnathostome embryos. *Evolution & development* **10**, 737–45.

YU Xiaobo, Zhu Min, Z. W. (2010). The Origin and Diversification of Osteichthyans and Sarcopterygians: Rare Chinese Fossil Findings Advance Research on Key Issues of Evolution. *Paleoichthyology* **24**,.

Zeller, R., López-Ríos, J. and Zuniga, A. (2009). Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nature reviews. Genetics* **10**, 845–58.

Zhang, X.-G. and Hou, X.-G. (2004). Evidence for a single median fin-fold and tail in the Lower Cambrian vertebrate, Haikouichthys ercaicunensis. *Journal of evolutionary biology* **17**, 1162–6.

Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J. and Gerstein, M. (2010b). Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome biology* **11**, R26.

Zhang, J., Wagh, P., Guay, D., Sanchez-Pulido, L., Padhi, B. K., Korzh, V., Andrade-Navarro, M. A. and Akimenko, M.-A. (2010). Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* **466**, 234–7.

Zhu, J., Sanborn, J. Z., Diekhans, M., Lowe, C. B., Pringle, T. H. and Haussler, D. (2007). Comparative genomics search for losses of long-established genes on the human lineage. *PLoS computational biology* **3**, e247.

Appendix 3A – Genomic sequence upstream of *and1* (zebrafish) containing the *1947and1* enhancer

GGTGAATTACAGCTTTAAGACACCTCTAGTTTAAATGAATTTAATTTGTTTAAAGACAGACCCACTATTAGTTCACT
ATATAAATATATACTGTATATTTAATGTTTAATTTTCCATGCTATTTTTCCATGTATTTTGTATTATTATATGCAAA
ACGCCTTTTTACCCCCCCAAAATTCTCAAACGTCCAAGTTGCTGTTTTTAGTTGTTTGCTTGGCAACAGAATATCGC
TATAAGTGCTCTAGCATGCACACGCTACACTTTTCTTAACGCTTTTCCGCCTCCTCTGTGTAGTGATAGTAAAGGTA
GTTTGAGCTTGGCACAGGATGGCCTAAAAACAGGCAGTTAAGAGTAAAGGAGGTATAAAGCTGAGCTCATGTAAAGC
GATTTTATGGGCCTTCATTGGCTCCAGCGCAGCAGGTCTTCTCTCTTTAGAGCCTTTATGCATCGGCAGCGCTTGAA
GTAGCTGAACACGTGTCACTCAGGGCTCTTTTATATCAGGGGGGGCCCTTCTGAAAGGTTTACGGAAAATAACCCGT
AGACCCCAATAAGAGCTCTTATTTTGATTTGACGGAGAACAGCACATCATAGATGAGCATTTATCTTTGCTGATGAC
TTTTACACTTTTTTTTTCCTCTATAGATTTCGAGGATTTATTCGATGACGATGATATTCAGTGACGAAGACGACAAA
CTGGATGAATTTCTCCTTTGTATAAAAATGTTCATTGTTGTAAATTTCAGTGATTGTCTTAACTGTCAAGTGTGTTA
ATAAACATTTGCACAAAACCCAAGTCTTCCTTTTTCCAGATGTTTCCACATTTTTTTTCGGAAACCCCAGACATTCC
TACTCAAAAGTTGGAATTATTATAAACGCTTTTTTCCCTAGGCTGAAATCTGTTTCGAGTGCTGATTGTAGGCTTTC
GCACATTCCTGCACACCTGTAAAGGGTGTATAATTACACATGCGATTGGCCGTTAACATAAAGCACAGATGTTTTAC
ACTCTGACATTGACATCAAAATATTTTTCCTTCCAGATATCGCTGAGCTGACATCATACTTCATTATTATTGAATTC
AAACGGCTTTTAATCTGTATCACTTCATCACAAACAGGGCGCTGATTGTGAATGATCATTTGTAGTGTTAGCTGTGA
GATGATGGATTGAGCTACTGCCCGGCAGATGTGCATGTGTTTAAACTCCCCATGAGGGATTTTACCCCCTTTTTCTT
GAGCACACAATGTTAGGTGTGTTAATTTTTGACACAATATTCAGTAAATGAAGATCTGTTTAATGCCTGTTATTATT
TTAATTGCAAGAAAAATAAATAAATAGATTAATAGATTAAATTTAGTTAAGATAAATAGAGATAGATGATGGACGGA
CATAAATAAACAGACTAGATAAAAAAAAAAATGTATTAATCAACAAATCAAAAATTAATAGTATAAAGTATATGTTTT
TACTTATTTAATTTTTATTCAATGTTTTTATCTTGTTAATGTTTGTAAAACTTTTCCAGTATTGGGTTGCAGCTGGA
AGGGCATCCGCTGCGTAAAACATATGCTGGATAAGTTGGTGGTTCATTCCGCTGTGGTGACCCCTGATTAATAAAGG
GACTAAGGCGAAAAGAAAATGAATGAATGAATGAATTTTTCTTAATTTTACATACGGAACATGGATTCTGCACCAAA
TATTTAAATCAATCCAGGTTATTATTAAATTTAATGTTAGTTCCTTACAATCATTGTTTGTTTTCGAACTAATCCTA
AGTATTCTGACCTGATTTTTCTCCTTCAGCATTTTCTTTAGCTTCTAAATGTAAAACAGTTGCTCTTTATGTTGTAC
GACAGCATCTGATTGGATGACTCCGGGCCCTATGGGATGCACCTTTTTTGTTTTTAATTATCAAGGTTGGGGTTTAA
ATTGTGGCCAGGTTCTCAACAGTGGTGCAGTCGGGTGTGTTGAGACTGTAGTGCACGAGGGGTTTGGCTCACCTTTG
TCTACCAGTCAGTCGCTCAGCTGTAAAGCTGAACTTCGAGGCCCAGTCGGACTGCCTCTCTTTAGAGGTTCACACCA
CCACATCCTAACACAGGTAAGATACACAATTGTATGTTCACTTATGTTTATGGTTAACCCAAAATGTAAAAATATAC
AGCAACAATTATTAATATGCTTGCAGTTCTTAAACACTGATTTGATCTGTTTGAAAAATAAAATAAAACCCTTGAAT
CTTTATTTAAGAAAGTATTTAATGTTGTTTCTATCTTTGAAAGCTGATAGTGACATTAACAGGAATTAAAACAATGT
AAAGAGCTTAAAAATGTACATTATTGTGATACATTTTAGTTCAATGTACATAATTTGTCTTTTGGGTGTTTTTTTTC
TTTATATAGAGACAATTTTGCTGGAAAATAGACAAAAAACAAAATCCTGTTGAATTGTCTTTTAATATTCGTATTTG
CAGTTTAATTTACTCCAACCTGATATAGTCACTCACTGAAATTTACATGATAAACTACCAATAAACAATAGGAAACT
TTACTCATTATAATCTTGATAGAAGTTTATCATGTCAAAGTACAGATATGAAAGAAAATAAACAAGCTGTAATTAAT
TTCATCCAAAGGAATCATGGCTCATTTGAGAGGATCTTCCATCTTCCATGTGTTACTGCTTGCAACATTAATATTAC
CAGGTGAGATTTTCAAAATATGACCTCAAGTGTTTATTTTTTCAGTAGTTCTATGCTAAAATGCCAGATCTGATTTA
GCATTTTTTGACTGAAATCCCTCTAAATATCCTACACCATATGAGCTGGTGTCAGGTTTCTGAGAACCAGCTCTTCG
TTTCTGCTCTTCAACTAGTAAAAATGATTCCTTGTTGTCAGCAAAACTCAAGTGCACAAAGTTTCCAGATGTGGCAG
AAGTTTACAAGCTCATAAAAAAAAAACAACAACTTGCTGATACCCTGAAAGAGCAACATTTCCTTTACCTCCCTCATT
TAAGACACAGAGATCTTCCGCTAAGTCAATATAACACAAACGTCCCCATTCCAGCCAAGATGAATGTCACTTTCAGA
GAGATTAAACGGTAAAGTGTCAGTTCATTCACTTTCCTGACGCTTGGATAATTAGGTTCACCTGTAGGTACAGAGGC
TCCGGTCCTTTGTAGATTTTATAGGTTTTGCTTATGGACAGGTTTGGTTATTGCAGCCCCCTGCTGCCAGGCATCAA
AGTGCTGATGTGTGCTGTTTGAAGATTAATTTCCTACAATACACTGTATATTACCACCTCCCCCCAAAAAAACTTGT
AAAACTGCATCCTGCTGCCATCAGCTTATTGGGTCTTTACCGTCTGTAAGCATTTTTCTACTTTGCTTTTGACGAAC
GCAACATCTGGTTGAGACCTGCACCCCCTAATTTTTGTTGTCTCTCTACTCTTCTTCCATAAAAGCATTTCTGCTGG
CTGGC

Fig 3A1. Genomic sequence upstream of *and1* (zebrafish) contains the *1947and1* enhancer and putative alternatively spliced exons. Various features of the *and1* upstream region include the start of the *1947and1* element (green), the end of the *1947and1* element (red), the *and1* Exon 1 as reported by Zhang et al. (2010) and Zfin.org (yellow), the *and1* Exon 1 as reported on Ensembl.org (light grey), the *and1* Exon 2 as reported by Zhang et al. (2010) and Zfin.org (light blue), the *and1* Exon 2 as reported on Ensemble.org or the start of Exon 3 as reported by Zhang

et al. (2010) and Zfin.org (dark blue), and the -*1064 to -1117* and -*971 to -1031* enhancer regions (dark grey).

Appendix 3B – Number of reporter gene-expressing cells after transient transgenesis

Table 1. Numbers of *eGFP*-expressing cells per microinjected embryo screened at 48hpf. Cell counting was not performed on all embryos with recorded *eGFP* expression.

Construct	Date of screening	Photo i.d. number	Total number of cells	Number of cells in fin fold	Percentage of cells in fold (%)	Notes
1947and1	13-Feb	1	32	24		Some GFP localizations are long streaks. One individual bright spot within the streak was counted as one cell. The maximum observed amount of bright spots in a streak was five.
		2	9	9		
		3	3	3		
		4	33	27		
		6	18	10		
		7	31	30		
		8	10	8		
		9	10	6		
		10	2	2		
		11	2	2		
	7-Jan	1	5	5		
		2	4	4		
		3	3	3		
		4	3	3		
	5-Oct	1	2	2		
		2	10	7		
		3	2	2		
		4	1	1		
		5	3	3		
		6	17	16		
	7-Oct	1	14	12		
		2	11	10		
		3	9	7		
		4	16	16		
	14-Oct	1	4	3		
		2	9	8		
		3	12	12		
		4	3	2		
		5	25	21		
		6	12	8		
		7	10	10		
		8	15	13		
		9	9	9		
		10	5	5		

Table 1 Continued

Construct	Date of screening	Photo i.d. number	Total number of cells	Number of cells in fin fold	Percentage of cells in fold (%)	Notes
	29-Oct	1	3	3		
		2	5	5		
		3	4	4		
		4	4	4		
		5	4	3		
		6	9	9		
		7	3	3		
		8	23	22		
		9	5	5		
		10	2	2		
		11	3	3		
totals (average)			419 (9.31)	366 (8.13)	87	
1527and1 (no data)						
1279and1	7-Oct	1	24	16		
		2	6	3		
		3	18	12		
		4	12	11		
		5	13	13		
		6	4	4		
	7-Jan	1	11	10		
		2	10	9		
		3	28	19		
		4	6	4		
		5	8	8		
		6	8	6		
		7	10	3		
		8	9	8		
		9	8	4		
		10	9	7		
		11	5	4		
		12	4	4		
		13	10	8		
		14	43	40		
		15	19	11		

Table 1 Continued

Construct	Date of screening	Photo i.d. number	Total number of cells	Number of cells in fin fold	Percentage of cells in fold (%)	Notes
		16	8	7		
		17	18	16		
		18	19	18		
		19	8	4		
		20	9	8		
		21	12	10		
		22	13	11		
totals (average)			352 (12.6)	278 (9.9)	79	
1199and1	21-Jun	1	4	4		
	22-Jun	1	14	12		
	26-May	1	10	10		
		2	5	5		
		3	9	9		
totals (average)			42 (8.4)	40 (8)	95	
1117and1	2-Jun	1	20	19		
	21-Jun	1	6	6		
		2	4	4		
		3	4	4		
		4	14	14		
		5	13	13		
		6	2	2		
		7	9	7		
		8	18	18		
		9	8	8		
totals (average)			98 (9.8)	95 (9.5)	97	
1063and1	2-Jun	1	19	19		
	21-Jun	1	6	6		
		2	4	3		
		3	3	3		
		4	7	7		
		5	13	13		

Table 1 Continued

Construct	Date of screening	Photo i.d. number	Total number of cells	Number of cells in fin fold	Percentage of cells in fold (%)	Notes
		6	2	2		
		7	7	6		
		8	12	12		
		9	7	6		
totals (average)			80 (8)	77 (7.7)	96	
1036and1 (no data)						
1297-1030Δ876and1	19-Feb	1	6	6		
totals					100	
1199-932Δ876and1	1-Feb	1	49	49		
		2	6	5		
		3	20	19		
		4	8	7		
	20-Feb	1	2	2		
		2	1	1		
		3	3	3		
		4	8	8		
		5	3	3		
		6	3	3		
		7	4	3		
	20-Oct	1	4	4		
		2	5	5		
		3	2	2		
		4	5	5		
		5	8	8		
		6	6	6		
		7	4	4		
		17	5	5		
totals (average)			146 (7.7)	142 (7.5)	97	

Table 1 Continued

Construct	Date of screening	Photo i.d. number	Total number of cells	Number of cells in fin fold	Percentage of cells in fold (%)	Notes
1117-971Δ876and1	3-Nov	1	3	3		
		2	24	23		
		3	15	11		
		4	3	3		
		5	1	1		
totals (average)			46 (9.2)	41 (8.2)	89	
1063-932Δ876and1	13-Feb	1	10	8		
		2	7	5		
		3	3	3		
		4	1	1		
		5	2	1		
	1-Nov	1	4	3		
		2	2	2		
		3	3	3		
totals (average)			32 (4)	26 (3.3)	81	
876and1 (no data)						

Appendix 3C – Results for Chi-squared test

Table 1. Results for Chi-squared test. DF=1, P≤0.05

Total number of scored embryos(n)	Observed embryos with fluorescence	X^2	P-value
101	39	25.4817	0
120	35	8.18435	0.004
109	26	1.71775	0.19
179	20	7.02543	0.008
177	4	32.0634	0
148	1	32.1462	0
137	19	2.29089	0.13
45	5	1.79453	0.18
95	8	6.84149	0.009

Appendix 3D - Predicted transcription factor binding sites of region -971 to -1947

Table 1. Predicted transcription factor binding sites of region -1063 to -1117 of any given core or matrix score.

Position (Strand)	Core Score	Matrix score	Sequence	Factor name
1125 (+)	1	0.969	ttttttCGGAAacc	Sp100
1118 (-)	1	0.987	GGAAAcccca	c-Rel
1108 (+)	1	0.937	gaCATTCctact	TEF
1106 (-)	1	1	cATTCC	TEF-1
1091 (-)	1	1	gttGGAAT	ZABC1
1089 (+)	1	0.9	ttgGAATTattataa	Tst-1
1088 (-)	1	0.865	ggaattATTATaaa	FOXJ2
1085 (-)	1	1	ATTAT	ZNF333
1084 (-)	1	1	tTATTA	PMX1
1082 (-)	1	1	ATTAT	ZNF333
1080 (-)	1	1	tATAAA	CDX-2
1069 (-)	1	0.988	tttTCCCTagg	Helios A
1069 (-)	1	1	tTTTCC	NFAT1
1069 (-)	1	1	tTTTCC	NF-AT4
1069 (-)	1	1	tTTTCC	NF-AT1
1046 (+)	1	0.938	tcgagTGCTGattgtaggc	c-Maf

Table 2. Predicted transcription factor binding sites of region -971 to -1031 of any given core or matrix score.

Position (Strand)	Core Score	Matrix score	Sequence	Factor name
1042 (+)	1	0.865	gtgctgATTGTaggctttcgca	Sox7
1027 (-)	0.8	0.79	tttcGCACAttcctgcacac	MTF1
1022 (+)	1	0.952	caCATTCctgca	TEF
1021 (-)	1	0.961	acATTCCtg	TEF-1
1020 (-)	1	1	cATTCC	TEF-1
999 (-)	1	0.955	ggtgtaTAATTacaca	dlx5
997 (+)	1	0.94	tgtatAATTAcacatg	dlx5
997 (+)	1	0.896	tGTATAattacaca	Zfp128
997 (-)	0.821	0.896	tgtataatTACACa	Zfp128
996 (-)	1	0.992	gtataATTACaca	41183
996 (+)	1	0.965	gtaTAATTac	IPF1
995 (-)	1	0.963	tataATTACacatgcg	Brn-2
994 (+)	1	1	ATAAT	ZNF333
994 (+)	1	1	aTAATTac	dlx-3
993 (+)	1	1	TAATTa	PMX1
993 (-)	1	1	tAATTA	PMX1
978 (-)	1	0.923	ttggccgTTAACataaa	Bbx
977 (+)	1	0.921	tggccgTTAACataaag	HNF-1alpha
977 (+)	1	0.961	tggccGTTAAcataaag	Bbx
976 (-)	1	0.946	ggCCGTTaac	v-Myb
975 (-)	1	0.99	gCCGTTaac	v-Myb
972 (-)	1	0.888	gttaaCATAA	POU3F2
963 (+)	1	0.957	aagcaCAGATgtttta	Tal-1beta:E47

Table 3. Predicted transcription factor binding sites of region -1117 to -1947 showing only the predictions with a core and matrix score of the highest possible value of 1.

Position (Strand)	Core Score	Matrix score	Sequence	Factor name
1908 (-)	1	1	tTTAAT	LHX3
1907 (-)	1	1	TTAATt	DRI1
1907 (-)	1	1	tTAATT	dlx5
1869 (-)	1	1	tATAAAt	CdxA
1869 (-)	1	1	tATAAA	CDX-2
1868 (-)	1	1	atAAATA	MEF-2C
1868 (-)	1	1	ATAAAt	SATB1
1850 (-)	1	1	tTTAAT	LHX3
1849 (-)	1	1	tTAATG	IPF1
1843 (-)	1	1	tTTAAT	LHX3
1842 (-)	1	1	TTAATt	DRI1
1842 (-)	1	1	tTAATT	dlx5
1838 (-)	1	1	tTTTCC	NFAT1
1838 (-)	1	1	tTTTCC	NF-AT4
1838 (+)	1	1	tTTTCCa	NFATC2
1838 (-)	1	1	tTTTCC	NF-AT1
1829 (-)	1	1	cTATTTt	MEF-2D
1828 (+)	1	1	TATTTtt	MEF-2C
1825 (-)	1	1	tTTTCC	NFAT1
1825 (-)	1	1	tTTTCC	NF-AT4
1825 (+)	1	1	tTTTCCa	NFATC2
1924 (-)	1	1	tTTTCC	NF-AT1
1808 (-)	1	1	ATTAT	ZNF333
1807 (-)	1	1	tTATTA	PMX1
1805 (-)	1	1	ATTAT	ZNF333
1742 (+)	1	1	tTGTTTgct	HNF-3beta
1735 (-)	1	1	cTTGGCa	NF-1A
1733 (-)	1	1	tgGCAACag	RFX
1730 (-)	1	1	CAACAgaa	BRCA1:USF2
1714 (+)	1	1	tAAGTG	Nkx3-2
1685 (+)	1	1	TTTCTt	PARP
1674 (-)	1	1	tTTTCC	NFAT1
1674 (-)	1	1	tTTTCC	NF-AT4
1674 (-)	1	1	tTTTCC	NF-AT1
1632 (-)	1	1	cTTGGCa	NF-1A
1593 (-)	1	1	aaGGAGG	GKLF
1587 (-)	1	1	gtATAAA	TBP
1586 (-)	1	1	tATAAA	CDX-2
1560 (-)	1	1	tTTTAT	HOXA13
1560 (+)	1	1	tTTTAT	SATB1

Table 3 continued

Position (Strand)	Core Score	Matrix score	Sequence	Factor name
1476 (-)	1	1	cACGTG	USF2
1457 (-)	1	1	tTTTAT	HOXA13
1457 (+)	1	1	tTTTAT	SATB1
1456 (+)	1	1	TTTATa	CDX-2
1455 (-)	1	1	GGGGGccct	LRF
1422 (+)	1	1	GGAAAa	NFAT1
1422 (+)	1	1	GGAAAa	NF-AT4
1422 (+)	1	1	GGAAAa	NF-AT1
1421 (+)	1	1	GAAAAta	IRF-4
1419 (-)	1	1	aAATAAcc	HMGIY
1350 (+)	1	1	aTTTAT	SATB1
1349 (-)	1	1	ttTATCT	GATA-4
1318 (-)	1	1	tTTTCC	NFAT1
1318 (-)	1	1	tTTTCC	NF-AT4
1319 (-)	1	1	tTTTCC	NF-AT1
1296 (+)	1	1	aTTTAT	SATB1
1295 (+)	1	1	TTTATt	Cdx-1
1287 (-)	1	1	aTGACG	CREB
1239 (-)	1	1	cTTTGTa	SOX10
1239 (+)	1	1	cTTTGT	SOX10
1235 (-)	1	1	gtATAAA	TBP
1234 (-)	1	1	tATAAA	CDX-2
1233 (+)	1	1	ATAAAa	HOXA13
1233 (-)	1	1	ATAAAa	SATB1
1222 (-)	1	1	cATTGTt	SOX10
1189 (+)	1	1	tcAAGTG	Nkx2-5
1189 (-)	1	1	tcAAGTG	NKX2B
1179 (+)	1	1	TAATAa	PMX1
1178 (-)	1	1	aATAAA	Cdx-1
1152 (-)	1	1	cTTCCT	ELF1
1152 (-)	1	1	cTTCCT	SPI1
1146 (-)	1	1	tTTTCC	NFAT1
1146 (-)	1	1	tTTTCC	NF-AT4
1146 (+)	1	1	tTTTCCa	NFATC2
1146 (-)	1	1	tTTTCC	NF-AT1
1143 (-)	1	1	tccAGATGtttc	RP58
1141 (-)	1	1	cAGATG	NMYC
1135 (+)	1	1	TTTCCac	NF-AT2

Appendix 4A - Multiple sequence alignment of Ecrg4 orthologs in human, mouse, rat, and zebrafish.

```
Hsapi        1 MAASP-ARPAVLALTGLALLLLLCWGPGGISGNKLKLMLQKREA-P--VPTKTKVAVDENKAKEFLGSLKRQ   68
Mmusc        1 MSTSS-ARPAVLALAGLALLLLLCLGPDGISGNKLKKMLQKREG-P--VPSKTNVAVAENTAKEFLGGLKRA   68
Rnorv        1 MGTSS-ARPAVLALAGLALLLLLCLGPGDVSGNKLKKMLQKREG-P--VPSKTNVAVSEHTAKEFLGGLKRA   68
Dreria       1 MLSEKF-HLRLLTL--LT-LLTALSLTDVASESKLEKLLMKRVDRD--VKPAAAVAVSPSKAKEFLTSLKRP   66
Drerib       1 MSLHSLCVPAILLI---S-VLSICLPSGGSSDSKLHRILIKRDAKEIDSRPKAYISVQQSKAKEFLSGLHRT   68
cons         1 *         :* :   : :*     ..  * .**. :* **        . : ::*   .***** .*:*    72

                 ↓Arg-70                                                          ↓Arg-132
Hsapi       69 KRQLWDRTRPEVQQWYQQFLYMGFDEAKFEDDITYWLNRDRNGHEYYGDYYQRHYDEDSAIGPRSPYGFRHG  140
Mmusc       69 KRQLWDRTRPEVQQWYQQFLYMGFDEAKFEDDVNYWLNRNRNGHDYYGDYYQRHYDEDAAIGPHSRESFRHG  140
Rnorv       69 KRQLWDRTRPEVQQWYQQFLYMGFDEAKFEDDVNYWLNRNQNGHDYYGDYYQRHYDEDAAIGPRSREGFRHG  140
Dreria      67 KRNLWDRSRPDVQQWIQQFMYMGFDEARLETDLAYWMDQHRSSDQ----GRQHHYDENAALGPRGAASYRHG  134
Drerib      69 KRNVWDRSRPDVQQWIQQFMYMGYDEARLETDLSYWMDQARSNDQ----GRQHHHDENAPMSQQDPRYNRHG  136
cons        73 **::***:**:**** ***:***:***::* *: **::: :...:      *:*:**::.:. :.    ***  144

Hsapi      141 ASVNYDDY  148
Mmusc      141 ASVNYNDY  148
Rnorv      141 ASVNYDDY  148
Dreria     135 ANVNYDYY  142
Drerib     137 ANVNYDYY  144
cons       145 *.***: *  152
```

Fig 4A1. Well-documented cleavage sites and their conserved amino acids are highlighted in green. Sequence conservation is shared at the predicted Furin cleavage site located at the human Arginine 70. Numerous studies (Mirabeau 2007, Baird 2012, Farrah 2011) have recognized cleavage at Arg-132 of human Ecrg4, despite the absence of a canonical cleavage motif. Alignments were produced using TCoffee, and cleavage predictions were verified using Neuropred. Zebrafish Ecrg4a and Ecrg4b sequences indicated as *Dreria* and *Drerib*, respectively.

Appendix 4B - Pairwise sequence alignments of zebrafish And1(44-103) with zebrafish Ecrg4a(55-110) and Ecrg4b(57-112).

```
D.rerAnd1          44 GLEEAPKKLIRNRRNISWYKQHSDFWNWYKYFTDNENKDGVAELDRVYLS     93
                      ..:|....|.|.:||: |.:...|...|.:.|......:...|.|   |:
D.rerEcrg4a        55 KAKEFLTSLKRPKRNL-WDRSRPDVQQWIQQFMYMGFDEARLETD---LA    100

D.rerAnd1          94 YLQNKNRAES    103
                      |..:::|:..
D.rerEcrg4a       101 YWMDQHRSSD    110
```

```
D.rerAnd1          44 GLEEAPKKLIRNRRNISWYKQHSDFWNWYKYFTDNENKDGVAELDRVYLS     93
                      ..:|....|.|.:||: |.:...|...|.:.|......:...|.|   ||
D.rerEcrg4b        57 KAKEFLSGLHRTKRNV-WDRSRPDVQQWIQQFMYMGYDEARLETD---LS    102

D.rerAnd1          94 YLQNKNRAES    103
                      |..::.|:..
D.rerEcrg4b       103 YWMDQARSND    112
```

Fig 4B1. Sequences presented correspond to fragments of the full-length proteins which are presented in Objective 2, Fig 4.1. Predicted convertase cleavage sites are highlighted in green. And1 and Ecrg4a sequence identity (top) was calculated as 14/60 (23.3%) and similarity calculated as 25/60 (41.7%). And1 and Ecrg4b sequence identity was calculated as 15/60 (25.0%) and sequence similarity was calculated as 24/60 (40.0%). Pairwise sequence alignments produced with Emboss.

Appendix 4C - Multiple sequence alignment of And1 orthologs in various fish species.

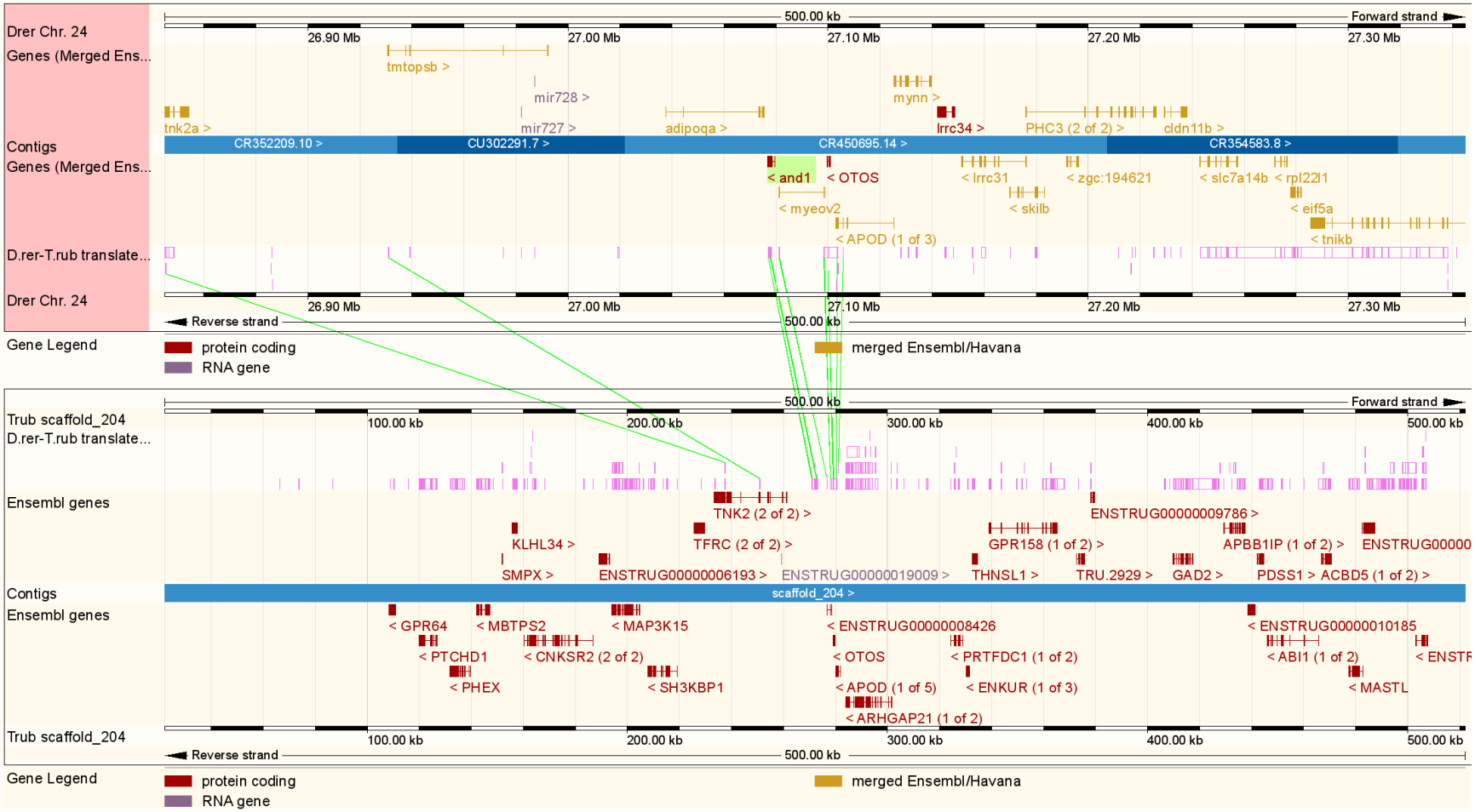
```
                                                                                                  ↓Arg-57
Drerio          1 MAHLRGS---SIFHVLLLATLILPAFLLAGPVTQRLKGDDGSD--DKTGLEEAPKKLIRNRRNIS-WYK   63
Xmaculatus      1 MAERRSC--YAVWITSLLIVMLLPAFLSAGPILQKSKGDVGESIQEKAEAAASHRRLVRNRRNIS-WYK   66
Oniloticus      1 MAERRRTSFSGVLITVVLAVMLLPAFLSAGPIVQKPRGDAGESIQEKASAAAAHRRLIRNRRNIS-WYK   68
Gaculeatus      1 ------------------------------------------------------------------------    0
Olatipes        1 ------------------------------------------------------------------------    0
Gmorhua         1 ------------------------------------------------------------------------    0
Lchalumnae      1 ----------------------------------------------------------VRQKRNPRMGYP   12
cons            1                                                                          69

                                                  ↓Arg-105
Drerio         64 QHSDFWNWYKYFTDNENKDGVAELDRVYLSYLQNKNRAESRRSYKMYLQHLGDIYKS-CAESDDPNCVS  131
Xmaculatus     67 QHSDFWAWYKYFTDNGNQDAVQEMDKIYLAYLQNKNRAEARRSYKAYLRHLGEIYKS-CANSDDPNCVA  134
Oniloticus     69 QHSDFWAWYKYFTDNGNQEAIQEMDRIFLAYLQNKNRAEGRRSYKAYLSHLGDVYKS-CSDSEDPNCVS  136
Gaculeatus      1 ------------------------MDRIYLAYLQNKNRAEGRRSYKAYLRHLGDVYKS-CADSDDPNCVS   45
Olatipes        1 ------------------------MDRIYLAYLQNKNRAEARRSYKEYLRHLGEIYKS-CADSADPNCVA   45
Gmorhua         1 ------------------------LDRIYLAYLQNKNRAEGRRSYKAYLSHLGDIYKS-CADSEDPKCVA   45
Lchalumnae     13 DNRNFWDWYQYFTDTGNAEKVHEMDKVYLEHLQKKNREEAKQSYVAYLQHLKELYGAQCDPLKDTHCAH   81
cons           70                        :*:::* :**:*** *.::**  ** ** ::* : *    *.:*.  138

Drerio        132 SYTNR-----------------PKPKAEAPVPAPV------KSCDPYRDPYCLYSK---GYSYY-PYLA  173
Xmaculatus    135 SYTGR-----------------PKSKPEPPKPAPI------KTCDPTKDTHCLYMALVQGKSPYLPLVV  180
Oniloticus    137 SYTSR-----------------ASPKAEPPKPAPV------KTCDPTKDPYCMYMALVQGKSRYKPLVL  182
Gaculeatus     46 SYTSR-----------------PK--PEAPKPAPV------KTCDPTQDAYCLYAALVQGKSPYQPLAL   89
Olatipes       46 SYTTR-----------------PK--PEPAKPAPL------KTCDPTKDPHCLYMALVQGKSPYLPLVV   89
Gmorhua        46 THTTR-----------------PT--PEPPKPAPV------KTCDPTKDAHCLYAALMQGKSPYLPLVL   89
Lchalumnae     82 YLTTMQLNKSPEPPLPGSPAPPPP--PTPPPPAPYRSYLRSKPCDPRKDLYCRSQPLRQP---------  139
cons          139  *                   .   . .. ***       *.*** :* :*                 207

Drerio        174 PAPVKSPAPAPAPV--KAPAYLHT-PVVKDPRSGQYYYSPLVQPFLSAEQRAELLRICDAEDVECLQYH  239
Xmaculatus    181 ------PVAPAAAVKAPAPLYVRAAAQSKEP--AHYHFAPSAVSFLTKEQKAELLRICSSDDVECLQYH  241
Oniloticus    183 ------PAA------TPAPQYIHSAAPAKDPQSGYYYYSPSTMPFLTMEQKAELLRICPSDDVECLQYH  239
Gaculeatus     90 ------PAAAPAPVKAPAPQYARSAAPAKDPQSGYYYYSPSAASFLSKEQKAELLRICASDDVECLQYH  152
Olatipes       90 ------PVAPAAPVKAPAPLYVRS-APAKAPQSGYYYYSPSAVSFLTKEQKAELLRICSADDVECLQYH  151
Gmorhua        90 ------PATK-AAVAAPAPVYARGPAPAKNPRSGYYYYSPSVQPFLTKEQKAELLRICSPEDVECLQYH  151
Lchalumnae    140 ----------------------------QQPPVAYYFSYASPRPYLTPKQQSELADLCDSSDTACIQYY  180
cons          208                            : *  . *.  .   .:*: :*::**  :* ..*. *:**:  276

Drerio        240 LRAAYGYRSAAALAPSYAHLGCDPKTDPQCVPHLVQKAPSGLYLRYPNCDPRVDPYCAYAAALASSQNP  308
Xmaculatus    242 LRAAFGYRASSGPLPSYSHLGCDPKKDPACKAKLAQKAPSGLYHQNPNCDPLRDPYCSYAASLVAPRAP  310
Oniloticus    240 LKAAYGYS--GGSRPSYAHLGCDPKKDPTCK--MAQKSPAGLYMN------------AYSASSMAPQA-  291
Gaculeatus    153 LRAAYGYGPSAGPMPSYSHLGCDPSKDPNCKPKLVQKAPSGLYLQYPNCDP-RDPRCAHAASLVAPRAP  220
Olatipes      152 LRAA------------------------------------------------------------------  155
Gmorhua       152 LKAAFGYKPAAGPLPSYAHLGCDPKKDPGCQPQLVQKAPSGVYLKYPNCDPLRDPRCAYEAALSA----  216
Lchalumnae    181 IKT-YVPRPAVVPAPIYNYQTCDPQKDPSCAQTIA---------NYPHCDPETDPRCRFLMLSYRSTPP  239
cons          277 :::                                                                  345

Drerio        309 E------SPPCNPLYEDNCNPLTATRIGSLAHENLKDKLNSEAGAMR-AP-QSPQHDPYAMFRDASAGA  369
Xmaculatus    311 NPSAHTGPGSCNPLYEDNCNPLTATRFSSPP-QEYRGDERDEAASIRAAPPPAEQNDPYAMFRDPYANA  378
Oniloticus    292 ------APGSCNPLFEDGCNPLTASRFSTPP-ESYKSNERDQAAAIRAAP-PAEQYDPYSMYRDSYSNV  352
Gaculeatus    221 NPPAAAGPGSCNPLYEEGCNPLTATRFATPP-EAYQSEEEDEAAAIRAAP-PAVQNDPYATFRDAYANA  287
Olatipes      156 ------------------------------------------------------------------------  155
Gmorhua       217 ------------------------------------------------AAPTQAPANDPYAMYRQASAA-  237
Lchalumnae    240 V------QTRCNPYYDSGCIPEPAPE----------------------PPCDTTYDPYCQVSSPLAS-  278
cons          346                                                                        414
```

```
Drerio          370 D------------------------------LRRHM----------------------------------  375
Xmaculatus      379 NRVTDPYAMYRQYP----TPAAPQSTDPYNILRQFMSQAQDNDPYAPRM-------SAPESNPNDPFSA  436
Oniloticus      353 NRVTDPYAMHRQAS----APASPPATDLFAILRQYMAQAHANDPYASRMG------GASESNSNDPYSA  411
Gaculeatus      288 NRVNDPYAMYRQASAAASAPATPPANDPYAMLRRYMAQAQGNNPYATHRGGGPHAEAAPEANPNDPFSA  356
Olatipes        156 ------------------------------------------------------------------  155
Gmorhua         238 -SRNDPYAMYRQASAA-AAPTEPPANDPYAMLRKLMSRAQSNDPHAPR----------------------  283
Lchalumnae      279 ----------RQKP----RPQCDPRYHPYCQV---------PS--------------------------  298
cons            415                                                                     483

Drerio          376 -------------PAQ-----LQPPRYH-QPE-EPEHPLGPRGKTKEGYECFVGYDKECIPLSSQNERR  424
Xmaculatus      437 VRE---AMNRQSPPSPRQQYPFSNPSYE-EPPQEERHSLGPPGKTKEGYDCYIGYDRECFPVRPSQP-H  500
Oniloticus      412 MRDAAGAMHRHTPPSPRQQHPYSYPSYE-EPAQEERHPLGPPGKTKEGYDCFIGYDRECYPAKPSAP-R  478
Gaculeatus      357 IREAAAAMHRRGQPAARQQFPFANPSYE-EPAHDERHPLGPPGKTKEGHDCFIGYDRECFPATPAEP-R  423
Olatipes        156 ------------------------------------------------------------------  155
Gmorhua         284 ------------------------------------------------------------------  283
Lchalumnae      299 -------------PVTSEQRP--EPQLECEDYYDLR--CRMMGKTKEGHDCYIYYDKDCHLVRPPNE-E  349
cons            484                                                                     552

                                                                          Arg-470↓
Drerio          425 PAVHRQPSYPA-------------------EAYEPHLNADGSRTGVIEPD-PDCDPEYDRNCRLRRYEP  473
Xmaculatus      501 SGAPRRTPYPV-------------------DPYQPHQNTDGDRRGVLEPDNPHCDPEYDRDCRLRRYEP  550
Oniloticus      479 SGVQRRIPYPA-------------------EAYEPHLNSDGTRNGVLEPENPHCDPEYDRDCRLRRYET  528
Gaculeatus      424 SGIHRRVPYPA-------------------EAYEPHLNADGSRSGVVEPSDPHCDP-------------  460
Olatipes        156 ------------------------------------------------------------------  155
Gmorhua         284 ------------------------------------------------------------------  283
Lchalumnae      350 K-EPERIPEPVRAPCNPRDPNCRAMQPPPKKTYPPN---YGYRKGVIEPD-PDCDPEYDPNCRLRRVDP  413
cons            553                                                                     621

Drerio          474 EPAEPVSISPQEDVAHPAQEPSHHEEIPQHREESYPETPGYDQQQYDAYMSGQEDPYAGYGPQEPGAAS  542
Xmaculatus      551 AQPQ----------------PELHTQEDRSQRTEESEQHGQDQYEVEPYQSGQEEPYMSYPPQSQGTPS  603
Oniloticus      529 EQPQS--------------QPERHAAADHSQTAAEREQVNQDQYEAEPYQSGQEESHTPYQPLSQGMPS  583
Gaculeatus      461 ------------------------------------------------------------------  460
Olatipes        156 ------------------------------------------------------------------  155
Gmorhua         284 ------------------------------------------------------------------  283
Lchalumnae      414 NSSTS--------------GNSQAPAAP----------------QKGSTGSGSEDSQSSPHPNEAQQPG  452
cons            622                                                                     690

Drerio          543 FQDVLRGYGGRYPGQDDHLAYNGDYRKK  570
Xmaculatus      604 LQDILRSYGDRFPEQDDHRAYADDYRKK  631
Oniloticus      584 LQDILRRYGDQYTRQDDHRAYGDNYHKK  611
Gaculeatus      461 ----------------------------E  461
Olatipes        156 ----------------------------Y  156
Gmorhua         284 ----------------------------T  284
Lchalumnae      453 AE--VKG------------------YNL  460
cons            691                              718
```

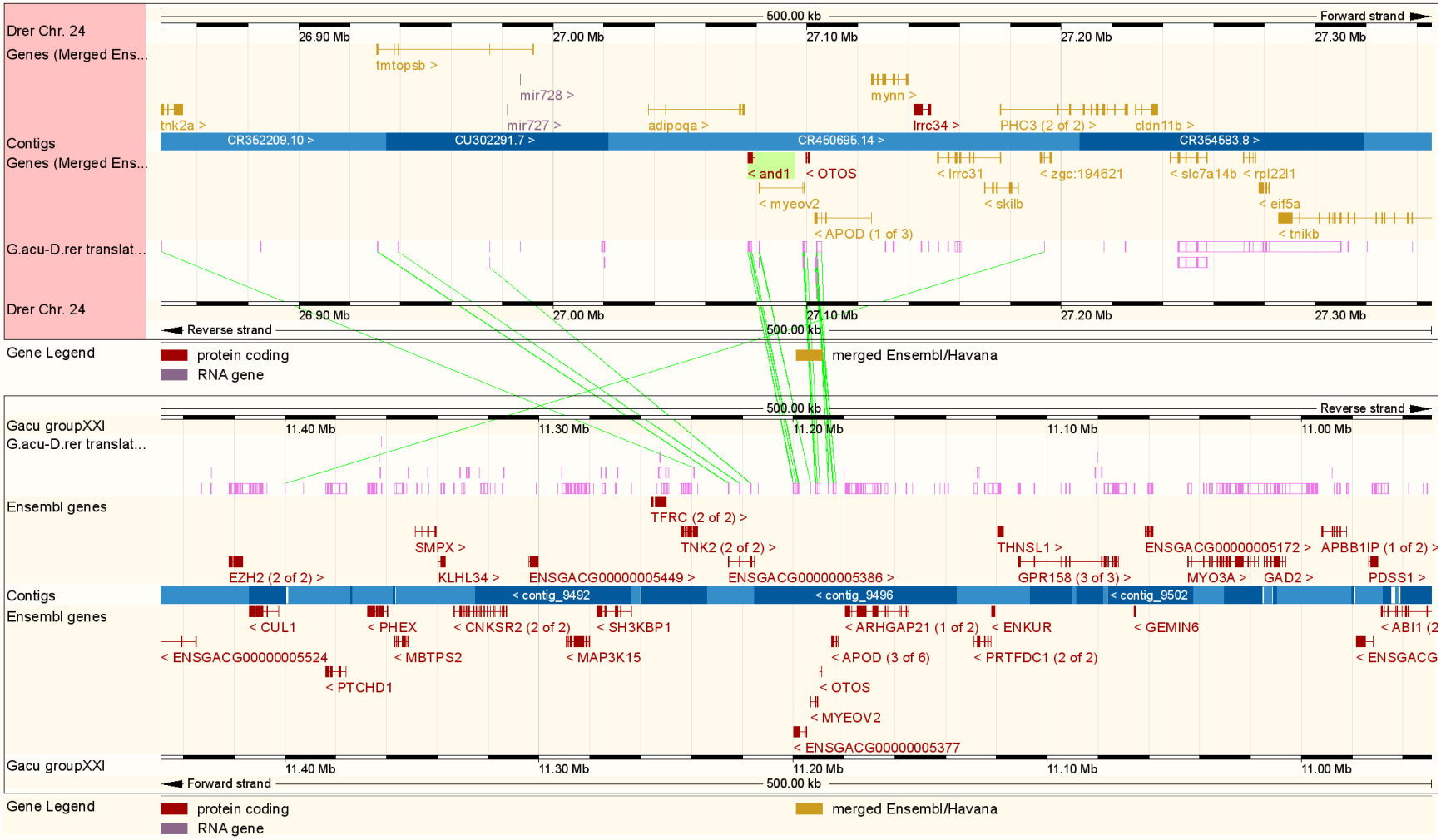
Fig 4C1. Predicted cleavage sites are highlighted in green. Sequence conservation at the location of And1 Arginine 105 is shared between all actinopterygian fish species. Coelacanth, a sarcopterygian, does not possess a dibasic arginine motif in this region. Predicted RXRR cleavage sites are conserved and indicated by their amino acid location in zebrafish at arginines 57 and 470. Sequences were obtained with a tblastn function in the Ensembl database. The identity of And1 for each species was verified based on chromosome synteny. Shortened sequences are due to incomplete genome sequences or annotation. Alignments were produced using TCoffee, and cleavage predictions were verified using Neuropred.

Appendix 5A. Synteny of *and1* loci in zebrafish (*Danio rerio*) compared to: fugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*), tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*), and coelacanth (*Latimeria chalumnae*). Synteny graphics produced with Ensembl.org "region comparison" tool.

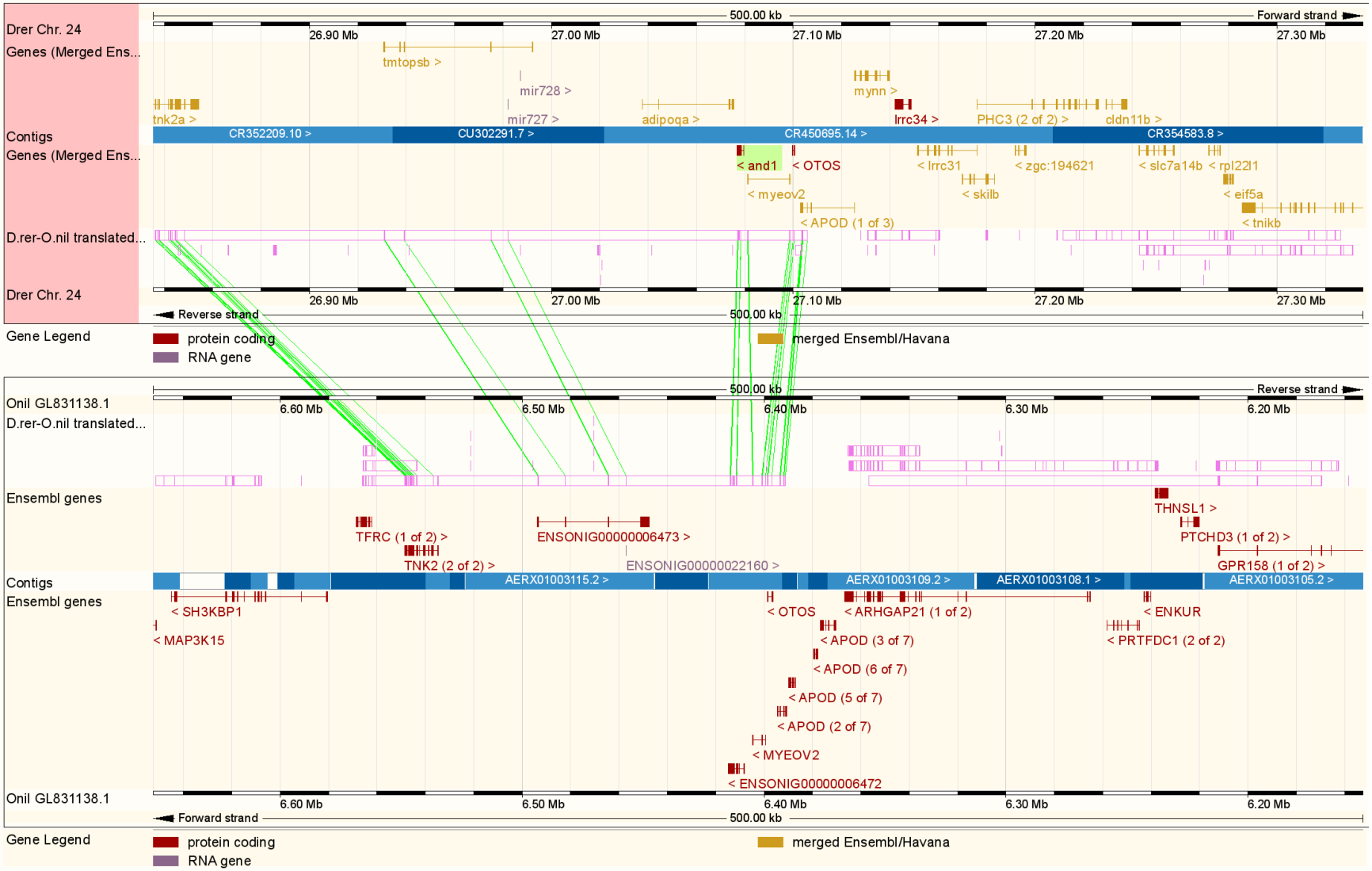
Zebrafish and Fugu



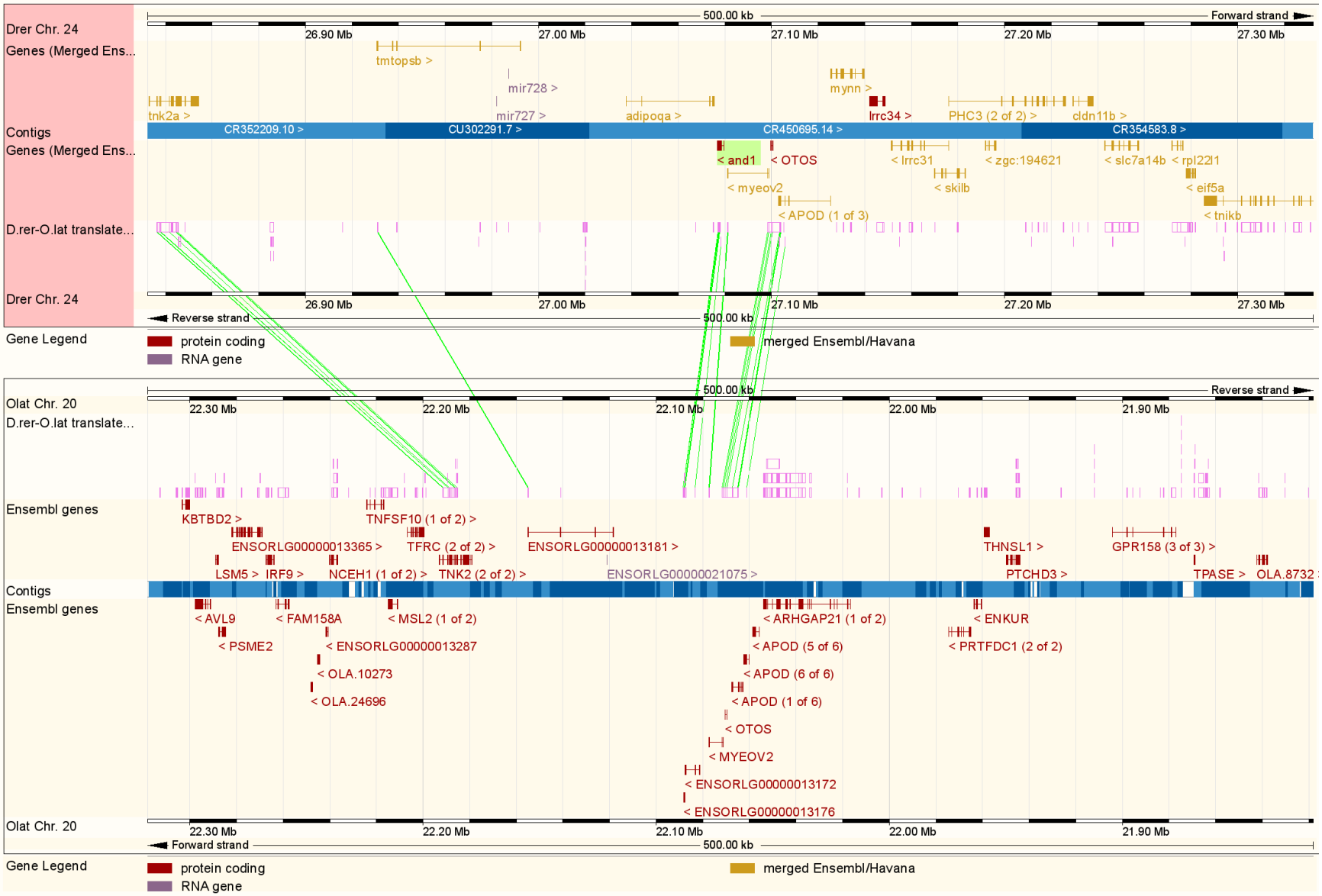
Zebrafish and Stickleback



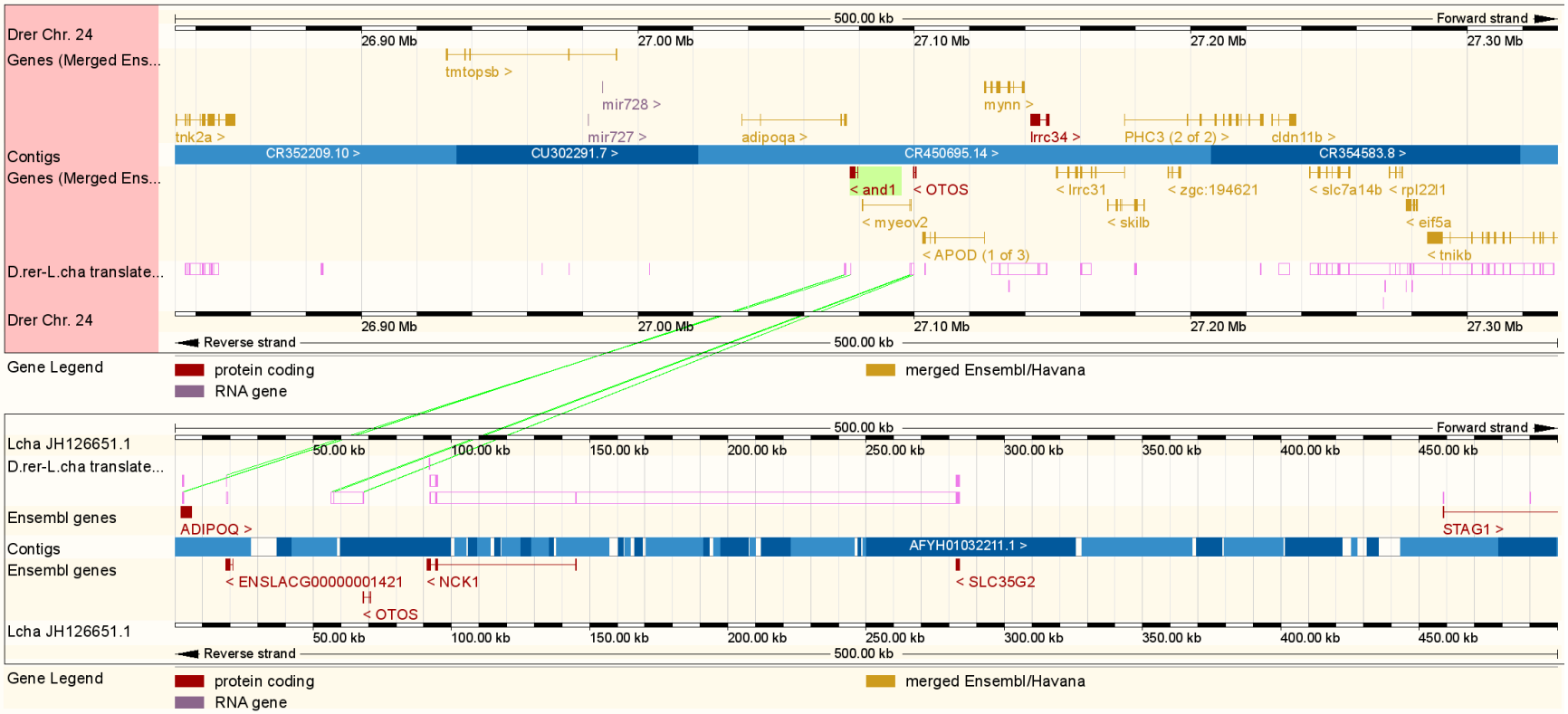
Zebrafish and Tilpaia



Zebrafish and Medaka

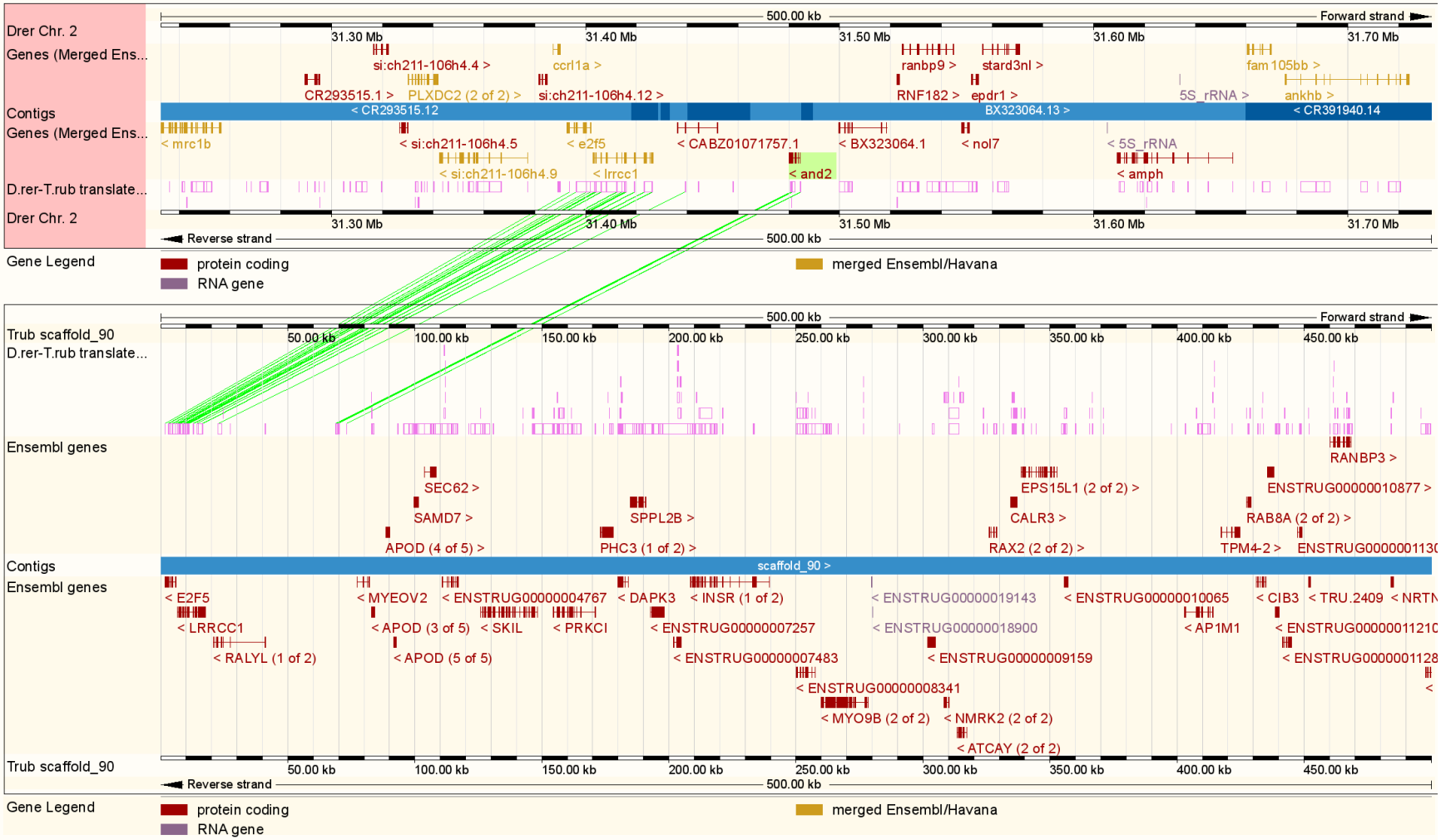


Zebrafish and Coelacanth

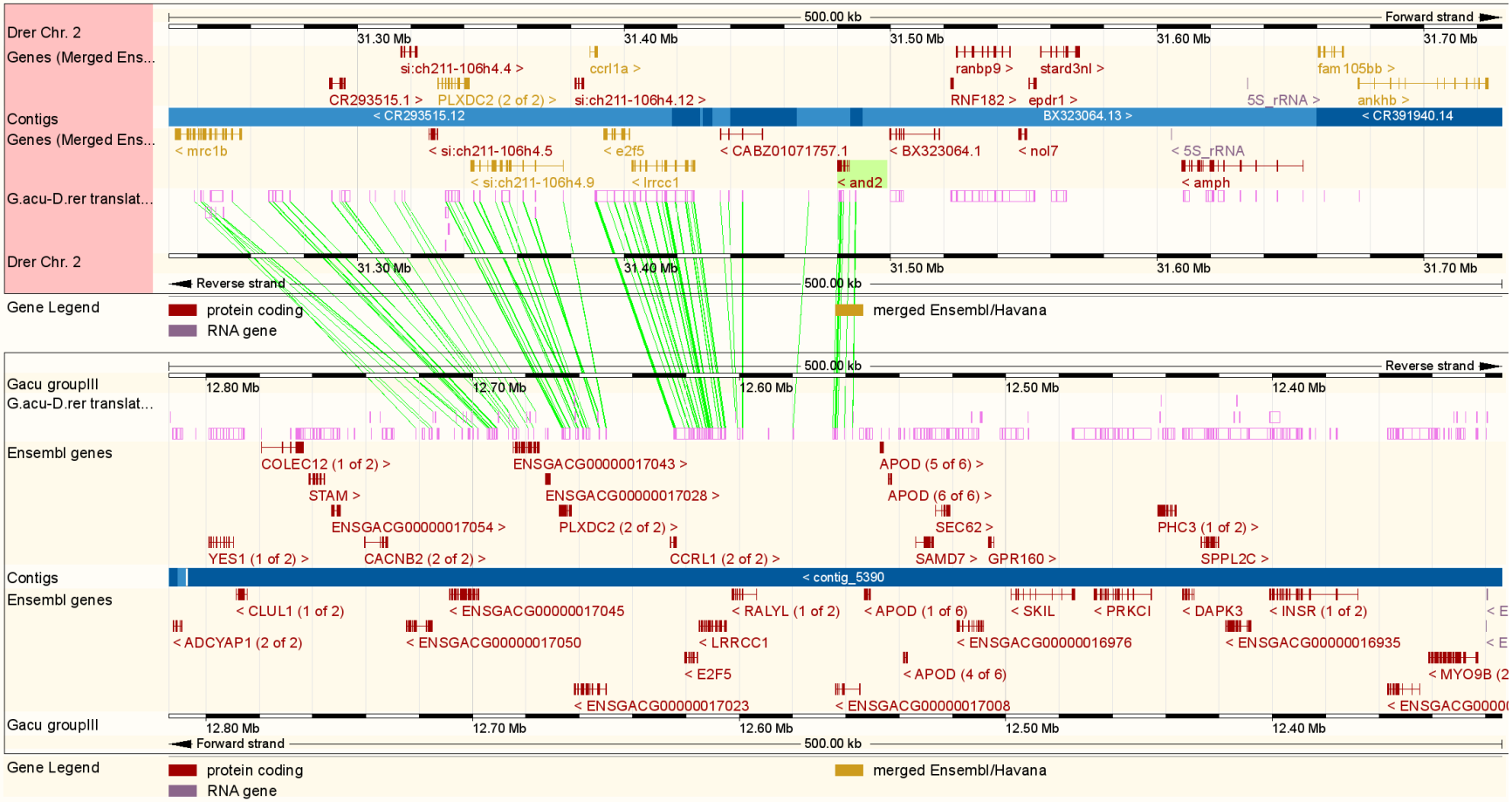


Appendix 5B. Synteny of *and2* loci in zebrafish (*Danio rerio*) compared to: fugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*), tilapia (*Oreochromis niloticus*), and medaka (*Oryzias latipes*). Synteny to coelacanth (*Latimeria chalumnae*) does not display same pattern as above species. Synteny graphics produced with Ensembl.org "region comparison" tool.

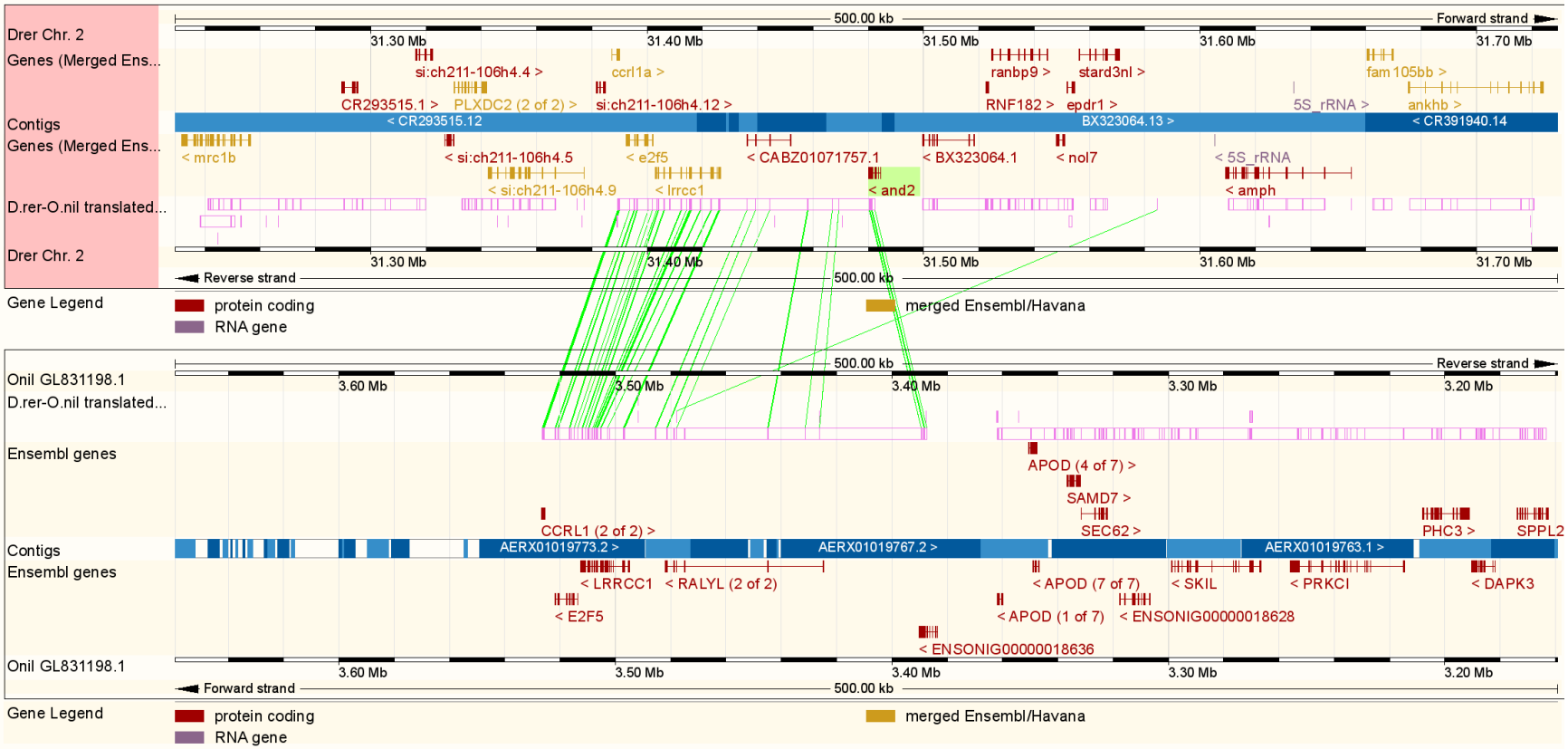
Zebrafish and Fugu



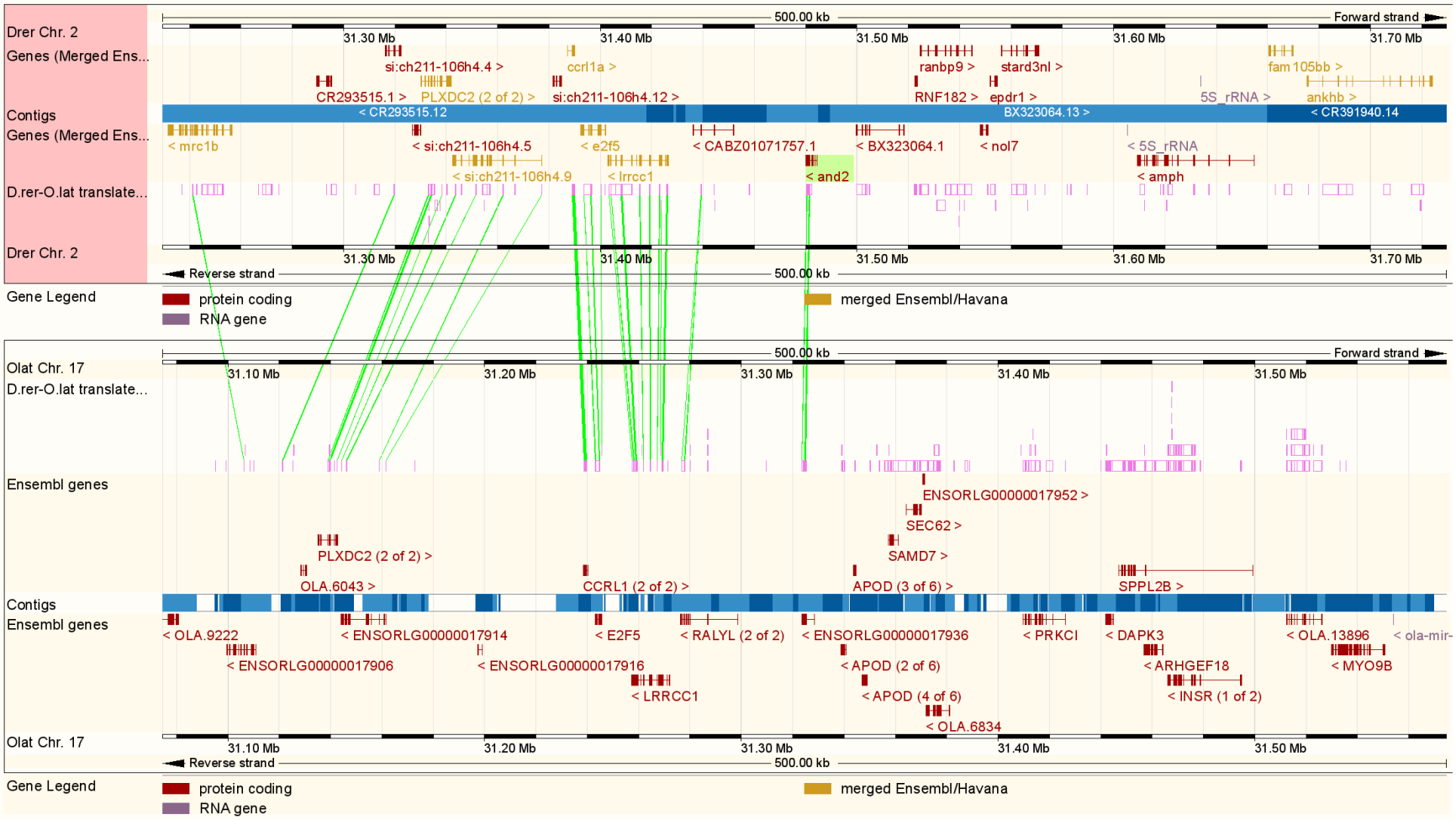
Zebrafish and Stickleback



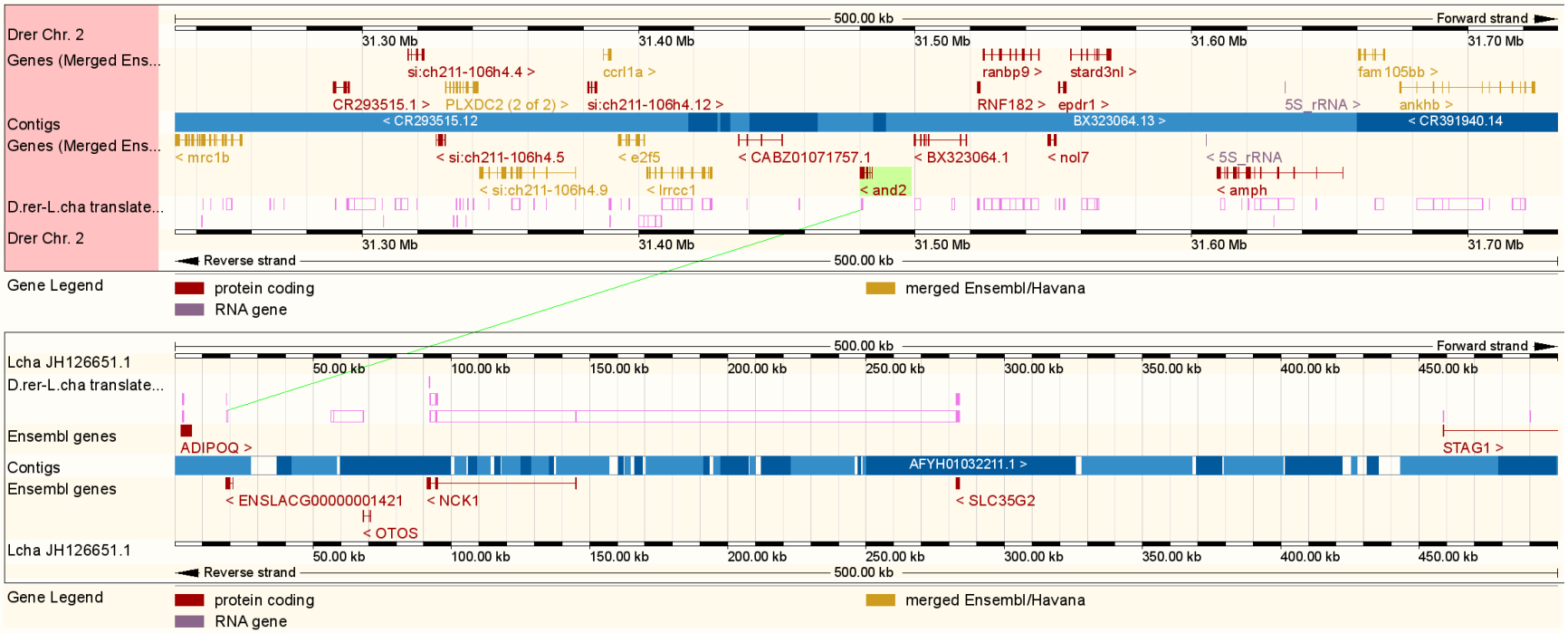
Zebrafish and Tilapia



Zebrafish and Medaka

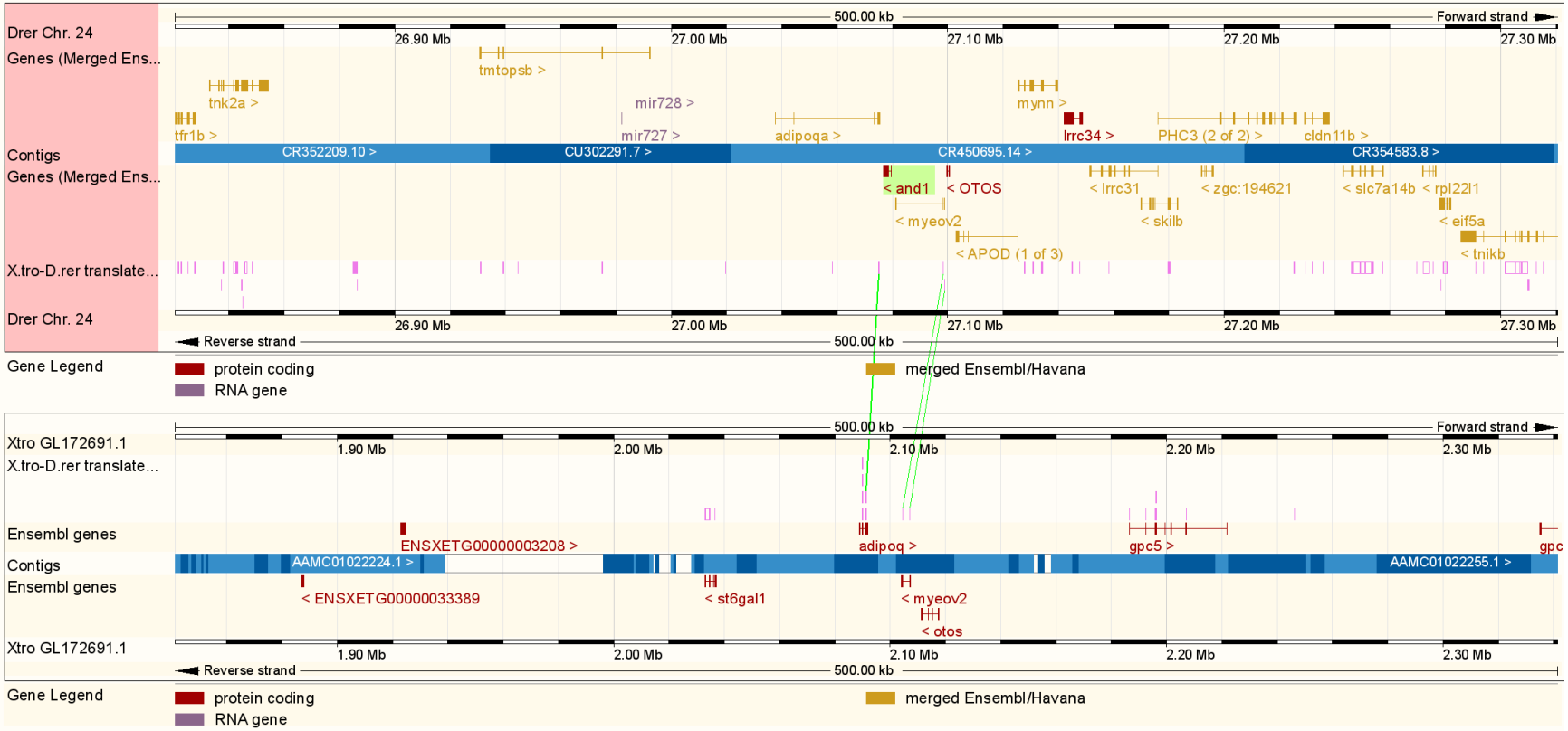


Zebrafish and Coelacanth

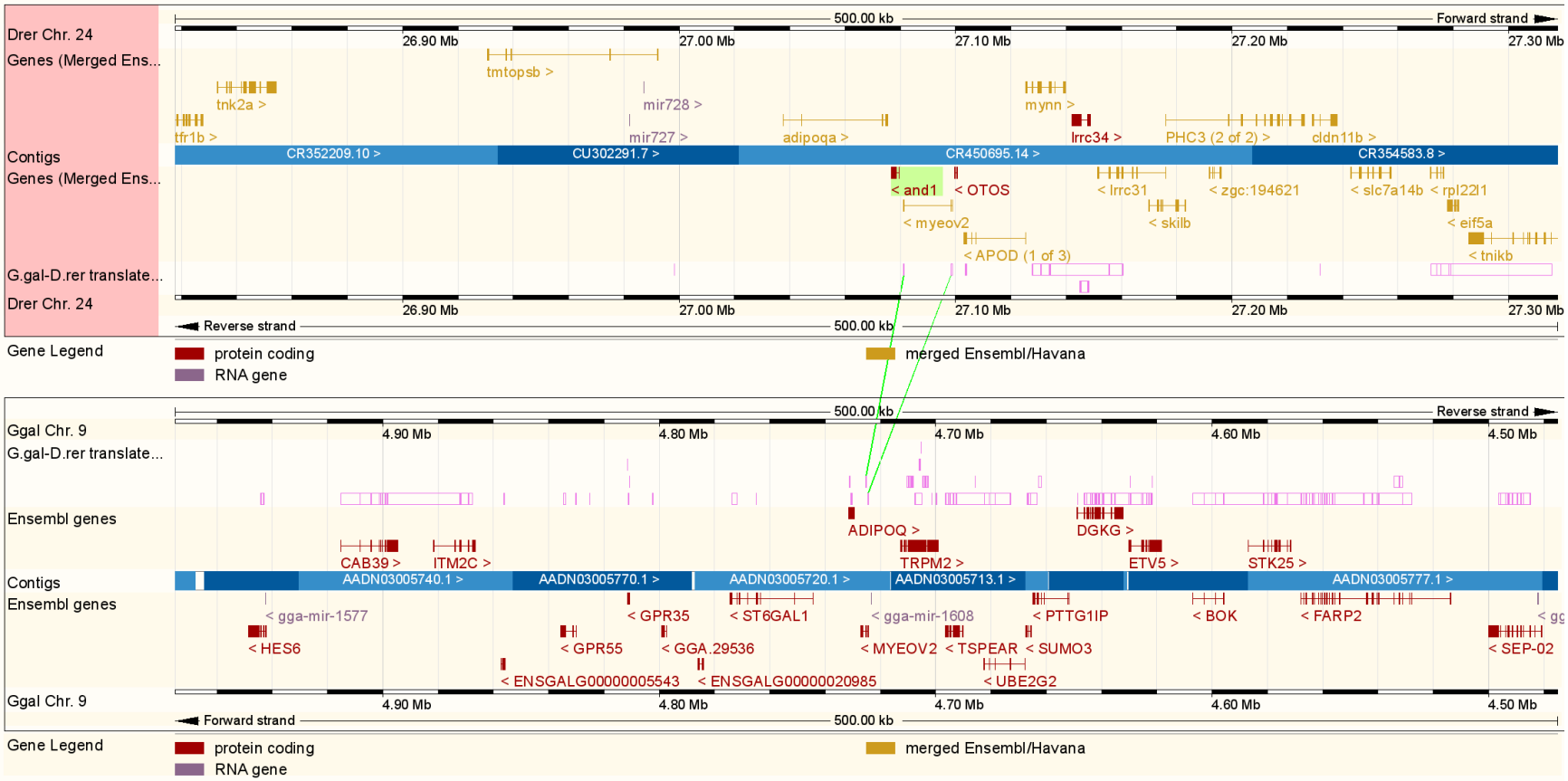


Appendix 5C. Synteny of *and1* loci in zebrafish (*Danio rerio*) compared to: *Xenopus* (*Xenopus tropicalis*), chicken (*Gallus gallus*), Rat (*Rattus norvegicus*) and human (*Homo sapiens*). Synteny graphics produced with Ensembl.org "region comparison" tool.

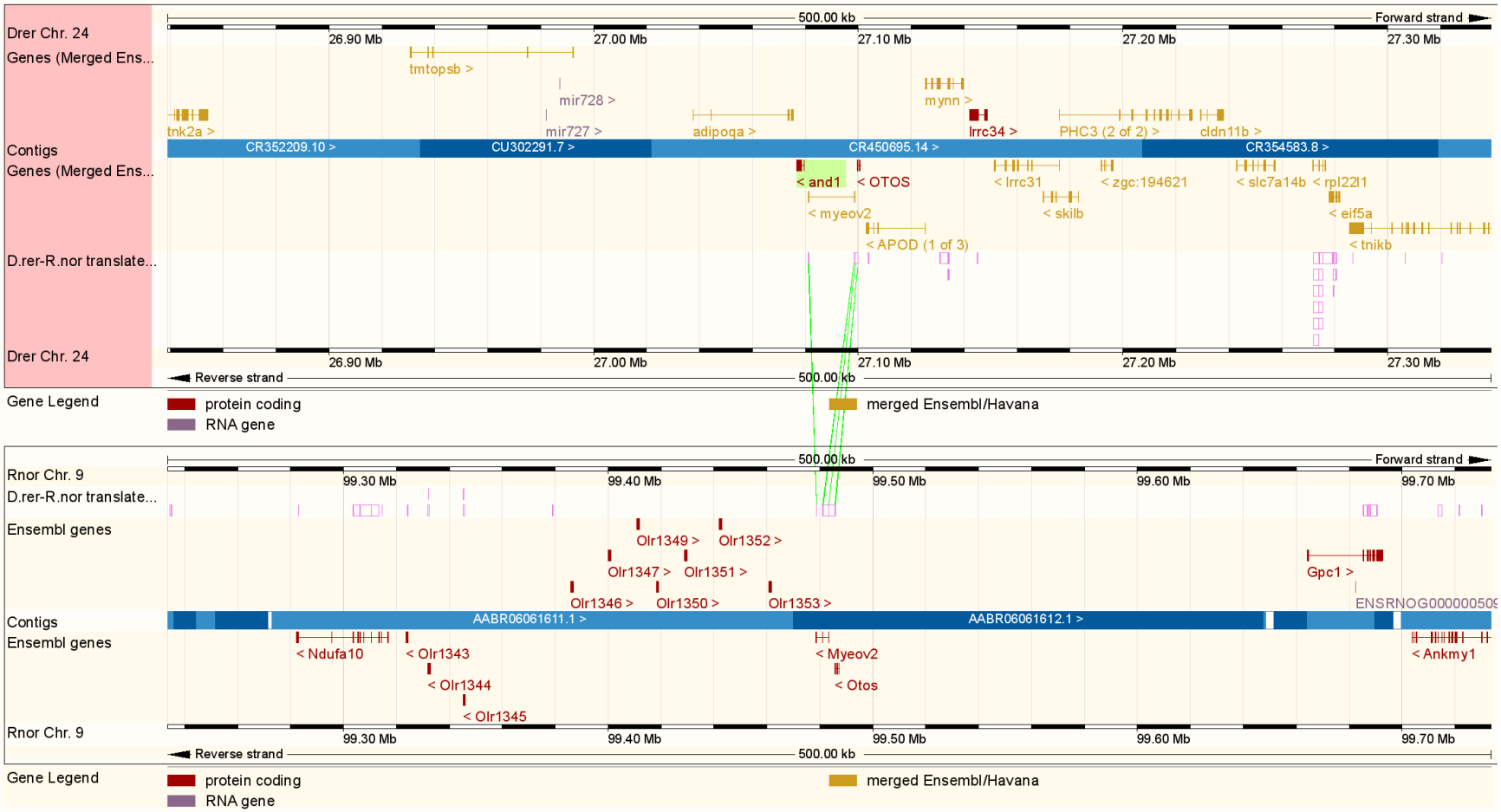
Zebrafish and *Xenopus*



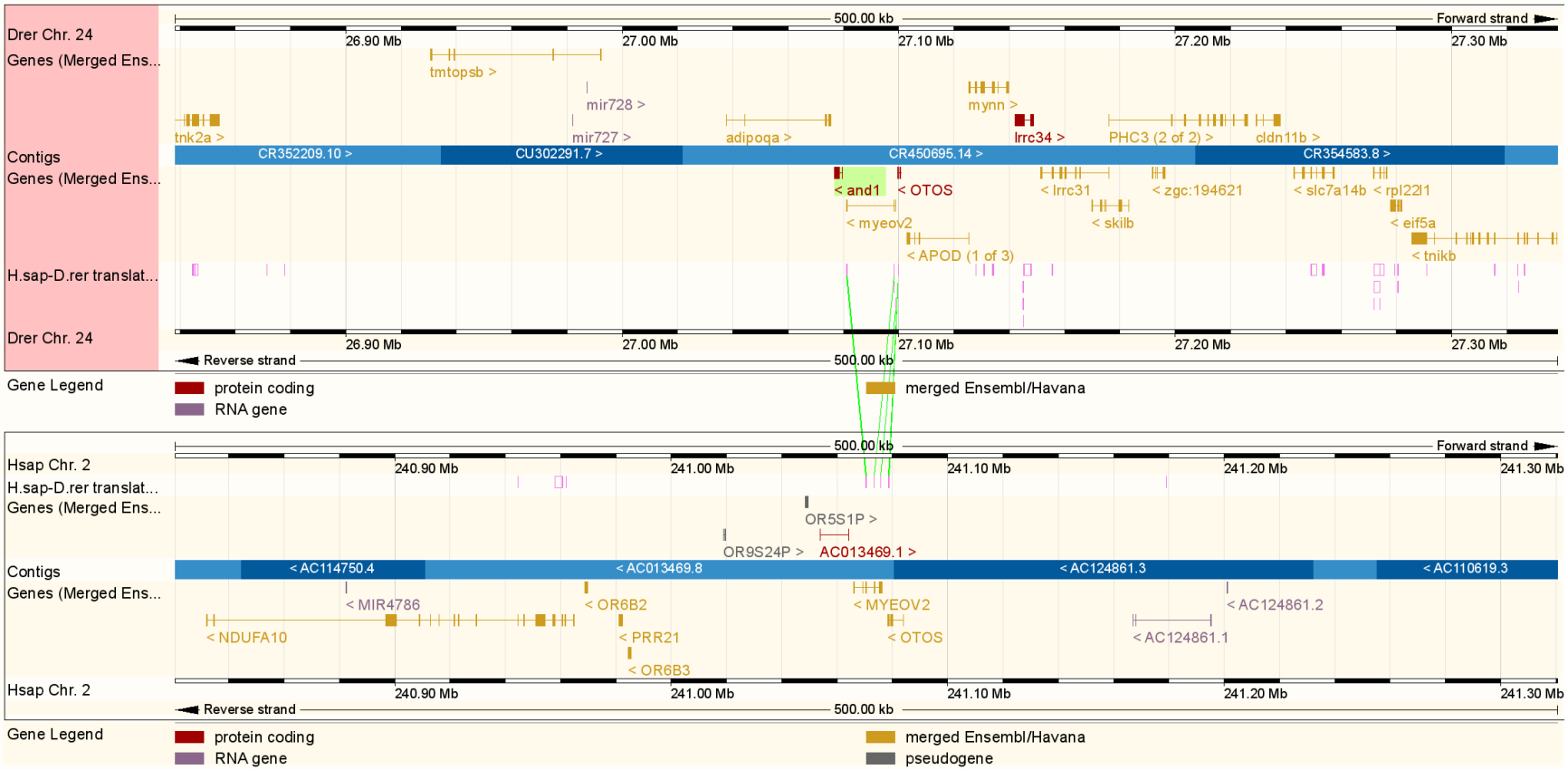
Zebrafish and Chicken



Zebrafish and Rat

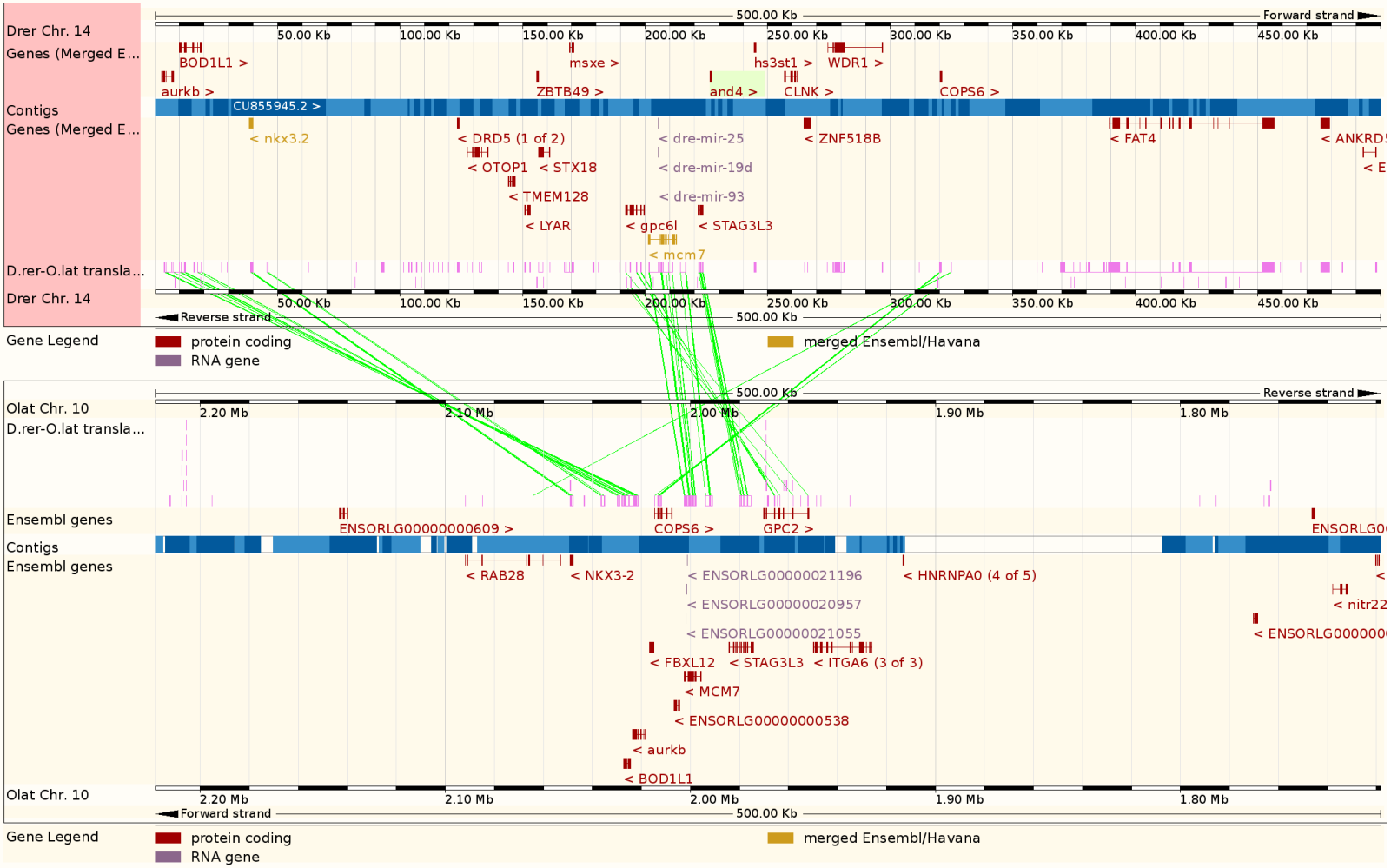


Zebrafish and Human

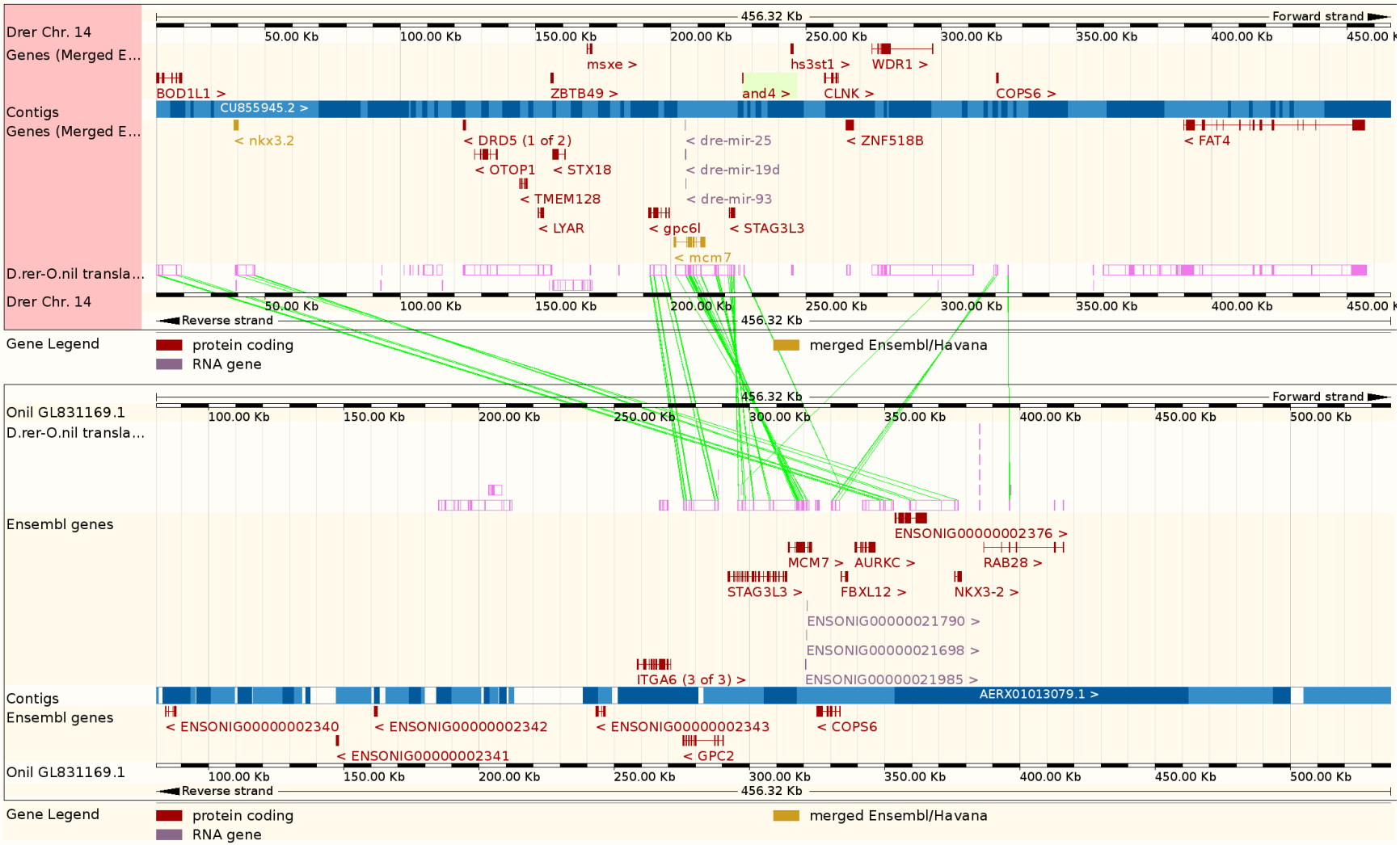


Appendix 5D. Synteny of *and4* loci in zebrafish (*Danio rerio*) compared to: medaka (*Oryzias latipes*), tilapia (*Oreochromis niloticus*), cod (*Gadus morhua*), fugu (*Takifugu rubripes*), platyfish (*Xiphophorus maculatus*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), and coelacanth (*Latimeria chalumnae*). Synteny graphics produced with Ensembl.org "region comparison" tool.

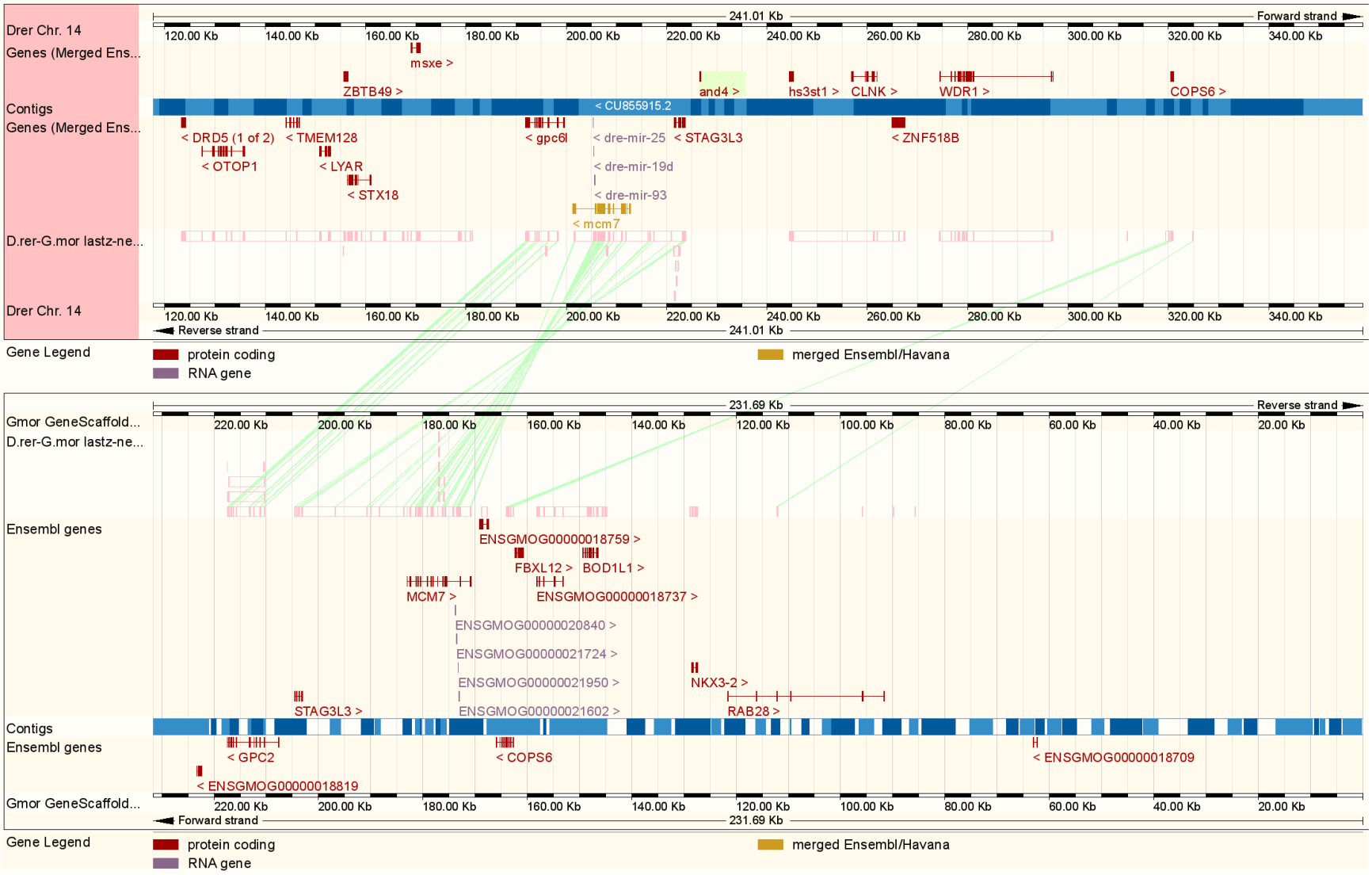
Zebrafish and Medaka



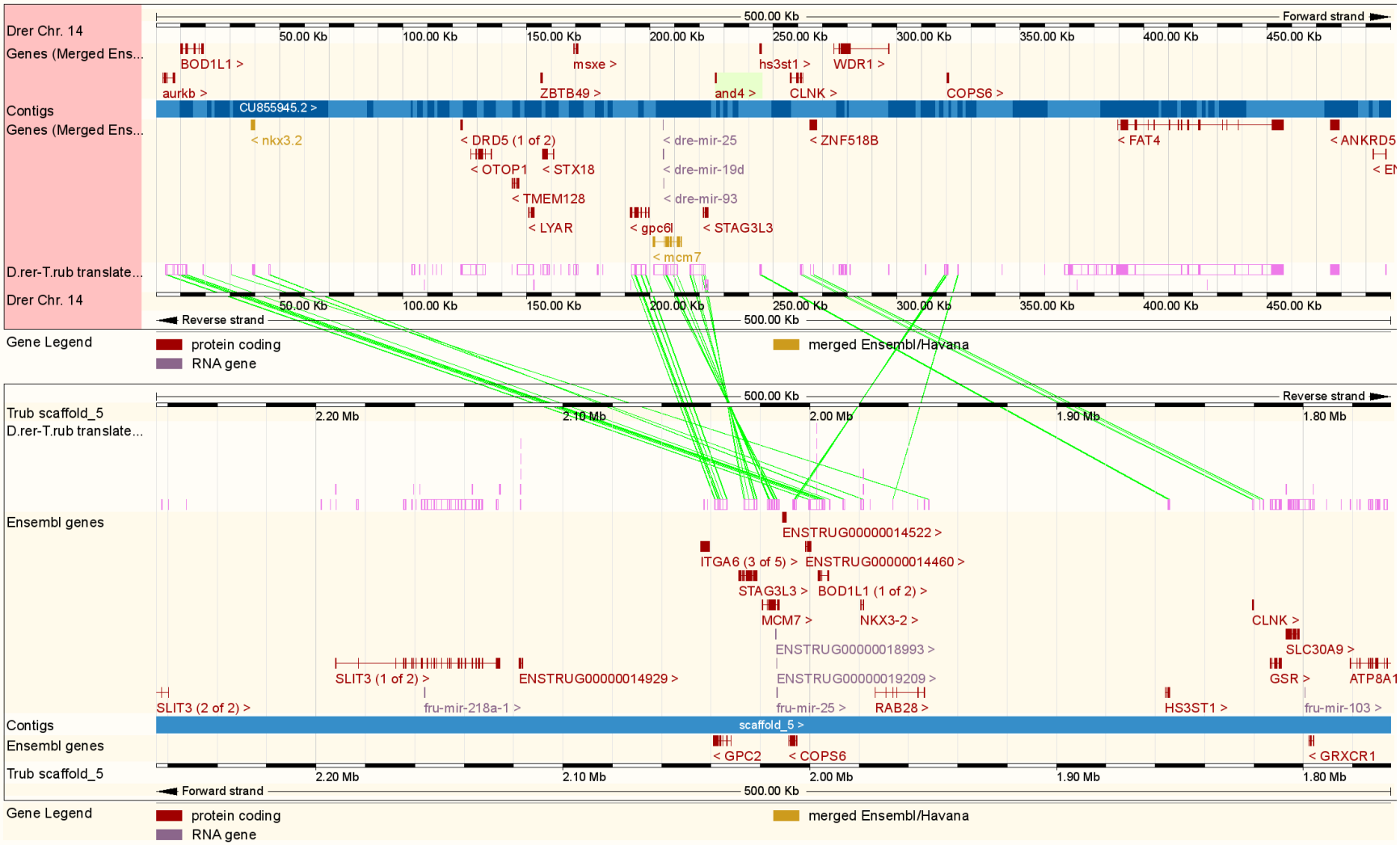
Zebrafish and Tilapia



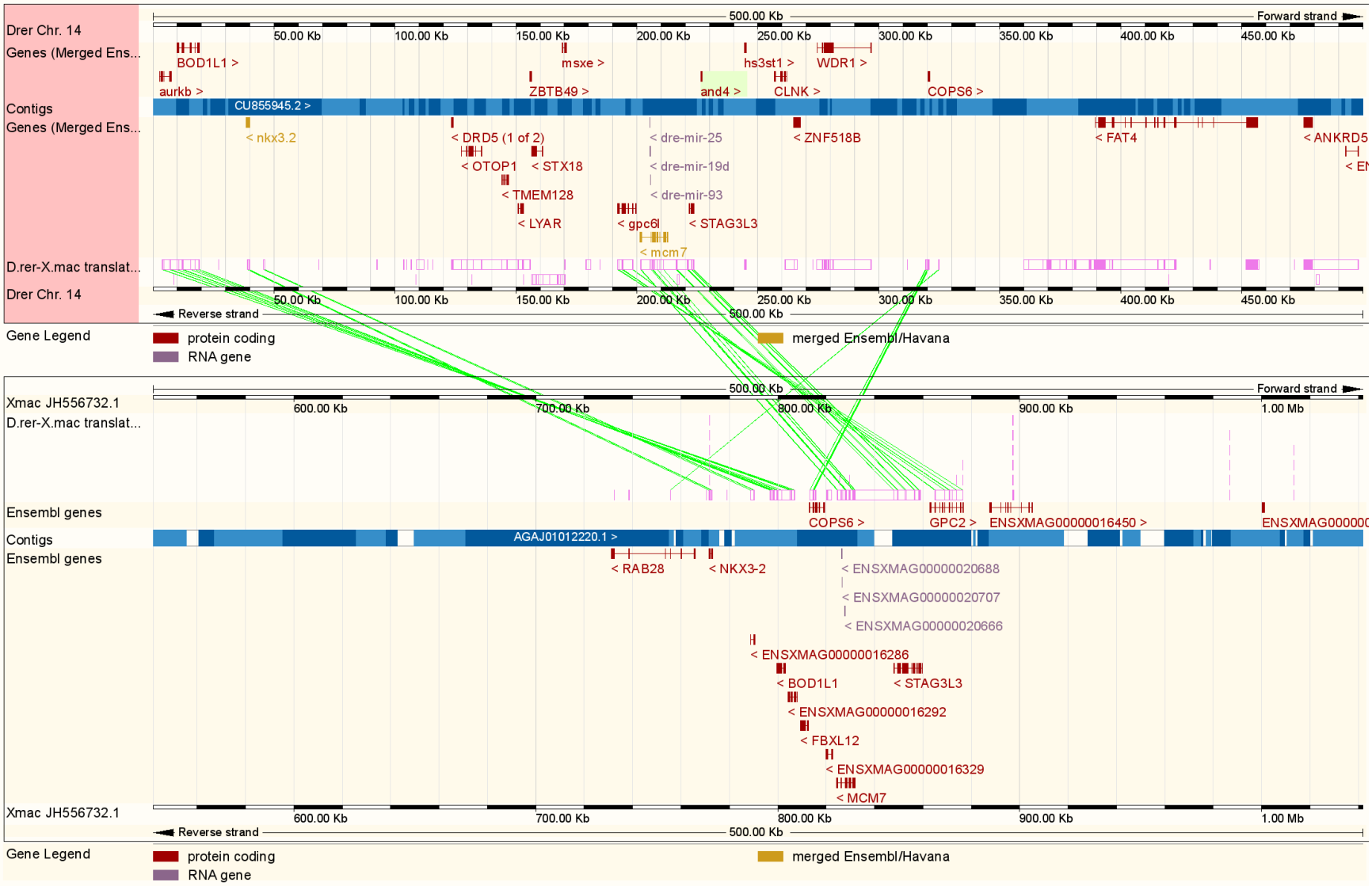
Zebrafish and Cod



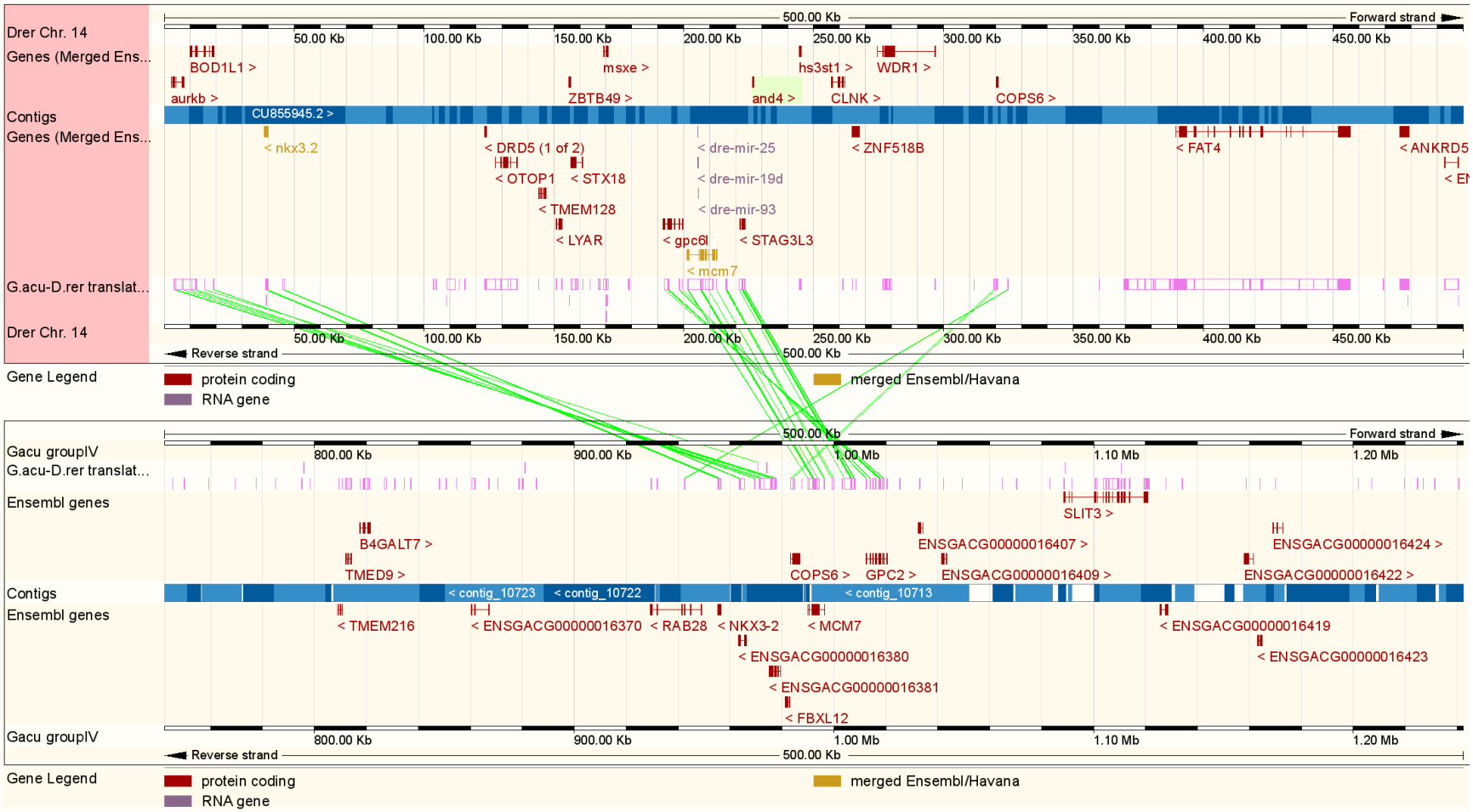
Zebrafish and Fugu



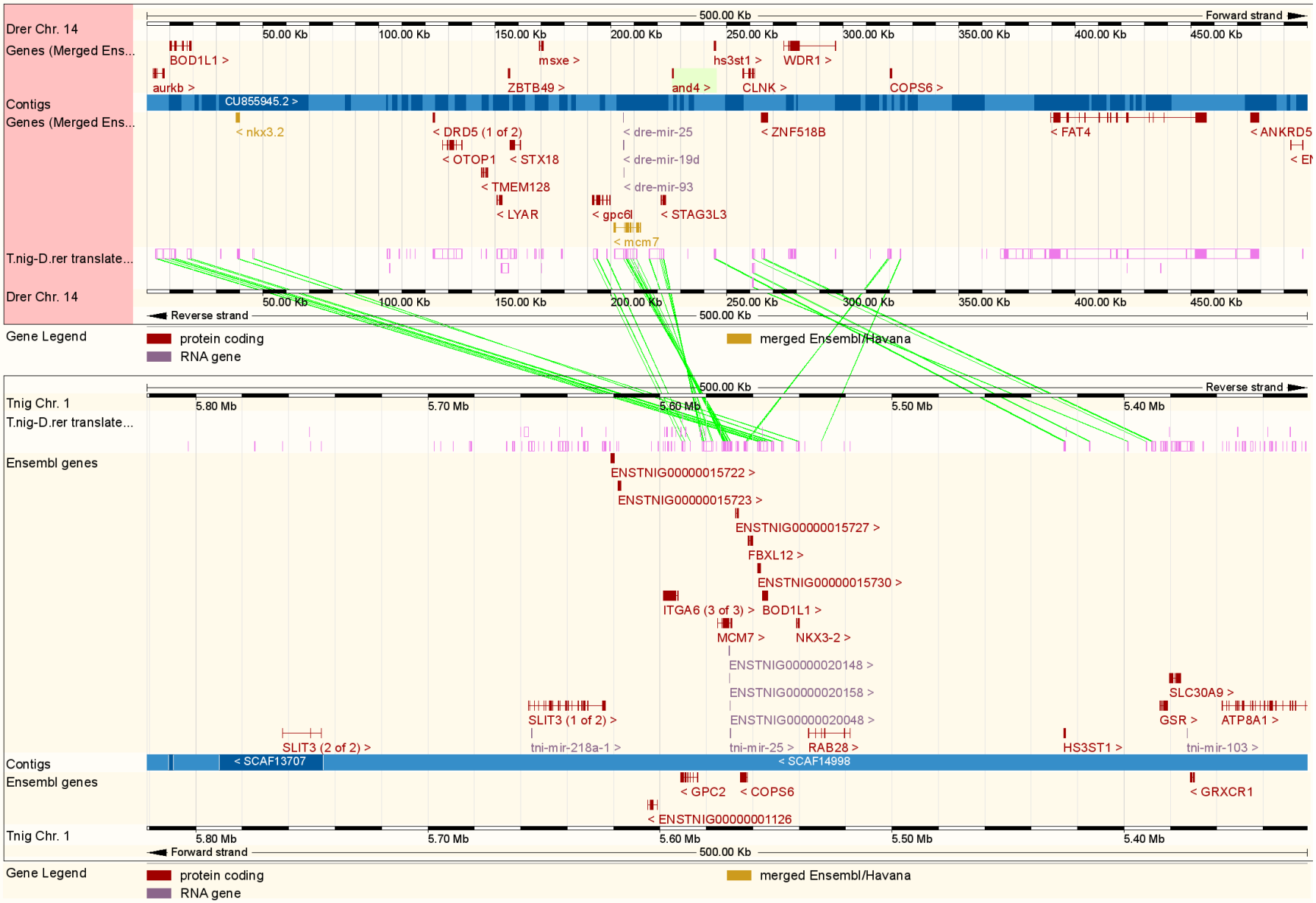
Zebrafish and Platyfish



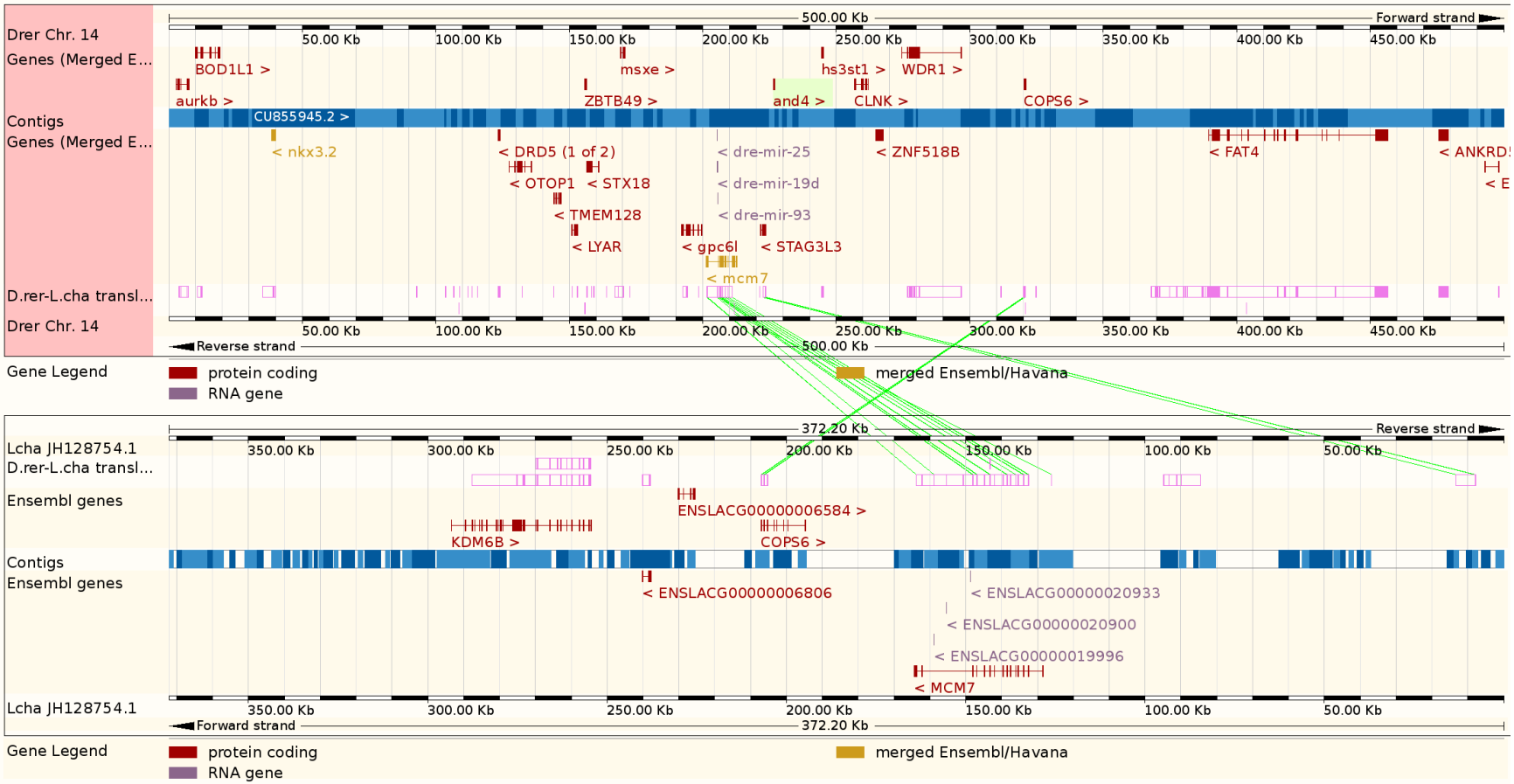
Zebrafish and Stickleback



Zebrafish and Tetraodon

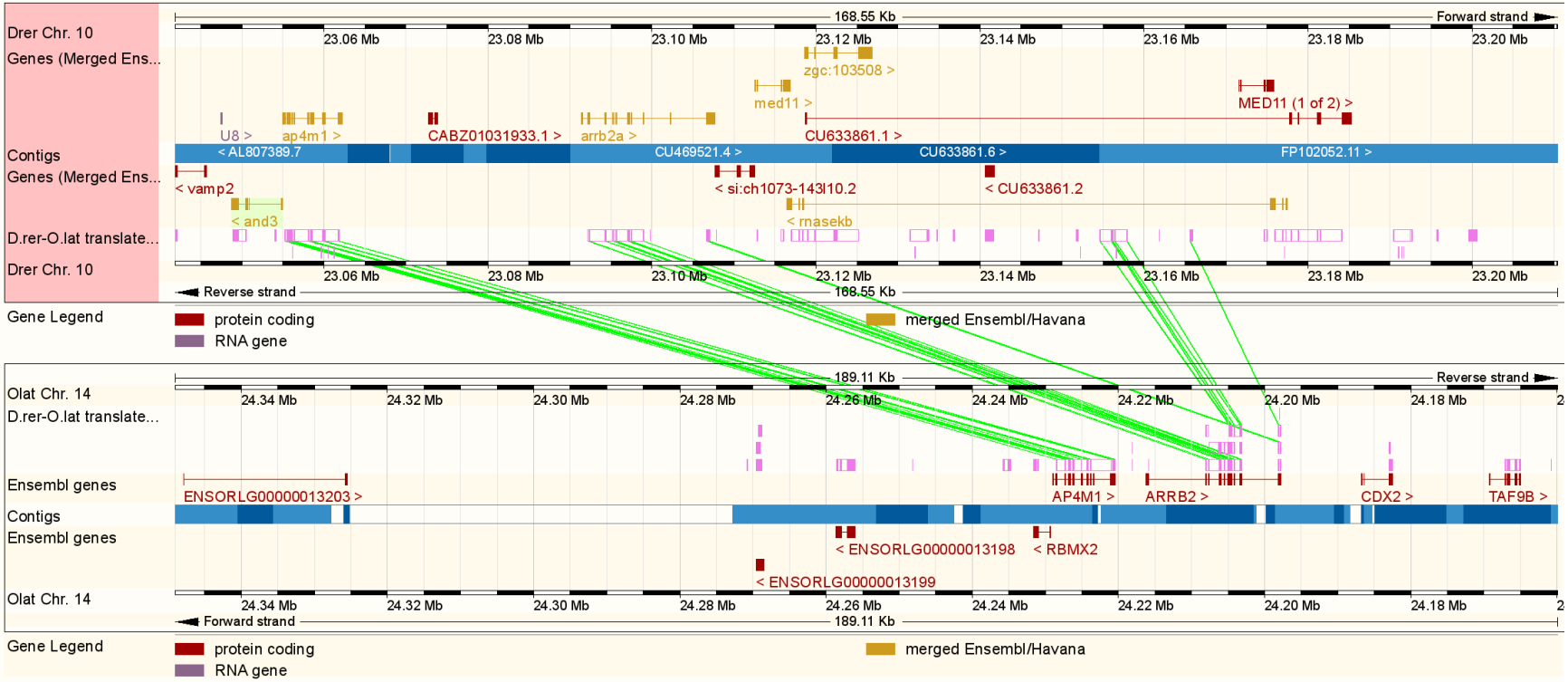


Zebrafish and Coelacanth

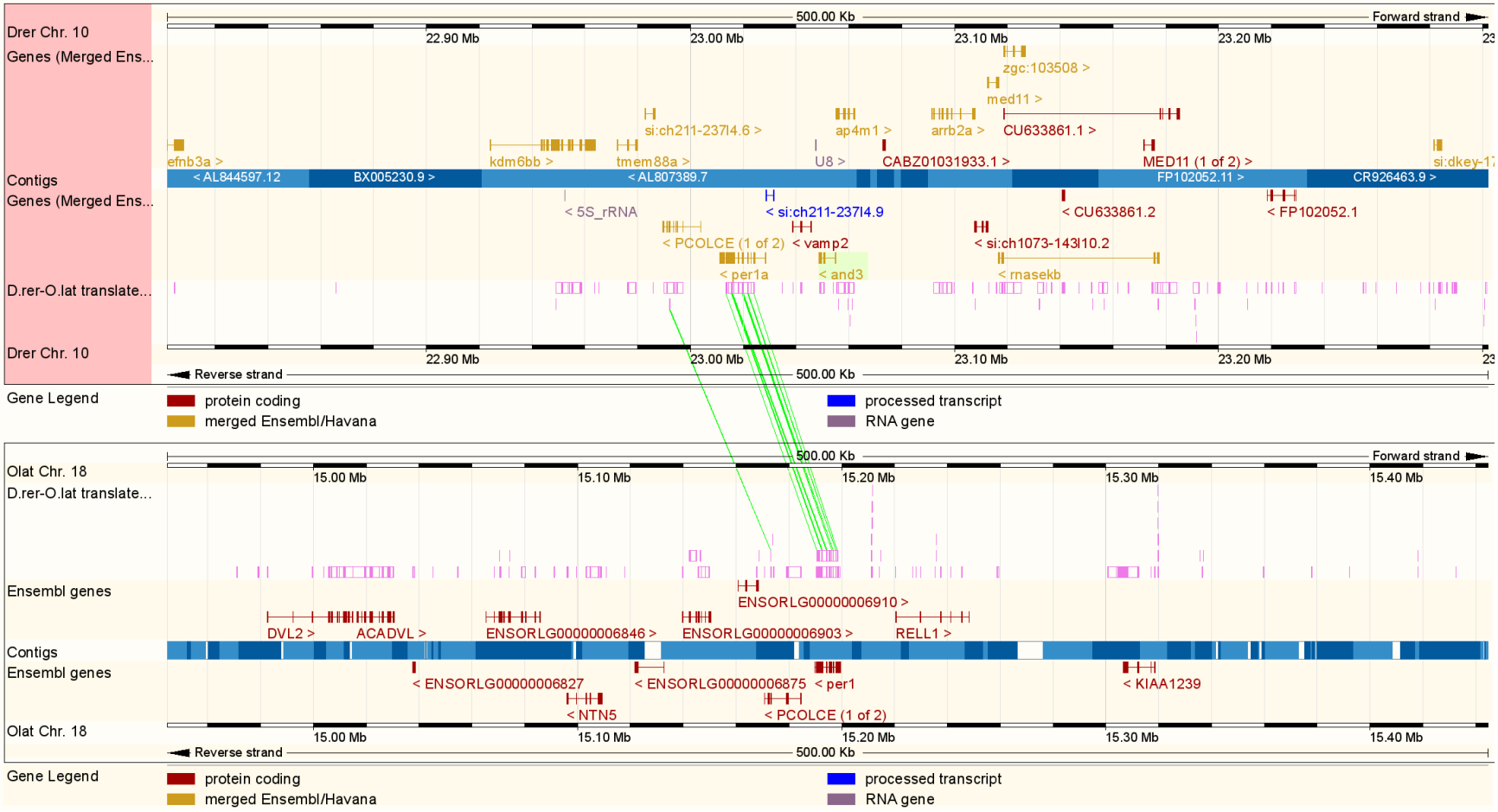


Appendix 5E. Synteny of *and3* loci in zebrafish (*Danio rerio*) compared to: medaka (*Oryzias latipes*), tilapia (*Oreochromis niloticus*), and coelacanth (*Latimeria chalumnae*). Synteny graphics produced with Ensembl.org "region comparison" tool.

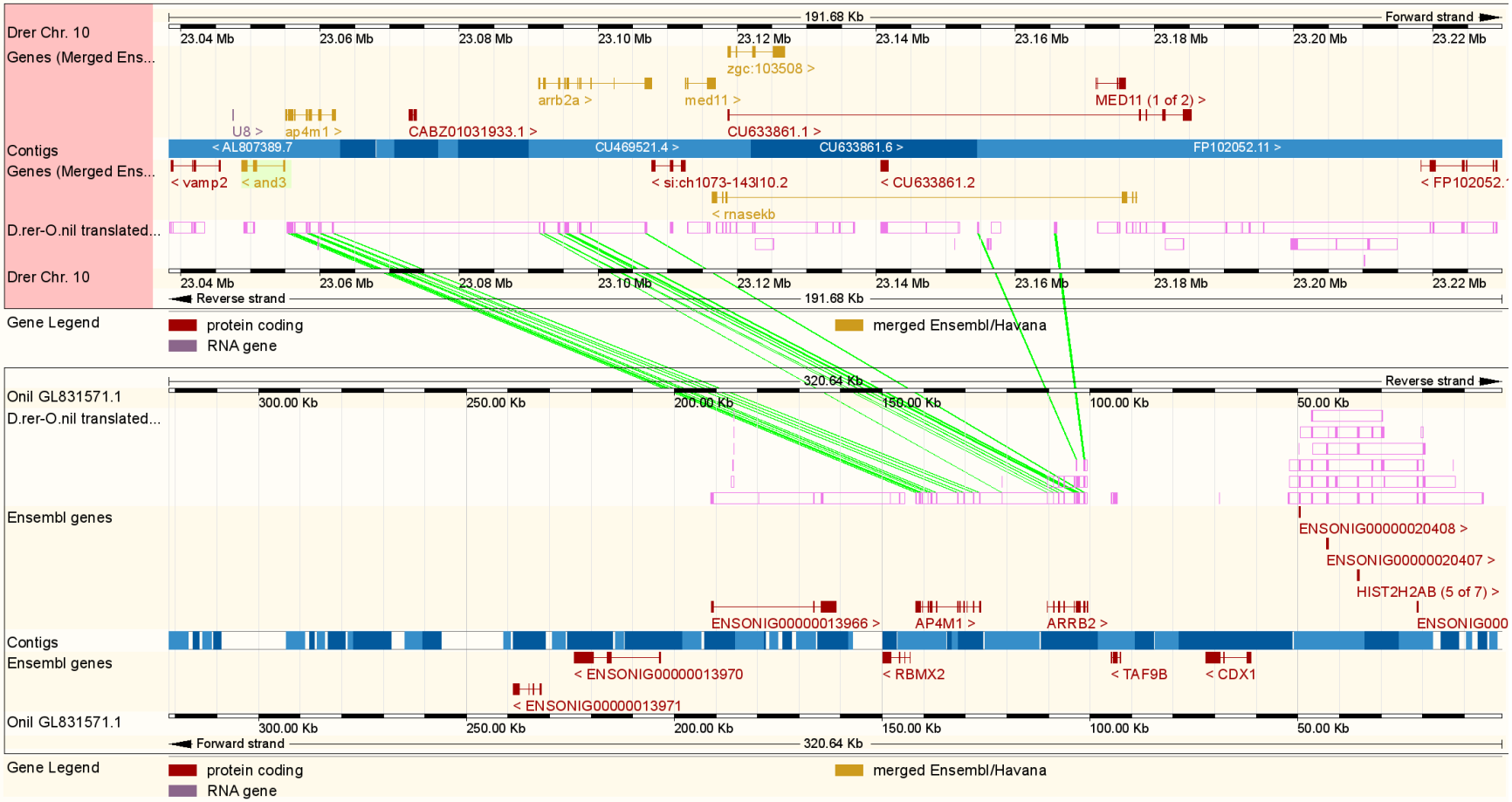
Zebrafish and Medaka a



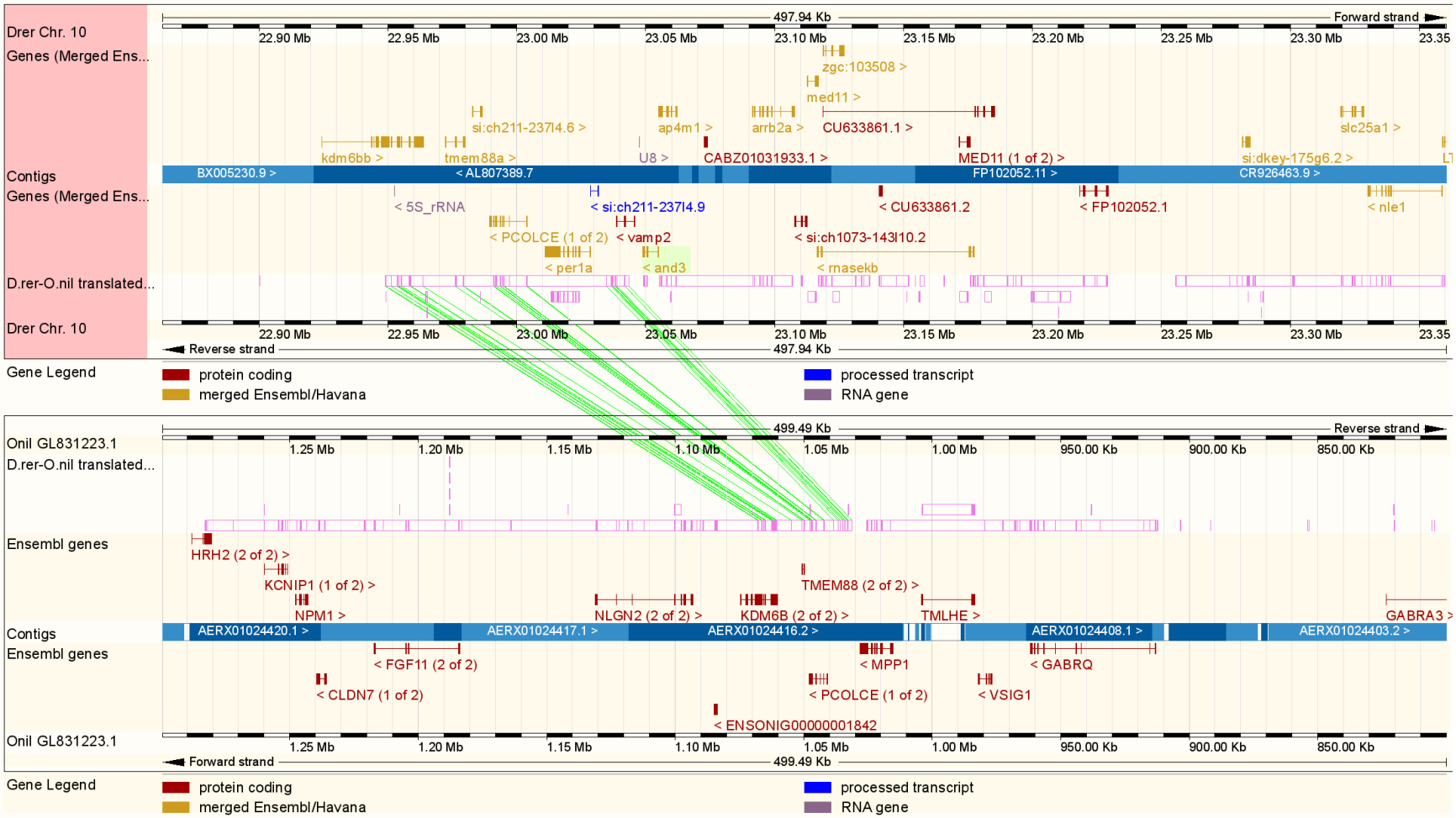
Zebrafish and Medaka b



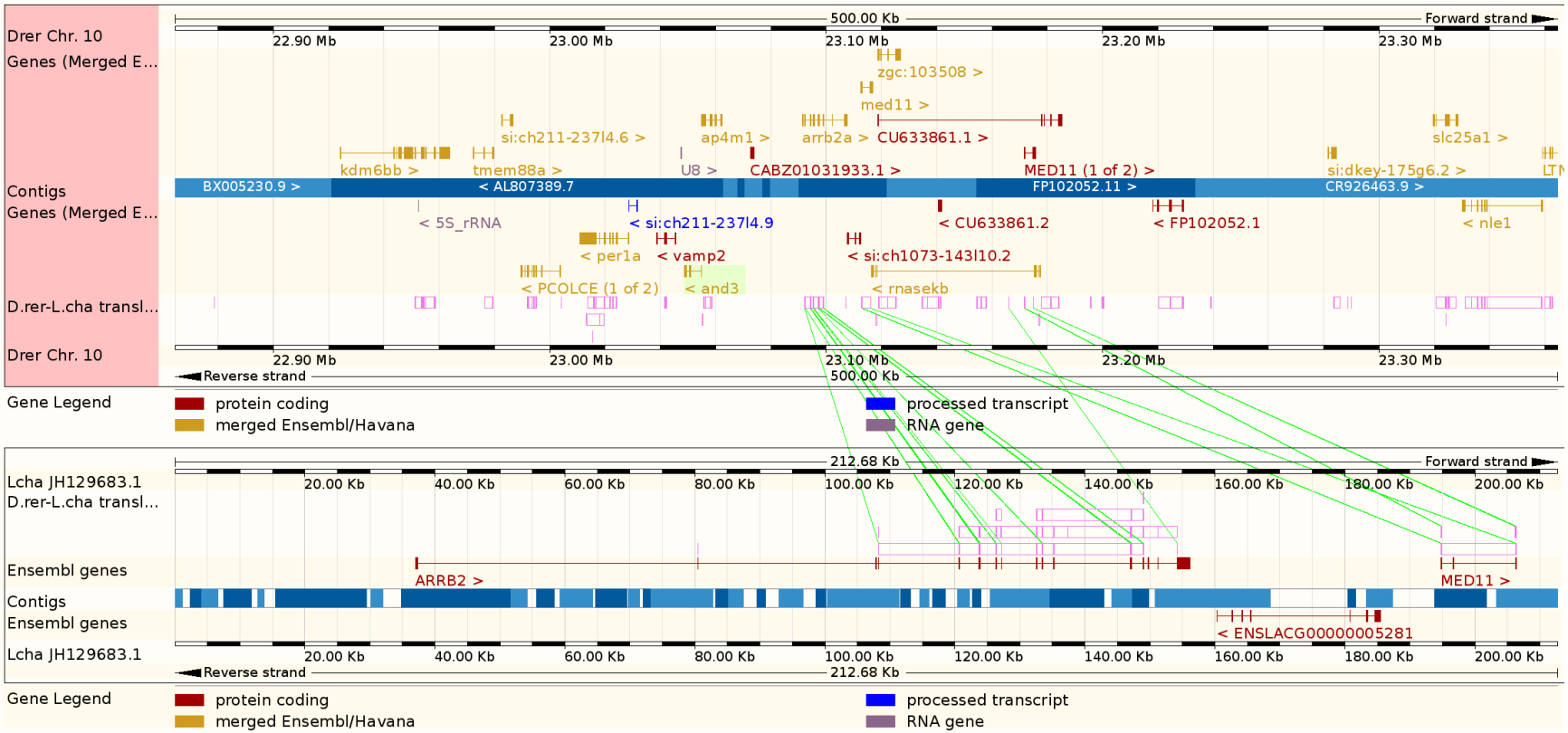
Zebrafish and Tilapia a



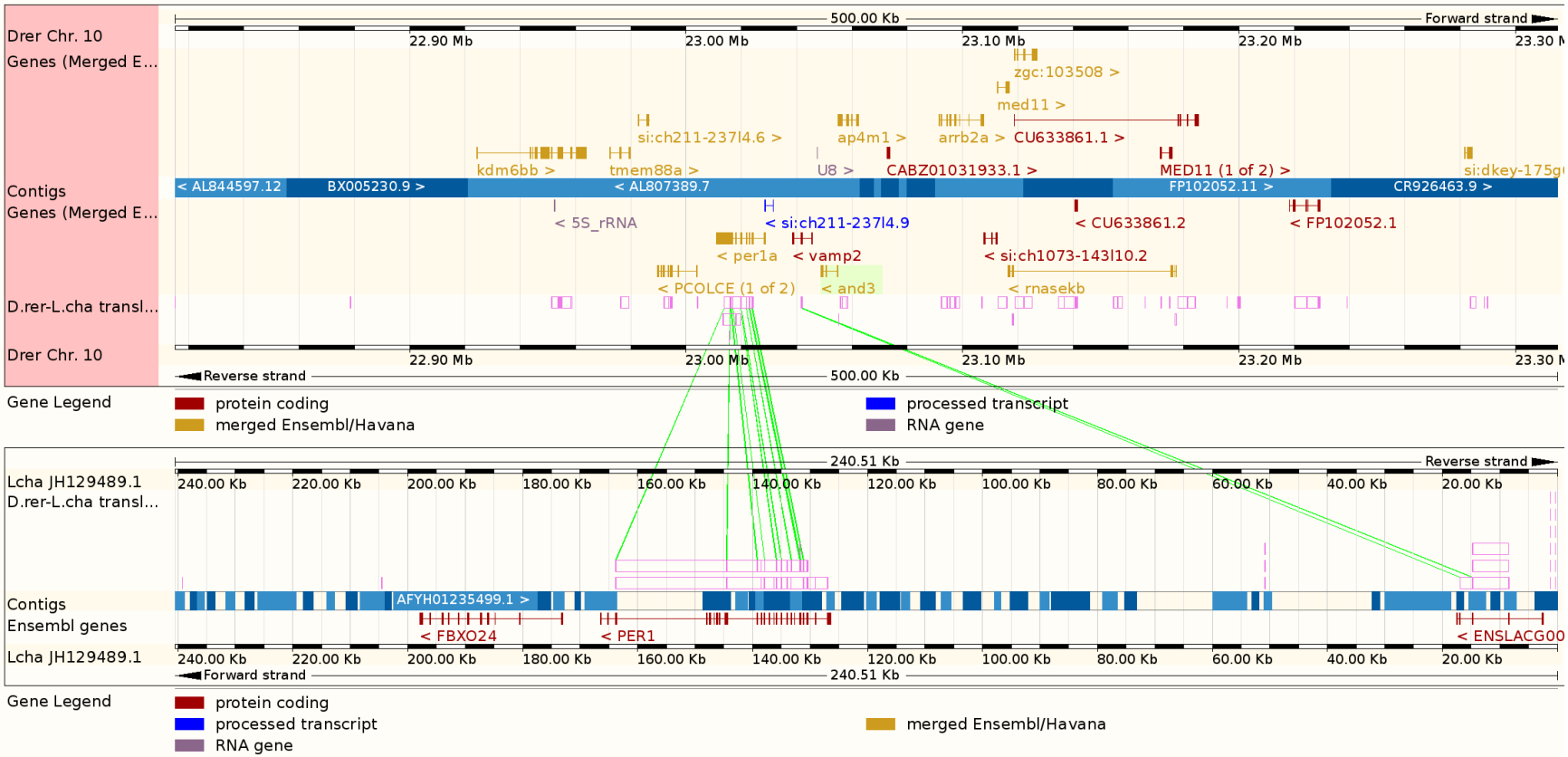
Zebrafish and Tilapia b



Zebrafish and Coelacanth a

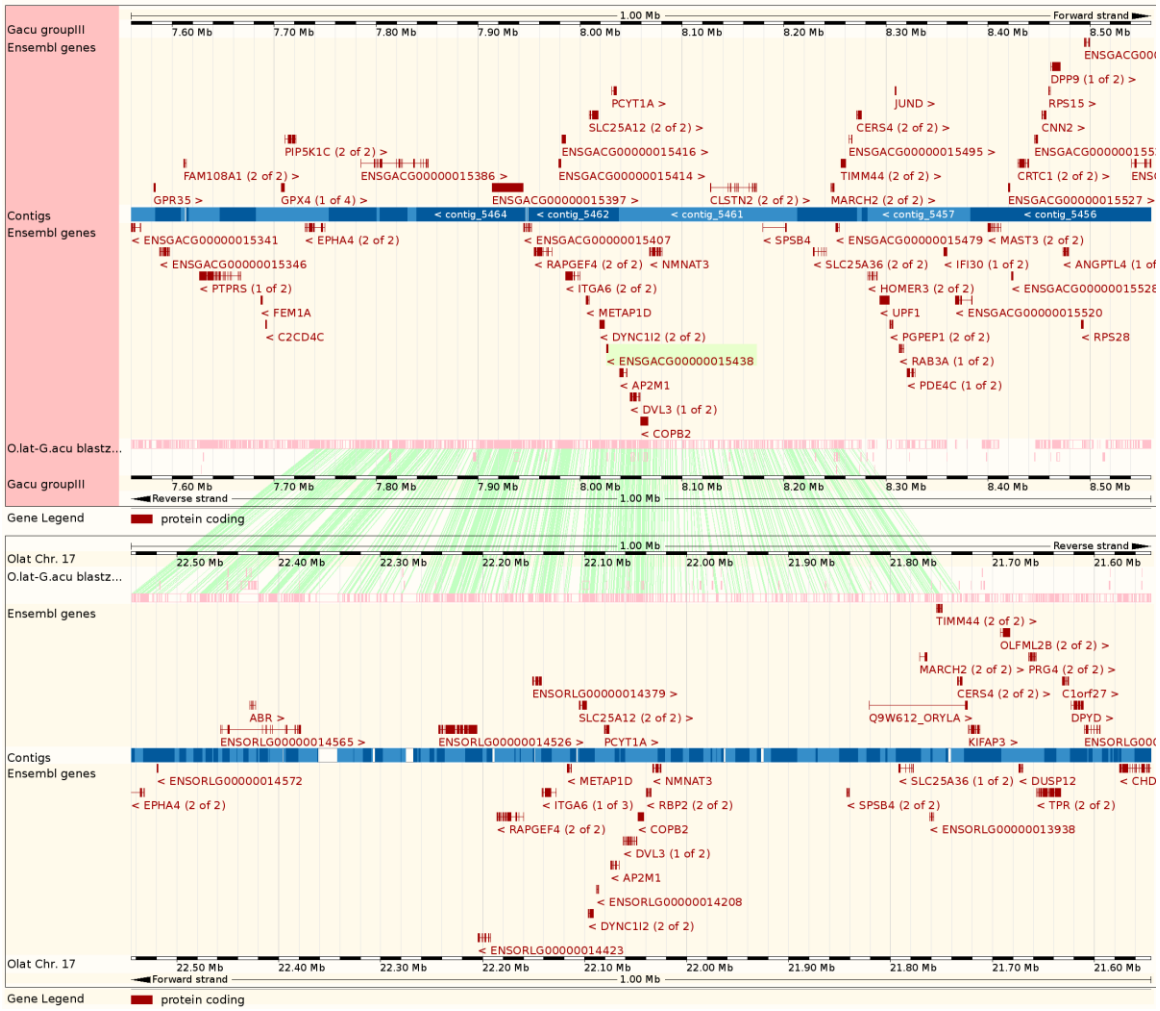


Zebrafish and Coelacanth b

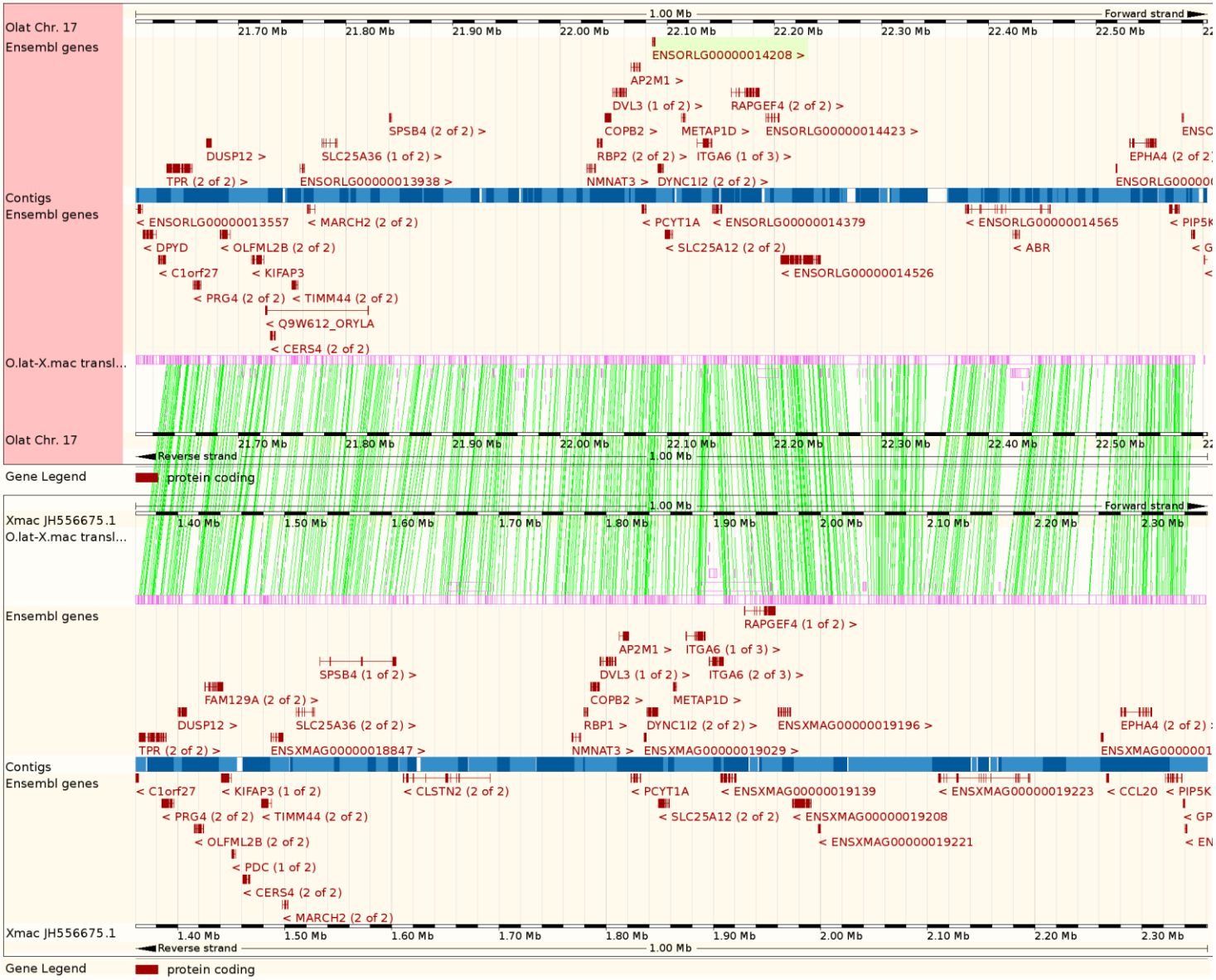


Appendix 5F. Synteny of *and3* loci in Acanthomorpha. Direct comparisons between: stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes*), platyfish (*Xiphophorus maculatus*) and coelacanth (*Latimeria chalumnae*). Loci shown for fugu (*Takifugu rubripes*) and tetraodon (*Tetraodon nigroviridis*) from which comparisons cannot be produced using the "region comparison" tool. All synteny graphics produced with Ensembl.org "region comparison" tool.

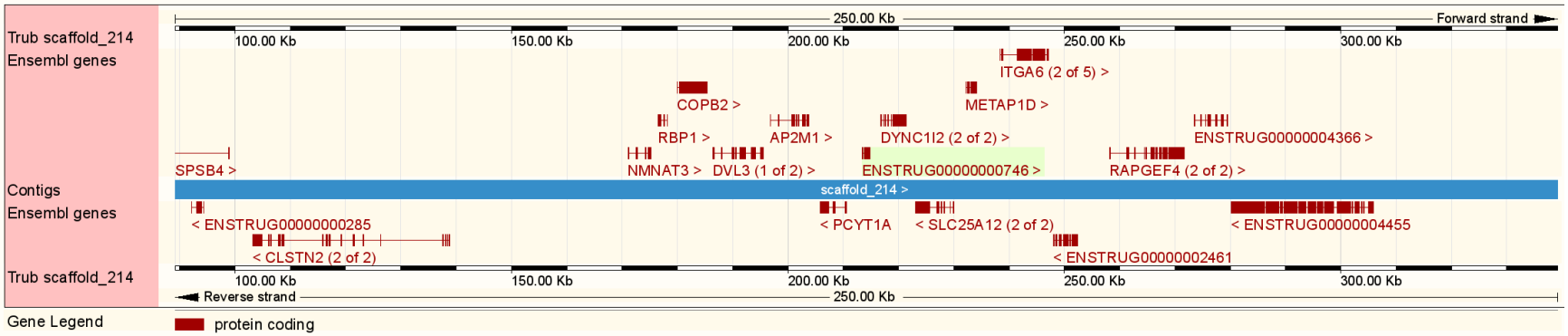
Stickleback and Medaka



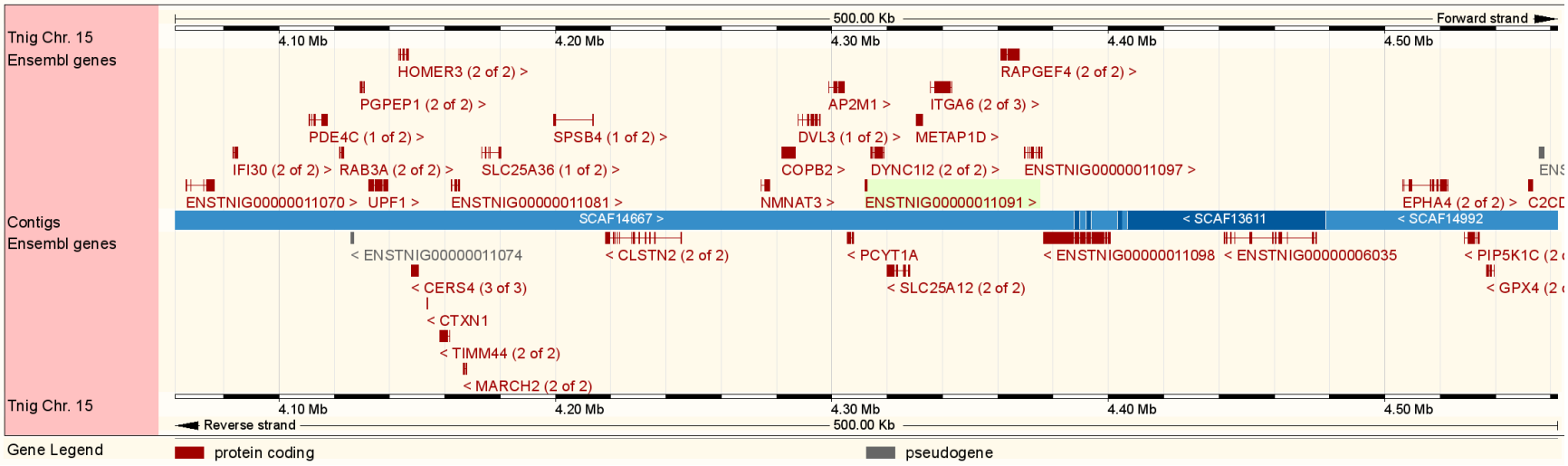
Medaka and Platyfish



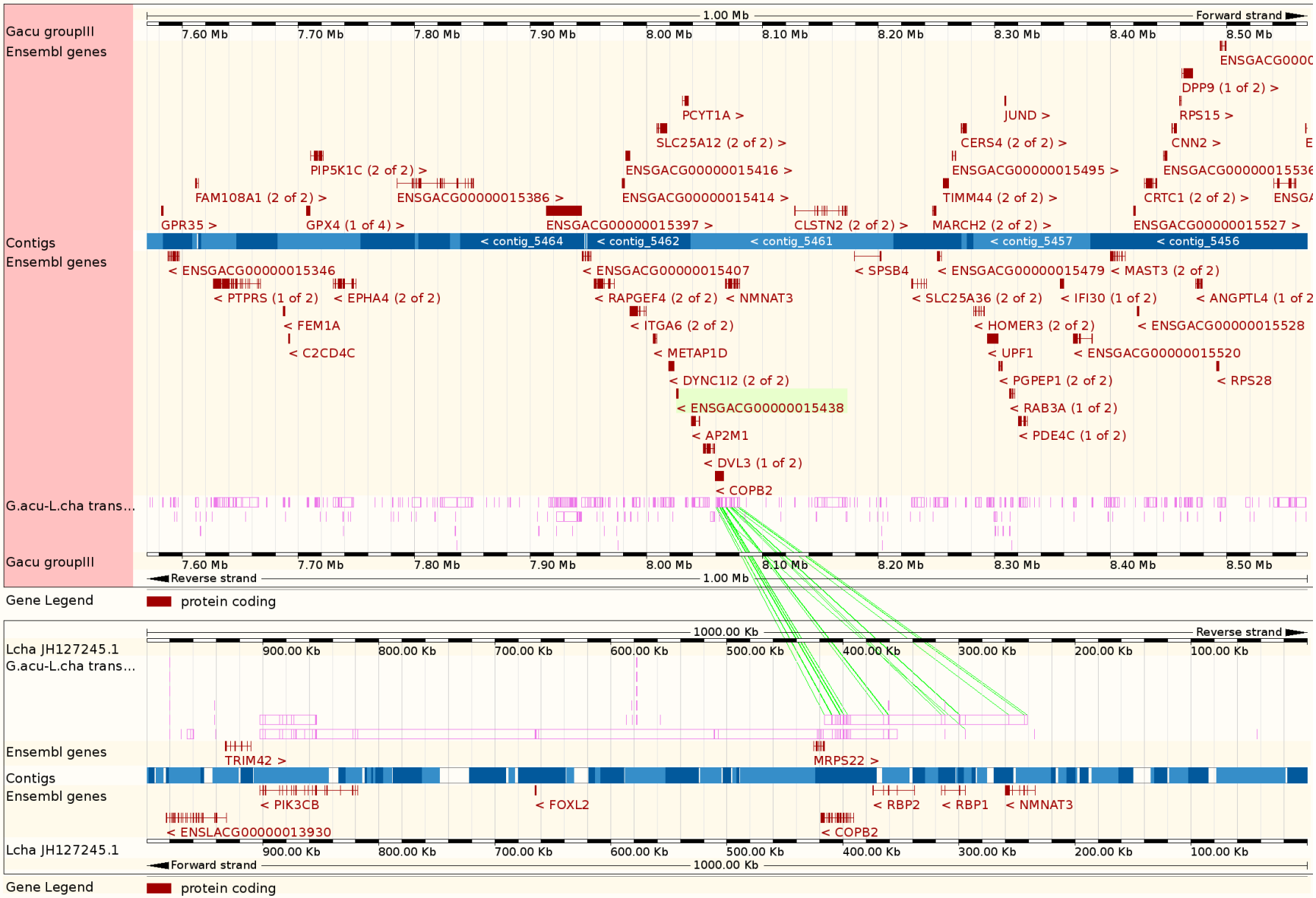
Fugu



Tetraodon

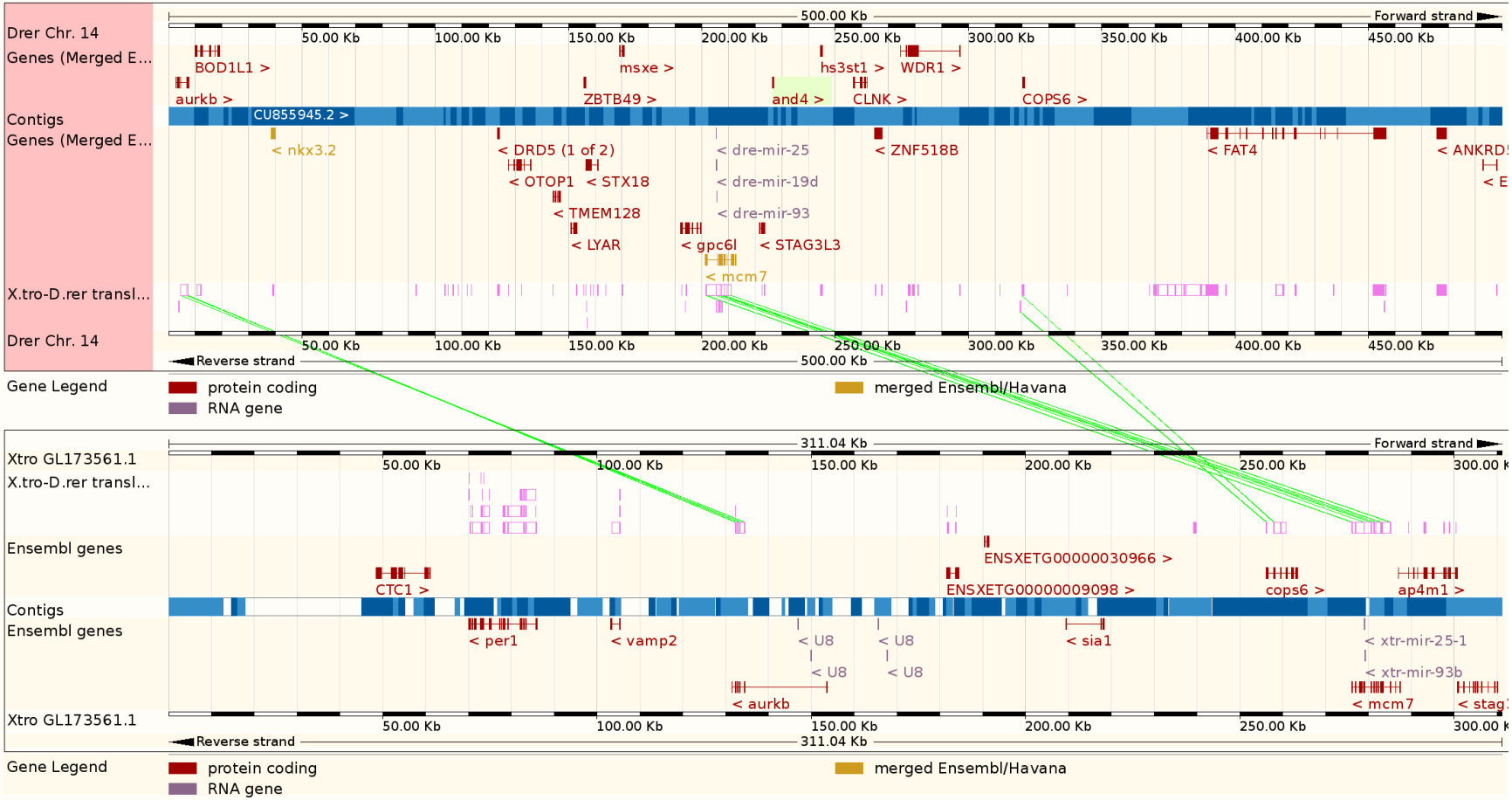


Stickleback and Coelacanth

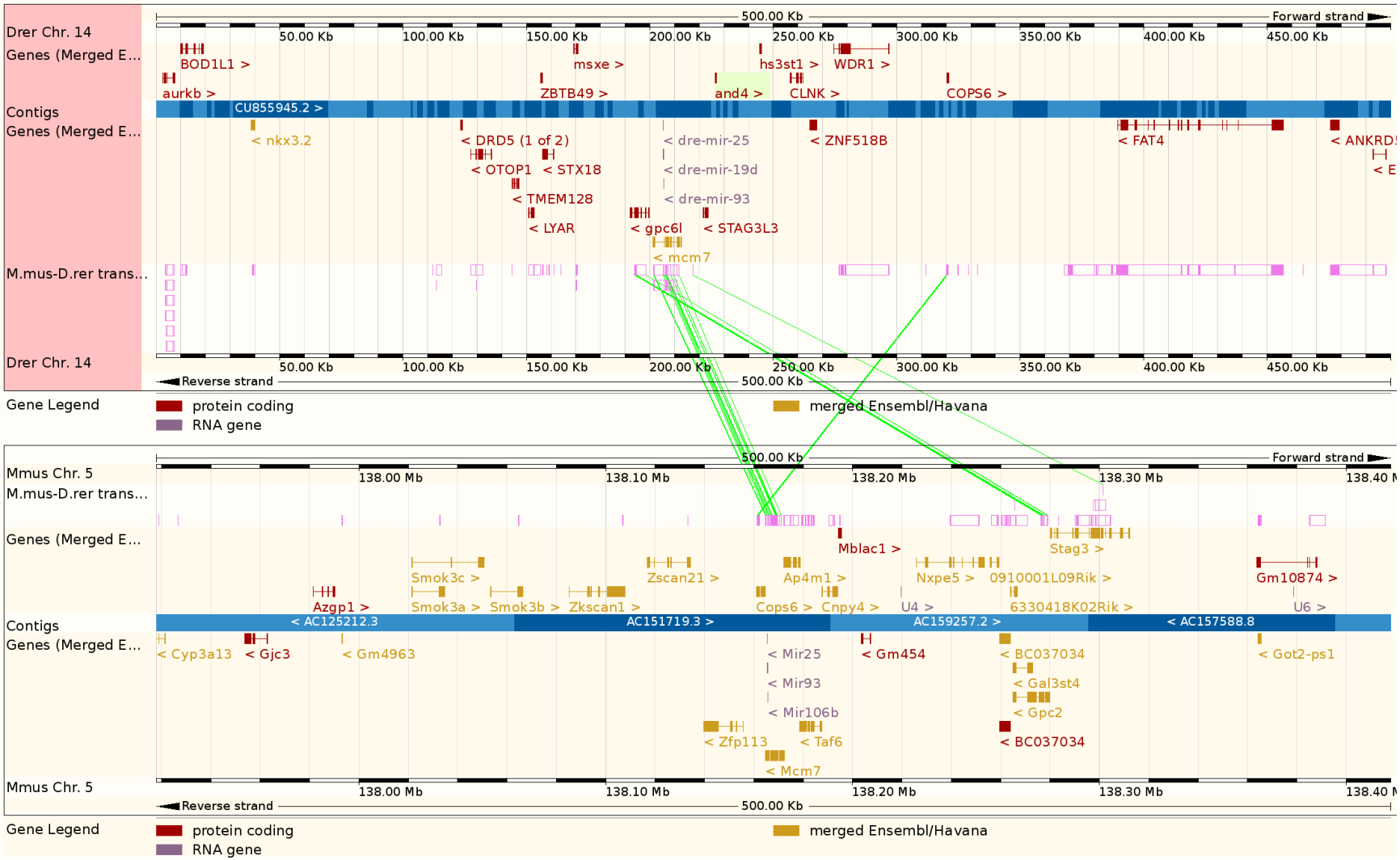


Appendix 5G. Synteny of *and4* loci in zebrafish (*Danio rerio*) compared to: *Xenopus* (*Xenopus tropicalis*), mouse (*Mus musculus*), rat (*Rattus norvegicus*) and human (*Homo sapiens*). Synteny graphics produced with Ensembl.org "region comparison" tool.

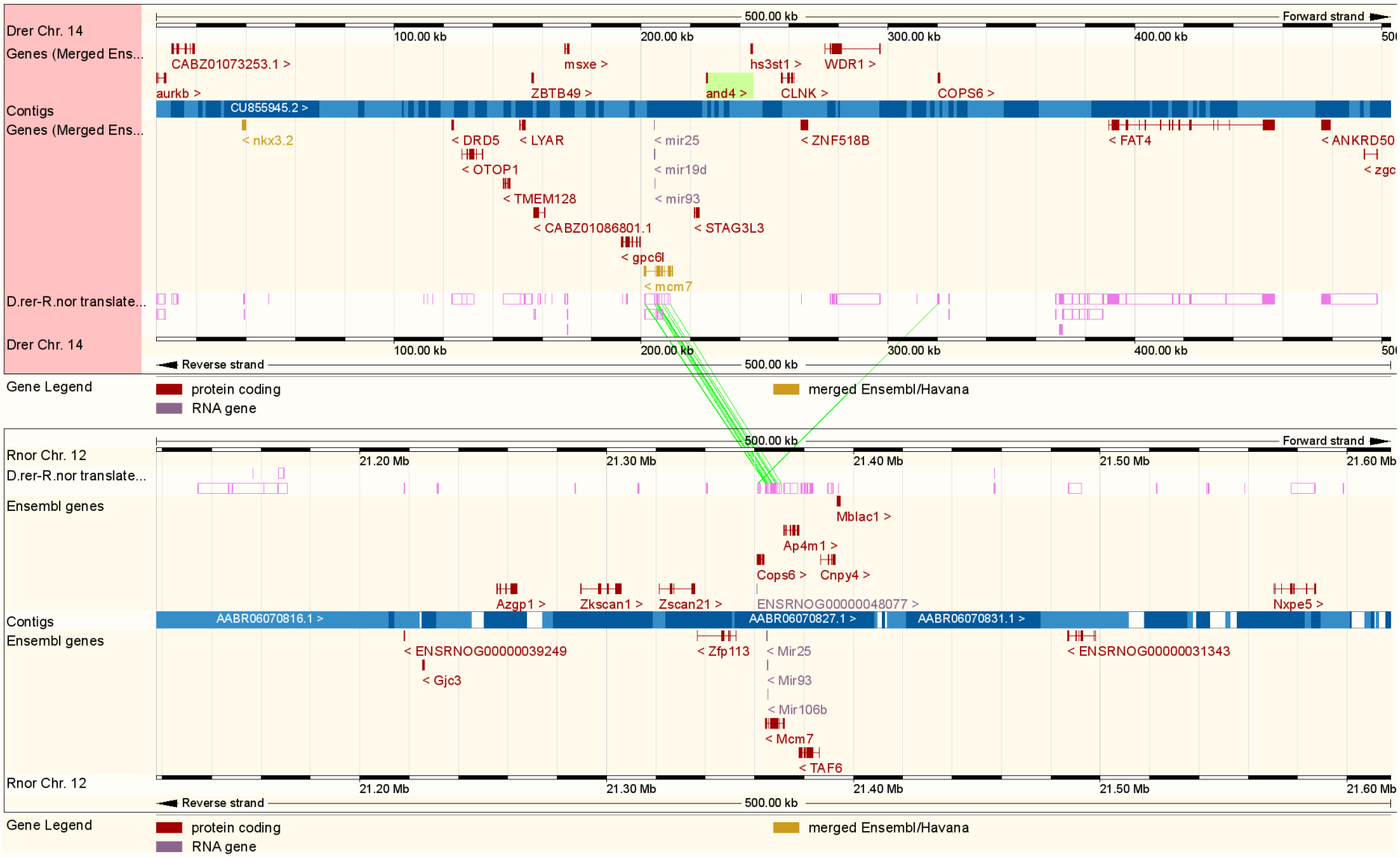
Zebrafish and Xenopus



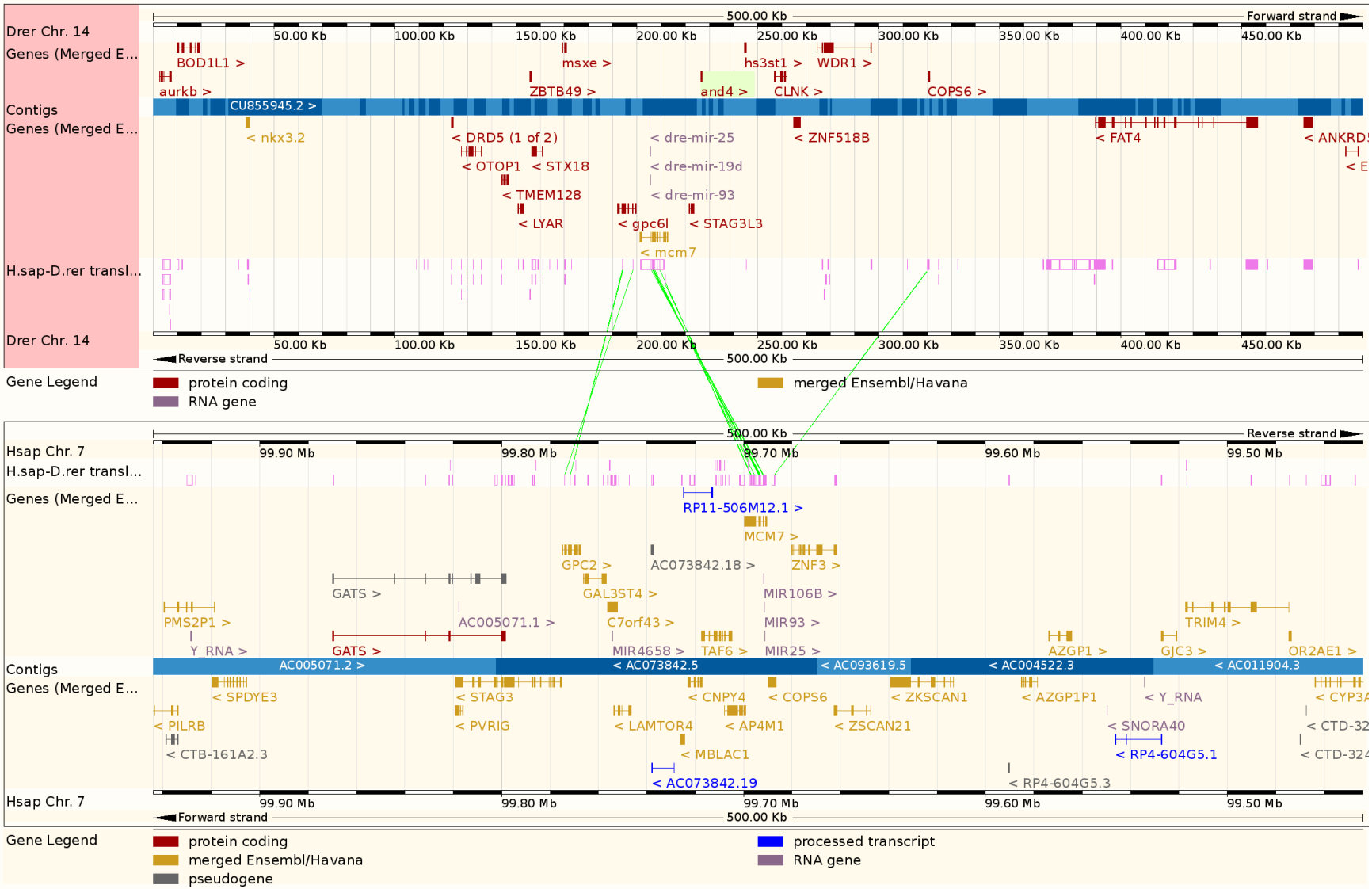
Zebrafish and Mouse



Zebrafish and Rat

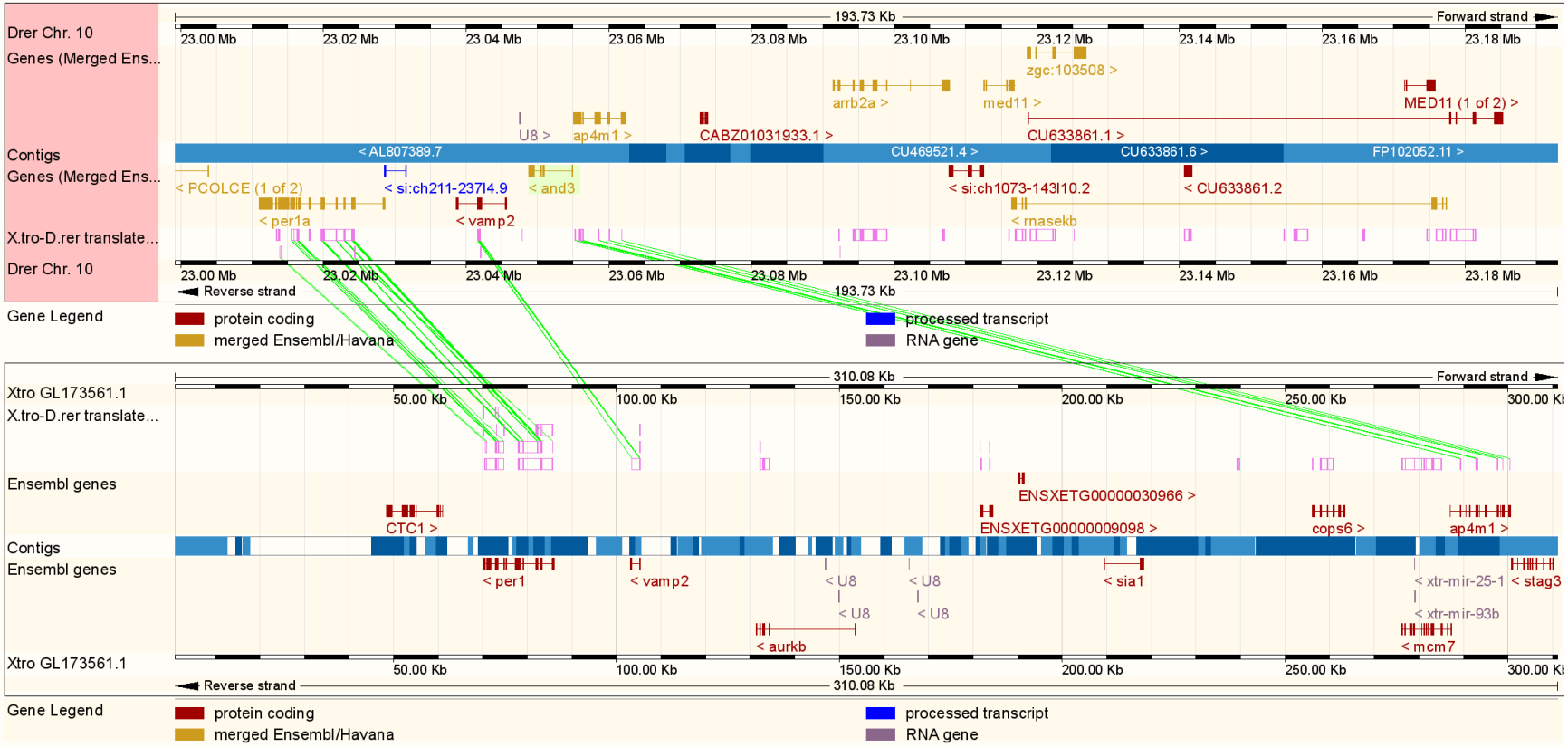


Zebrafish and Human

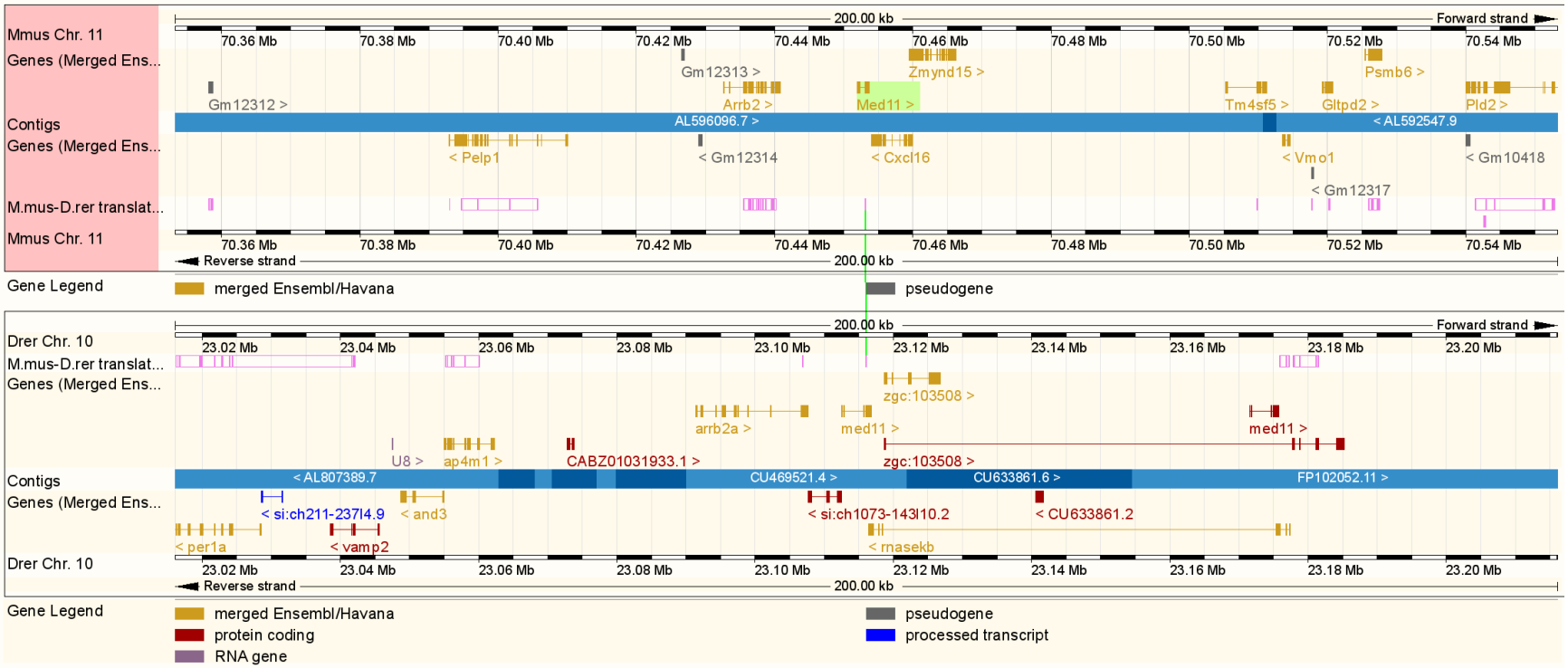


Appendix 5H. Synteny of *and3* loci in zebrafish (*Danio rerio*) compared to: *Xenopus* (*Xenopus tropicalis*), mouse (*Mus musculus*), and human (*Homo sapiens*). Synteny graphics produced with Ensembl.org "region comparison" tool.

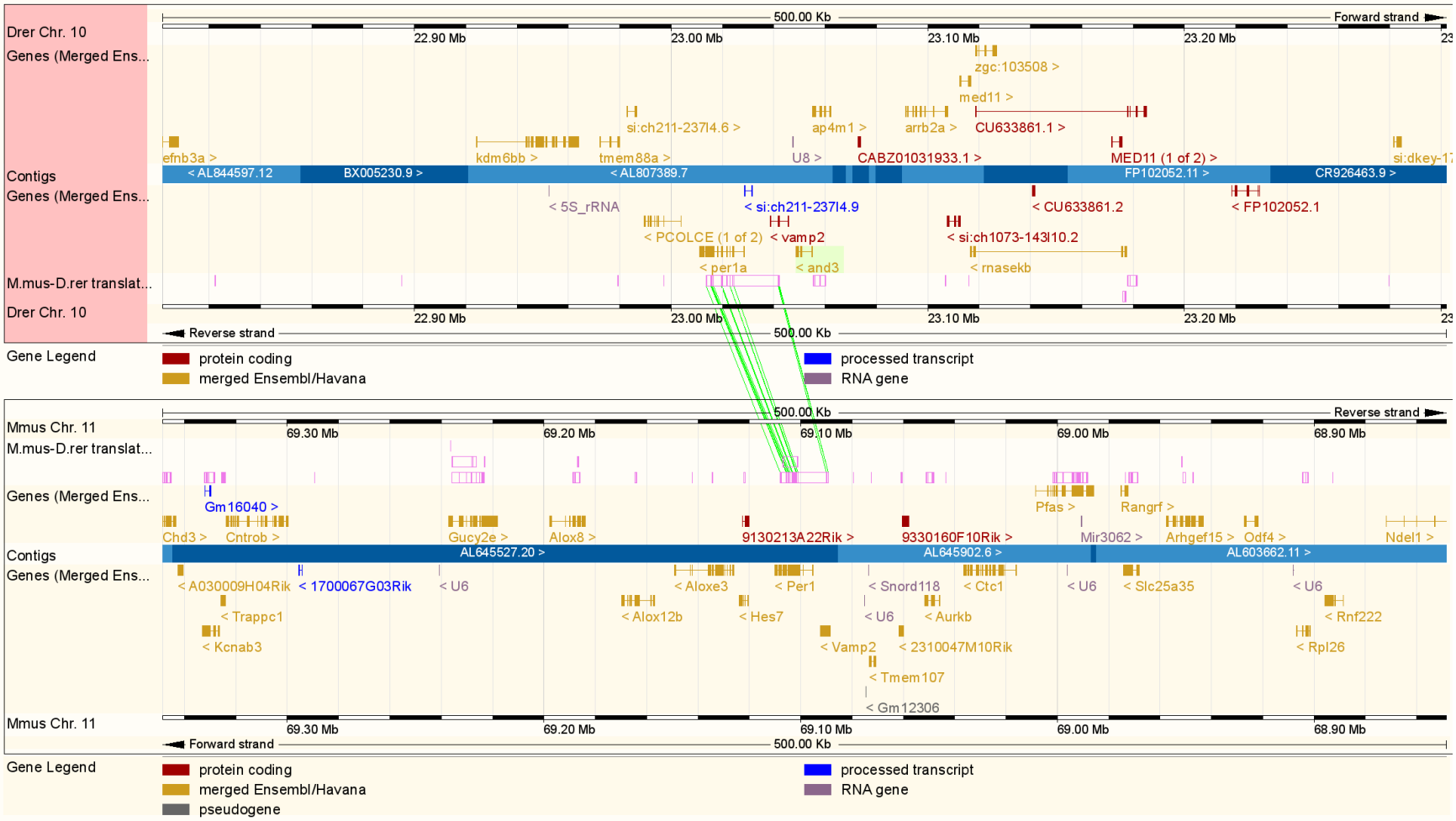
Zebrafish and *Xenopus*



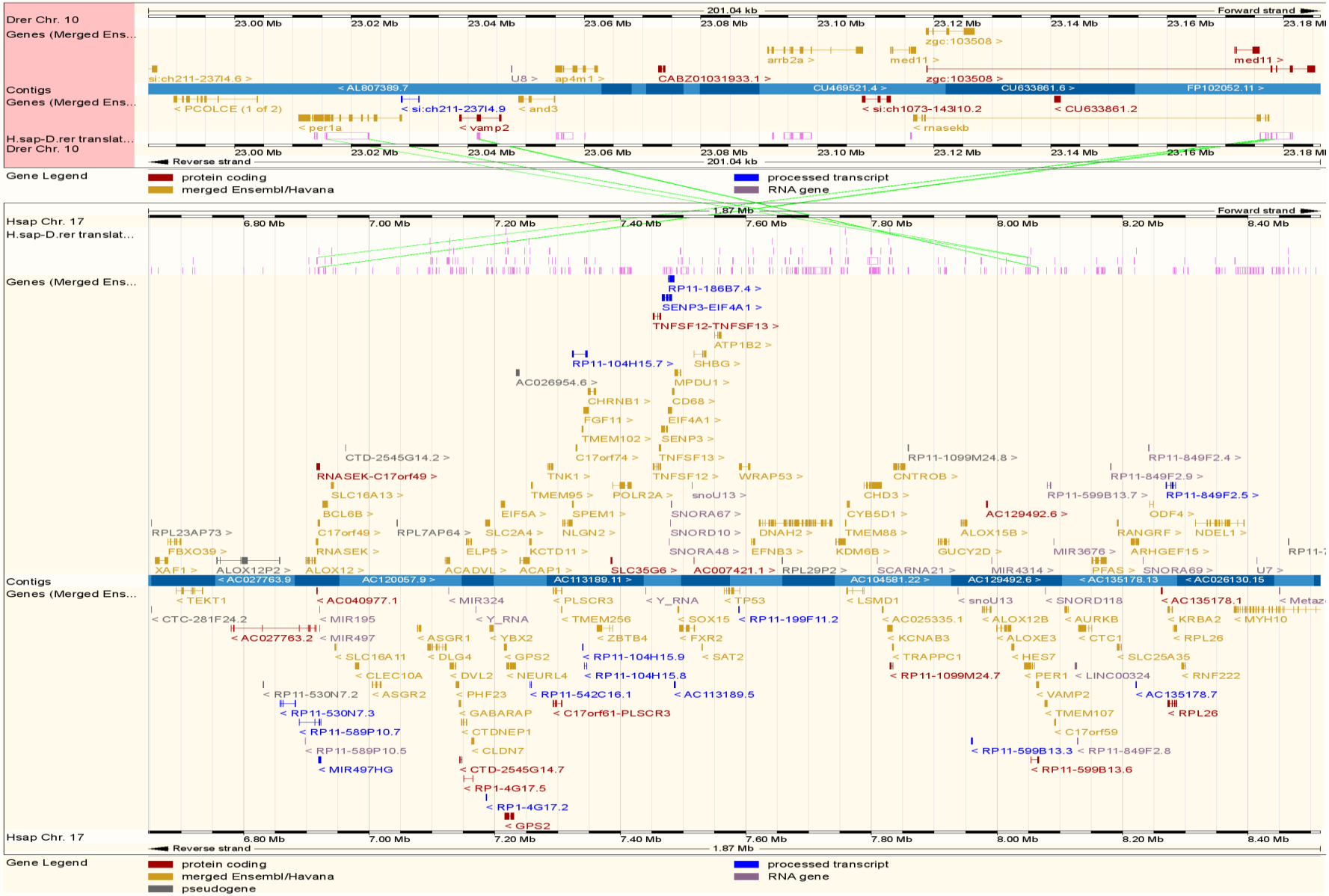
Zebrafish and Mouse a



Zebrafish and Mouse b



Zebrafish and Human a



Zebrafish and Human b