Hedmark University
of Applied Sciences

Faculty of Education and natural science
Department of Natural Science and Technology

Deependra Yadav

Development of gender-specific PCR test for Salmonids

M.sc. in Applied and Commercial Biotechnology

Master's Thesis 2016

Consent to lending by University College Library YES☐ NO☐

Consent to accessibility in digital archive Brage YES☐ NO☐

Acknowledgement

I would like to express my special appreciation and thanks to my supervisor Professor Robert Wilson, Associate Prof. Wenche Johansen, without your guidance and encouragement I could never achieve what I have done so far on my project. Thank you for your guidance in writing and suggestion to interoperate my results.

I would like to thank my supervisor Senior Scientist Jan Sunde for the help with the samples and valuable advice.

I also like to thank my friends and all members of the Biotechnology Group for their love, help support and encouragement.

Finally yet importantly, I thank my God, my parents for helping me through all the difficulties. I have experience your guidance day by day. I will keep on trusting you for my future. Thank you.

Hamar 2016,
Deependra Yadav

Abbreviations

AMH	anti-mullerian hormone
amhy	Y-linked anti-mullerian hormone gene
dmrt1	doublesex and mab-3-related transcription factor 1
dmy	DM domain gene on Y chromosome
ESD	Environmental sex determination
FGF9	Fibroblast growth factor 9
egfr	Epidermal growth factor gene
gsdf	Gonadal soma derived growth factor
GSD	Genetic sex determination
HMG	High mortality group
hiTAIL	High efficiency Thermal asymmetric interlaced
mya	Million year ago
RLM-RACE	RNA ligase mediated Rapid amplification of cDNA ends
sdY	Sexually dimorphic on the Y chromosome
Sry	Sex determining region Y
WGD	Whole genome duplication
Wnt4	Wingless related MMTV integration site 4

Table of Contents

ABSTRACT **7**

1. INTRODUCTION **8**

1.1 Sex determination and sex differentiation 8

1.2 Sex determination mechanism 8

1.2.1 Genetic sex determination (GSD) *9*

1.2.2 Male heterogamete *9*

1.2.3 Female heterogamete *10*

1.2.4 Allelic sex determination *10*

1.3 Environmental sex determination 11

1.4 Sex determining pathway 12

1.4.1 Sex determination in mammals *13*

1.4.2 Sex determination in fishes *14*

1.5 Genome duplication in fish 16

1.6 Sex determination in Salmonidae 18

1.7 Aim of the study 20

2. MATERIAL AND METHODS **21**

2.1 HiTAIL PCR (High Efficiency Thermal Asymmetric Interlaced PCR) 21

2.2 RNA Extraction and quality analysis 23

2.3 cDNA synthesis and 3' RLM-RACE 24

2.4 Cloning and Transformation 25

2.5 Plasmid isolation and Restriction digestion 27

2.6 Sequencing and capillary electrophoresis 27

2.6.1 Direct Sequencing of PCR products *28*

2.7 Polymerase Chain Reaction (PCR) analysis 28

2.8 High Resolution Melting curve (HRM) analysis 31

3. RESULTS 32

3.1 Polymerase chain reaction (PCR) to test presence of Exon 2 and Exon 3 in the sdY gene 33

3.2 Sequencing the sdY gene 34

3.3 hiTAIL PCR and cloning 42

3.4 Genomic DNA sequencing and assembling sequence 44

3.5 PCR based genotyping tests were developed to differentiate male and female salmonid species 48

3.5.1 Duplex PCR genotyping test and High resolution melting curve analysis 48

4. DISCUSSION 59

4.1 Primer development 59

4.2 RLM-RACE 60

4.3 Sequencing and assembling 61

4.4 Duplex PCR and HRM analysis 62

4.5 Triplex PCR and HRM analysis 63

5. CONCLUSION 65

6. REFERENCES 66

APPENDICES 71

Abstract

Most of the aquaculture industry relies on the female production stock because male rapidly leads to sexual development which causes poorer flesh quality and has increased chances to disease susceptibility. The sdY gene was recently discovered as a master sex-determining gene in Rainbow trout and the gene is considered to be conserved in most of the salmonid species. Four species of Salmonid (*Salmo trutta, Salmo salar, Salvelinus alpinus* and *Salmo marmoratus*) was selected for this study. Partial mRNA sequences of the sdY gene are available of *Salmo trutta, Salmo salar* and *Salvelinus alpinus* at NCBI. These partial mRNA sequences were used to design primers with the aim to amplify and determine the full-length sdY mRNA (cDNA) of *Salmo trutta, Salmo salar* and *Salvelinus alpinus*. hiTAIL PCR and 3' RACE was performed to obtain 5' and 3' end of the sdY gene and based on this new sequence information, primers were designed for PCR-based and HRM analysis genotyping methods to differentiate male and female samples. Duplex and triplex PCR-based genotyping were developed. Both the duplex and triplex PCR-based genotyping methods do not able to differentiate the sex of the selected species. HRM analysis using the same sets of primers as in duplex and triplex PCR-based genotyping can readily able to differentiate the sex of the selected species. The developed methods of genotyping cannot be concluded as the best method until we will not test these methods with more sex identified samples of these species.

1. Introduction

Sex determination and differentiation are an important aspect of study in the field of vertebrate reproductive biology. Sexual reproduction is a biological feature of most vertebrates. But the mechanisms of sex determinations-, i.e. the production of individuals with male or female phenotypes are diverse and affected by both genetic and environmental factors (Graves & Peichel, 2010). There are more than 2800 species of fish distributed in the large variety of habitats around the world (Nelson, 1994). In some species, an individual develops as either male or female and remains male and female for its lifetime (gonochoristic species). In other species, some individuals are a hermaphrodite, a combination of both sexes. Some hermaphroditic individuals can be either sequential or synchronous. Sequential hermaphrodites individuals develop first as one sex (male–protoandrous/female-protogynous) and later convert to the opposite sex. While in a case of synchronous hermaphrodites individuals produce and maintain functional male and female gametes simultaneously throughout their life (Devlin & Nagahama, 2002).

1.1 Sex determination and sex differentiation

Sex determination is defined as the genetic or environmental processes that determine the sexual characteristics in an organism. The sex differentiation refers to the process of gonadal development once sex has been determine, i.e. the gene determine whether the gonads follows the ovarian or testicular differentiation pathway (Hayes, 1998).

1.2 Sex determination mechanism

There are two mechanisms of sex determination recognized in vertebrate, genetic sex determination (GSD) and Environmental sex determination (ESD) mechanisms. GSD is a mechanism in which genes directly determine whether the gonads will differentiate into testes or ovaries without environmental (external) influences while in ESD mechanism sex of individuals is determined by environmental factors.

1.2.1 Genetic sex determination (GSD)

Genetic sex determination involves male heterogamete (XY) as in mammals and female heterogamete (ZW) as in bird. Male heterogametic species have chromosome X and Y, and all female have two X chromosomes. Zygotes most have X and Y, chromosome to become male or two X chromosomes to become a female. In the case of female heterogametic system, it is reversed to male heterogametic system i.e. females have Z and W chromosomes and two Z chromosomes for all males. There are other genetic sex determining mechanism, as XO system observed in the mole-vole (*Ellobius lutescens*), allele specific mechanisms identified in two fish species, namely- *Takifugu rubripes* (Kamiya et al., 2012) and *Oryzias luzonensis* (Myosho et al., 2012), or haplo-diploidy, where males arise from unfertilized eggs and females arise from fertilized eggs (Freeman, Herron, HHodin, Miner, & Sidor, 2007).

1.2.2 Male heterogamete

In male heterogametes, males will always be the heterogametic sex. This mechanism can exist in one of two forms i.e. either a female sex-determining factor on the X chromosome is required in a dosage-dependent manner where two copies are required for female gonad development or a male sex-determining factor on the Y chromosome initiates the male pathway. Currently different heterogametic male master genes have been discover like the Sry gene in mammals (Koopman, Münsterberg, Capel, Vivian, & Lovell-Badge, 1990; Sinclair et al., 1990), DMY in the Medaka fish (Matsuda et al., 2002), DMRT1 in chicken (Smith et al., 2009).

The male sex-determining factor which acts as an activator for a male pathway (Sry in mammals) or a suppressor of the female pathway or both are not conserved among vertebrates (Devlin & Nagahama, 2002; Kikuchi & Hamaguchi, 2013). For example, in Medaka (Oryzias latipes) the sex-determining master gene Dmy is not found in many closely related species that also possess a male heterogametic sex-determining system (Kondo et al., 2003; Myosho et al., 2012). In many more species of the Oryzias genus lack a known sex-determining gene even though they have an identified male or female heterogametic sex determining system (Takehana, Hamaguchi, & Sakaizumi, 2008; Takehana, Naruse, & Sakaizumi, 2005). Freshwater fish Patagonian perjerry (*Odontesthes hatcheri)* possesses a duplicated Y-linked anti-mullerian hormone gene (*amhy*) which is the first master sex-

determining gene at the hormone level to dictate phenotypic sex (Hattori et al., 2012). Till now no male heterogametic systems are known to possess a female promoting master sex-determining gene (Cutting, Chue, & Smith, 2013).

1.2.3 Female heterogamete

Birds have female heterogametic (ZZ/ZW) system in which a male promoting gene *dmrt1* does act in a dosage-dependent manner where two copies are required for the male and a single copy of *dmrt1* results in female development (Hudsonb & Sinclaira, 2007; Smith et al., 2009). The function of *dmrt1* has not been tested on other birds rather than the chicken although dmrt1 conserved presence on the Z and absence from the W across birds suggest a conserved function in birds (Chue & Smith, 2011; Smith et al., 2009). The African clawed frog (*Xenopus laevis)* which possesses Dm-W gene which acts like a female promoting master sex-determining gene (Yoshimoto et al., 2008). Above mention, both cases are opposite then the male heterogametic system observed in mammals. Females are heterozygous (ZW) and males are homozygous (ZZ).

1.2.4 Allelic sex determination

Some recent publication shows that a single base pair change in a gene can determine the sex of the species as in the tiger puffer fish (*Takifugu rubripes*) and *Oryzias luzonensis,* a closely related species to Medaka. An allele of Amhr2 in the tiger puffer fish where the presence of a G/C SNP denotes male phenotype while C/C denotes female phenotype (Kamiya et al., 2012). In *Oryzias luzonensis*, a closely related species to Medaka, Dmy is absent and instead an allele of gsdf (gonadal soma derived growth factor) controls maleness (Myosho et al., 2012).

1.3 Environmental sex determination

Several environmental factors are capable of influencing the proportion of sexes, such as temperature as seen in many reptiles (Shoemaker & Crews, 2009), social interactions are mainly observed in hermaphroditic species (Devlin & Nagahama, 2002) and the level of pH influence offspring sex.

1.3.1 Temperature

The temperature fluctuation influences fish biochemical pathway of sexual determination and act upon individual to induce male or female, as in reptile and amphibians. In many reptiles like turtles, crocodiles, tuataras and some lizards incubation temperature of the developing embryo can determine the sex. For example, in the red-eared slider turtle (Trachemys scripta) increase in incubation temperature from 26 ºC to 31 ºC promotes female development while the American alligator (*Alligator mississippiensis*) temperature range from 32.5 ºC to 33 ºC is female developing temperatures (Shoemaker & Crews, 2009).

Temperature is the factor with highest influence in the sex determination in fish (J. Baroiller, d'Cotta, & Saillant, 2008; Ospina-Alvarez & Piferrer, 2008). Male to female ratio increases with high temperatures and low temperatures have no effect or produce more females (J.-F. Baroiller, Guiguen, & Fostier, 1999; Ospina-Alvarez & Piferrer, 2008). This type of response can be observed in various freshwater and marine families such as: the Atherinids (*Menidia menidia*, *Odonthestes bonariensis*, *O. argentinensis* , *Patagonina hatcheri*); Poecilids (*Poeciliopsis lucida, Poecilia melanogaster*); Cichlids (various *Apistogramma* sp. and tilapias); Cyprinids (*Misgurnus anguillicaudatus, Carassius auratus*); Callichthyidae (*Hoplosternum littorale*), Pleuronectidae (*Verasper moseri*) (J.-F. Baroiller et al., 1999). In some rare species like channel catfish (*Ictalurus punctatus*) high temperatures tends to produce more female and low temperatures produce more male (Patiño et al., 1996). In case of the Japanese hirame (*Paralichthys oliaceus*) both high and low temperatures produce male population while the intermediate temperatures produce both male and female in 1:1 ratio (Yamamoto, 1999).

1.3.2 Social factor

The influence of social factor in sex-determination and differentiation play a significant role in hermaphroditic species. In some species sex is controlled by the level of dominance with respect to other population like, size compared to other in the population, number of individuals of males or females in the population and potential pheromones or chemical stimuli present (Devlin & Nagahama, 2002). In a case of protogynous species Red sea fish (*Anthias squamipinnis*) presence of a dominant male prevent the sex reversal of large group of females and when the male is moved to another tank one of the larger females in the group will sex-reversed to become male (Fishelson, 1970). Likewise, bidirectional inversion has been observed in Gobiidae (*Trimma okinawae*) largest female of the breeding group changes sex following removal of the dominant male, such resultant males retain both ovarian and testicular tissue and able to undergo a second sex reversal back to female if another larger male is placed in the group (Sunobe & Nakazono, 1993).

1.3.3 pH

Influence of pH on the sexual proportion has been reported in *Pelvicachromis pulcher* , 90% of males develop at pH 4-5 and 90 % of females develop in neutral pH (7.0) (Heiligenberg, 1965). A similar result has been observed for the three species of *Pelvicachromis* and for two species of *Apistogramma* , low pH (< 6.0) induce high proportion of males and neutral pH (7.0) produced a high proportion of females in the population (Rubin, 1985). However, the effect of pH is less effective than the impact of the temperature.

1.4 Sex determining pathway

The sex determining pathway means a complex series of activation of signal that lead to sex determination and differentiation. Sex determination and differentiation among vertebrate shares genetic pathway of sex determining although follow various sex determining mechanisms. Many genes responsible for common sex determining genetic pathway are known for activation of testis differentiation in all vertebrates (Graves, 2008).

1.4.1 Sex determination in mammals

Sex determination and differentiation share many conserved components among vertebrate the analysis of the sex-determining pathway in one vertebrate can discernment the pathways of many other vertebrates (Devlin & Nagahama, 2002). The gene involved in the sex-determining pathway in mammals is illustrated in figure 1.

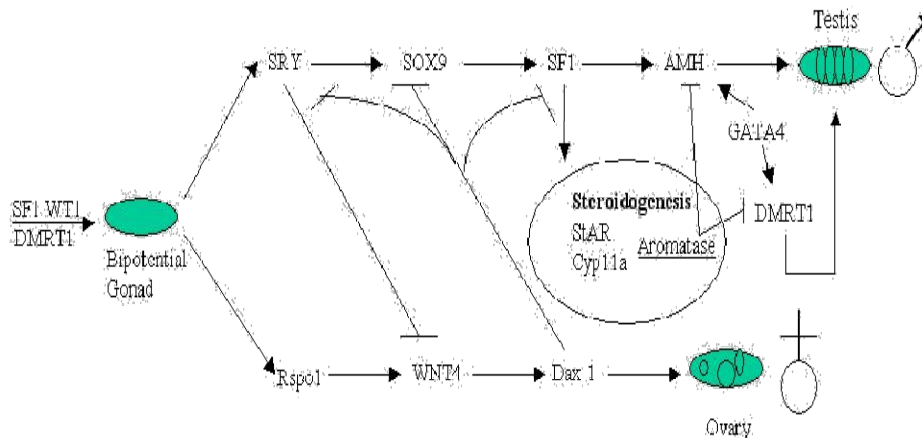


Figure 1. A simplified model highlighting some important genes involved in mammalian gonad differentiation pathway. Figure from (Li, 2010)

Gonad is initially undifferentiated and bipotential, sex of the individual is determined by the presence or absence of Sry. There are several antagonistic protein that influences the sex of individual such as wingless related MMTV integration site 4 (Wnt4) is a female promoting gene which is predominantly expressed at one end of the gonad. Fibroblast growth factor 9 (FGF9) is a male promoting gene which is predominantly expressed at the other end (Munger & Capel, 2012). These antagonistic protein influence neighbouring cells by the appropriate receptor leading to respective pathways (Bernard et al., 2011; Eggers & Sinclair, 2012). Gain and loss of experiments suggest FGF9 and WNT4 act as opposing signals to regulate sex determination (Kim et al., 2006). Some finding shows that both male and female pathways appear to be primed to go in both XX or XY embryos suggest the gonad is capable of becoming male or female even in the absence of Sry (S. A. Jameson et al., 2012). In the absence of Sry the female pathway becomes dominant after a short period of time (~1.5 days in mice) and as a result is termed the default pathway(Hiramatsu et al., 2009).

The Sry protein is a transcriptional factor containing the high mortality group (HMG) domain that bind to the a six base consensus DNA sequence will introduce a bend which leads to activation of downstream genes (Graves et al., 1995; Waters, Wallis, & Graves, 2007). Its intronless gene structure plays the important role for finding sex-determining master genes in other species. Even though the function of Sry is crucial, it is poorly conserved within the HMG box region. SOX9 is closely related to Sry, which is another essential gene that influence testis development (Graves, 2008). Genes like Dax-1 are necessary for the initiation of the female pathway in sex determination (Sinisi, Pasquali, Notaro, & Bellastella, 2002). SOX9 or DMRT1 genes that encode a protein with a zinc-finger DNA binding domain (DM domain) involve in the pathway according to the dosage dependency (J. L. Jameson, Achermann, Ozisik, & Meeks, 2003; Raymond et al., 1999). If SOX9 does not reach the pre-determined level, the supporting cell will follow female pathway (Munger & Capel, 2012). Even after the particular pathway has been chosen that path must be maintained; therefore Wnt4 and other female promoting factor must continually repress male promoting genes and likewise the reverse is for maintenance of testis in males (Matson et al., 2011; Uhlenhaut et al., 2009).

1.4.2 Sex determination in fishes

Ray-finned fishes (Actinopterygii) are the common ancestor of mammals which diverged from lobe-finned fishes approximately 450 Mya (Kumar & Hedges, 1998). There are about twenty four thousand teleost fish species worldwide; these are a monophyletic group in ray-finned fishes (Christoffels et al., 2004). Most of the fish species utilizing genetic mechanisms for sex determination have homomorphic sex chromosomes and most of the YY males are viable like in Tilapia (Kobayashi & Nagahama, 2008). YY males of many species are fertile; this indicates that the Y-chromosomes still contain functional autosomal genes. The sex chromosomes might still be in the early stage of evolution for the fish species that utilize genetic sex-determining mechanism. About only ten percentage of the cytogenetically characterized fish species have well-defined sex chromosomes (Devlin & Nagahama, 2002). Environmental factor plays a significant role in sex determination for those species which carry homomorphic sex chromosome. A hypothesis proposed by Devline and Nagahama (2002), autosomal gene could cause the replacement of the original sex chromosome once a new mutation occurred which could dominants the effect of existing sex-determining gene.

In case of Medaka, has XX/XY sex-determining system as in human. Oogenesis of Medaka early begins in females and the number of germ cells in females is much greater than in males only after 8 to 10 days post-fertilization. On the other hand, spermatogenesis begins after 5 to 6 weeks of post fertilization (Saito & Tanaka, 2008). Two methods were performed to identified the sex determining master gene in medeka, one method was to find candidate genes that are known to be involved in sex-determination and differentiation in other vertebrates and check if any candidate was located on the sex linkage group. Another method was positional cloning by chromosome walking from sex-linked markers towards the end of the sex chromosome. Both method identified Dmy (DM domain gene on Y chromosome) as the sex determining master gene in medaka. Mutation in *dmy* cause XY fish to females (Matsuda, 2003; Otake, Shinomiya, Matsuda, Hamaguchi, & Sakaizumi, 2006). Dmy is ortholog of DMRT1, is the only sex determining gene has been found in fish species. It is concluded that Dmy was transposed on to the chromosome 1 after a gene duplication and gained the new function of being sex determination switch (Graves, 2008). Many closely related medaka species lack Dmy gene which suggests that duplication of DMRT1 was recent in *Oryzias latipe*. Even some closely related species of *Oryzias latipe* like *Oryzias javanicus* and *Oryzias hubbsi* utilize ZZ/ZW system, which suggests that the new sex determination mechanism can evolve among closely related fish species (Takehana et al., 2008).

In case of Zebra fish (*Danio rerio*) no sex chromosome has been identified. Some study suggest that sex chromosome does not exist in zebrafish (Wallace & Wallace, 2003). Both sex determining system (XX/XY) has been suggested and this inconstancy of sex determining system indicate that zebrafish have unstable sex determining system (von Hofsten & Olsson, 2005). Oogenesis begins in all individuals regardless of their future sex, sex differentiated later in gonadal development. Oogenesis is completed in females while ovarian degradation initiated the testis development in males (Saito & Tanaka, 2008; von Hofsten & Olsson, 2005). High water temperature (37 ºC) or aromatase (cyp19a) inhibition can cause a sex reversal female (XX male) in Zebrafish (Uchida, Yamashita, Kitano, & Iguchi, 2004). Anti-Mullerian hormone (AMH) expressed in Sertoli cell of testis and in the follicular layer of ovaries (Rodríguez-Marí et al., 2005). *ftz-f1* gene is co-expressed with AMH both in testis and ovaries (von Hofsten & Olsson, 2005). *ftz-f1* gene regulates AMH and in turn AMH inhibits aromatase during gonadal development in mammals (di Clemente

et al., 1992). Some of the other genes like SOX9 or WT1 also play important role in sex determination of Zebra fish.

1.5 Genome duplication in fish

The family Salmonidae includes three sub-families: Coregoninae, Thymallinae and Salmoninae (Nelson, 1994). The sub-family Salmoninae includes species important for both wild fisheries and aquaculture industries like Salmon, trout and char (figure 2).

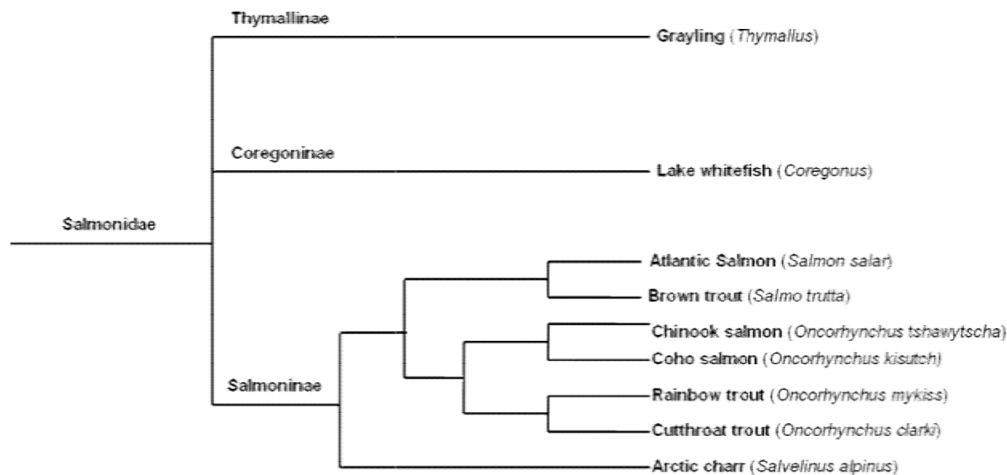


Figure 2. Phylogeny of family salmonidae, (Davidson, Huang, Fujiki, Von Schalburg, & Koop, 2009)

The duplication of genes, chromosomal segments or the entire genomes provides the raw genetic materials necessary for microevolution. These are important for generating novel gene functions and expression patterns, which possibly play significant role in evolution of metazoans, vertebrates and mammals from unicellular organisms (Ohno, 1970). There are One or two rounds of whole genome duplication (WGD) took place during the early evolution of vertebrate, this idea later known as the 2R hypothesis. The first round of WGD happen before the emergence of vertebrate and in the second round of WGD leads to the appearance of jaw vertebrates (Kasahara et al., 2007; Sidow, 1996). Third round of WGD (3R duplication) leads to give rise to the teleost lineage (Christoffels et al., 2004; Meyer & Van de Peer, 2005). The fourth WGD occurred 25 to 100 million year ago (Mya) in the

common ancestor of salmonids (Allendorf & Thorgaard, 1984). The studies of *HOX* gene clusters suggest strong evidence supporting the 2R and 3R duplication. Invertebrates have only one *HOX* gene cluster, while mammals have four gene clusters and teleosts contain seven copies of *hox* gene cluster (Amores et al., 1998; Hoegg, Boore, Kuehl, & Meyer, 2007; Naruse et al., 2000). In case of epidermal growth factor receptor gene (*egfr*), invertebrates such as *D. melanogaster* and *C. elegans* have one egfr gene while mammals have four and teleost fishes have seven egfr genes (Volff, 2005). The above observations all are in favours of 2R and 3R duplication, since only seven gene clusters are identified in teleosts rather than eight genes for both *HOX* and *egfr*. The identification of only seven cluster of genes rather than eight implies loss of one of the clusters during the teleost evolution (Hoegg et al., 2007). Atlantic salmon and rainbow trout have thirteen *HOX* gene cluster which is twice that found in Zebra fish and medaka (Moghadam, Ferguson, & Danzmann, 2005a, 2005b). Salmonids also contain 96-104 chromosome arms which is twice as many than other telosts (Mank & Avise, 2006; R. B. Phillips et al., 2009).

Genome duplication plays important role in evolutio; once a whole genome duplication has occurred the resulting gene called paralogous genes (Zhang, 2003). These paralogous genes may have three type of fate: (1) become silence or lost (nonfunctionalization), (2) acquire novel beneficial function to organism (neofunctionalization), or (3) mutation in both gene copies so that both have complement gene expression that corresponds to their single ancestral gene (subfunctionalization) (Innan & Kondrashov, 2010). After the genome duplication one gene is free to accumulate mutation since only one gene is necessary to survive and thus face one of the three fates above. Over times, the modification of new genomes through mutations, natural selection and genetic draft would produce novel gene functions which could enable speciation, i.e. known as divergent resolution. Divergent resolution means after genome duplication, one population loses function of one copy of the gene and the other population loses function of the second gene copy. This would cause the hybrid species to produce gametes that completely lack the functional genes, making the organism less viable and/or fertile. Such interspecific genomic incompatibility would mark the beginning of speciation (Lynch & Force, 2000). Divergent resolution could be one of the strong reason for huge diversity of teleost (more than 24000 species) making them the most species rich order of vertebrate.

Sex determination is one of the major problems for an organism that has undergone genome duplication (Davidson, Huang, Fujiki, Von Schalburg, & Koop, 2008). Neofunctionalization

is one way to overcome this problem, the duplicated gene undergoes neofunctionalization and become new sex determining master gene that shift the original sex determining gene, as in case of Dmy in medaka. Alternatively, one of the duplicated sex determining gene may be pseudogenized or lost from the genome in order to allow the sex determination system to return to the diploid state as it seems in rainbow trout (Yano et al., 2012). It is reasonable to expect that different lineages was evolved following different sex determining mechanisms after WGD and also assume that sdY gene was the master sex determining gene in common ancestor of salmonids.

1.6 Sex determination in Salmonidae

Salmonidiae are an important component of human food supply all over the world, due to their economic and social value some species of salmonidae, such as salmon, trout, char, grayling and fresh freshwater whitefish received attention of researcher. Salmonids are capable of changing their original sex with respect to the surrounding environment. Thus are ideal for studying the process by which sex chromosomes turnover and also for the early stage of sex chromosome evolution (Devlin & Nagahama, 2002).

Salmonids follow male heterogametic systems (Davidson et al., 2009). This has been proved via Cytogenetic studies, where some salmonid species contain heteromorphic Y chromosome, like in rainbow trout (*Oncorhynchus mykiss*), sockeye salmon (*Oncorhynchus nerka),* lake trout (*Salvelinus namaycush*) and least cisco (*Coregonus sardinella*) (Ruth Phillips & Rab, 2001). Some rainbow trout populations do not have heteromorphic sex chromosome and even some populations have different sized chromosomes, which suggest that the sex chromosomes are still in early stage of differentiation (Thorgaard, 1977).

Although salmonids have male heterogametic system, still it is not clear whether a single master sex determining gene dominates or not like Sry in mammals (Koopman et al., 1990; Sinclair et al., 1990), or each of the species of this family have their own master sex determining gene as in the genus *Oryzias* (Myosho et al., 2012) and stickleback (Urton, McCann, & Peichel, 2011). Genetic marker linked to the sex-determining locus (SEX) were design for the Artic salmon, Brown trout and Arctic char. The comparative studies using marker linked to SEX revealed that there is no sex linkage group conservation amoung these species(Woram et al., 2003), which suggest multiple gene may exist as a result of an

autotetraploidization event in the common ancestor of the salmonids (Allendorf & Thorgaard, 1984).

The recent discovery of sdY as male specific gene linked on Y chromosome in rainbow trout (Oncorhynchus mykiss) and in most of the salmonids (Salmoninae, Coregoninae and Thymallinae subfamilies) species, suggest that sdY gene may be the conserved master sex determining gene of these species, although sdY is not located at homologous genomic position among different salmonid species. Which suggest its movement associated to mobile elements, sdY gene was consider to be evolve through neofunctionalization from the interferon regulatory factor 9 (Irf9) by losing type I interferon (IFN) signalling pathway (Yano et al., 2012; Yano et al., 2013). The structure of the gene concluded from the alignment of the sdY cDNA (GenBank accession number AB626896) with Oncorhynchus mykiss Y chromosome genomic sequence (EU081756.1) shows that it was composed of four exons that align from 11034 to 18728 bp of the 21126 bp of Oncorhynchus mykiss Y chromosome genomic sequence (figure 3).

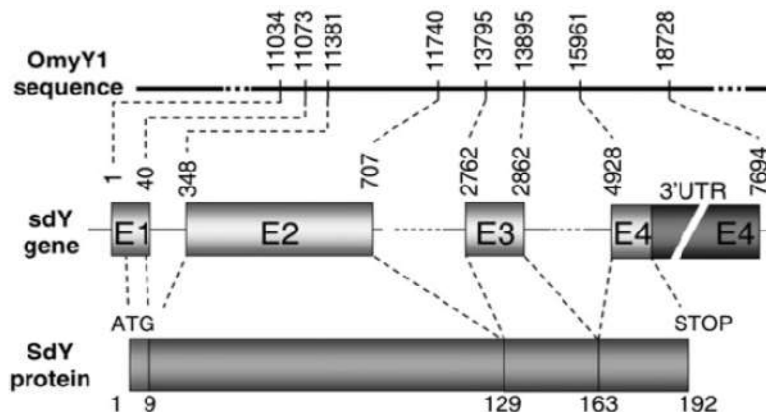


Figure 3. Number of exon and intron of sdY gene predicted in Rainbow trout

The alignment shows the position of the intron and exon boundaries and the codon number are indicated. The sdY gene encodes a putative protein of 192 amino acids and consist of four exon that span 7694 bp of Oncorhynchus mykiss Y chromosome genomic sequence. However the sdY gene was not discovered in other species of salmonids, only the partial mRNA of sdY gene is known for *Salmo salar*, *Salmo trutta* and *Salvelinus alpinus,* and these species of are very important for salmonid aquaculture industries in Norway and other part of the world.

1.7 Aim of the study

The overall aim of my M.sc thesis was to study and understand the sexually dimorphic on the Y chromosome (sdY) gene in the Salmonidae family especially in Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta*).

There were two major objectives

- To elucidate the full-length sequence of the sdY gene of Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta*) by performing hiTAIL PCR , 3' RACE and 5' RACE based on the published information of sdY mRNA sequence.
- To develop gender specific PCR test for salmonids, by designing primer targeting the conserved region of the sdY sequence.

2. Material and Methods

All the Fish samples (family: Salmonidae) used in this study were supplied by, Cryogenetics As and Professor Francesco Nonnis Marzano, University of Parma, Italy. Four species were selected for this thesis: Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*), Brown trout (*Salmo trutta*) and Italian marble trout (*Salmo marmoratus*). DNA was isolated from the fins of the samples according to the Genomic DNA 2 kit. RNA was isolated from the gonad of the samples with the target to sequence full length of male specific sdY gene for all the selected species. Based on the information of (Eisbrenner et al., 2014b) work, published complete sdY gene of (*Oncorhynchus mykiss)* Rainbow trout (Yano et al., 2012) and the published partial mRNA of Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta*) primers were design within the exon 2 and exon 3 of sdY gene. These sets of primers are test on selected species to screen the presence of sdY gene in selected species.

DNA and quality was verified using the Nanodrop ND-1000 spectrophotometer (Saveen Werner, v3.8.1).

2.1 HiTAIL PCR (High Efficiency Thermal Asymmetric Interlaced PCR)

hiTAIL-PCR is an efficient technique to amplify unknown DNA sequences adjusted with partial known sequence. The hiTAIL-PCR combines the advantages of the TAIL-cycling and suppression-PCR by blocking the amplification of non-target sequence, and the small target sequences while allowing efficient amplification of large target sequences. It uses the combination of known sequence specific primers and degenerated primers binding to unknown sequence (Liu & Chen, 2007). The hiTAIIL-PCR reaction is divided into three separate PCR cycle i.e. pre-amplification, primary and secondary amplification. Pre-amplification is performed on a genomic DNA extract, followed by primary and secondary amplification on a diluted fraction of the pre-amplification cycle. The polymerase (Hot Fire Pol® from Solis Biodyne™) is activated with a hot start step at 94 °C for 10 minutes.

Pre-amplification

Pre-amplification PCR cycle performed with degenerated primers on genomic DNA, which allows tosynthesize many amplicons. The reaction is cleaved into four reactions with homologue primer; the primers are constructed with a common leader (AC1 section) which contains a NotI cleavage site for the optional cloning afterwards. The AC1 sequence is followed by the degenerate region. The aim is to amplify plenty of fragments that contain the desired fragments. Ideally the degenerated primer should bind in the UTR region. Since the primers are degenerated, mean a mixture of primers are present. The stock concentration is higher (10 µM) than the standard 3 µM. The PCR was performed in a Veriti® 96-Well Thermal Cycler (Applied Biosystems) with the following parameters: the denaturation step at 94 ºC for 10 minutes (min), 93 ºC for 2 min and 95 ºC for 1 min, followed by 10 cycle of 94 ºC for 30 sec, 60 ºC for 1 min, 72 ºC for 3 min, followed by 1 cycle of 94 ºC for 30 sec, 50 ºC for 2 min, 72 ºC for 3 min, followed by 25 cycle of 94 ºC for 30 sec, 58 ºC for 1 min, 72 ºC for 3 min. Final extension at 72 ºC for 5 min.

Primary PCR

The amplicon obtained from the pre-amplification was diluted and with the specific primer (Liu & Chen, 2007) only the desired fragment will be amplified. The use of nested primer make certain that only the fragment of known sequence are synthesized, the degenerated primer are replaced by the primer with only AC1 common leader to make sure that no other fragment are synthesized. Since the AC1 sequence also present in the nested primer, the small fragment will be inhibited due to hairpin formation between the two AC1 oligos. The larger fragments are less inhibited than the small fragments. Cycle parameter: the denaturation step at 94 ºC for 10 min, followed by 1 cycle of 94 ºC for 20 sec, 65 ºC for 1 min, 72 ºC for 3 min, followed by 10 cycle of 94 ºC for 30 sec, 68 ºC for 1 min, 72 ºC for 3 min, 94 ºC for 30 sec, 68 ºC for 1 min, 72 ºC for 3 min 94 ºC for 30 sec, 50 ºC for 2 min, 72 ºC for 3 min and final extension at 72 ºC for 5 min.

Secondary PCR

The amplicon obtain from the primary PCR was diluted and further amplification was done to isolate the desired amplicon. Which was done in similar way with a nested primer in the known region and the common leader AC1 assuming the known sequence. Cyclic parameters used for PCR was as follow: Denaturation at 94 ºC for 10 min, followed by 7 cycle of 94 ºC for 30 sec, 68 ºC for 1 min, 72 ºC for 3 min, 72 ºC for 3 min, 94 ºC for 30 sec, 50 ºC for 2 min, 72 ºC for 3 min and the final extension at 72 ºC for 5 min.

Bands from the gel are purified by following protocol GFX™ PCR DNA and Gel Band Purification Kit from Illustra™ . The bands of interest were cut using a clean scalpel, weigh the agarose slice was weighed and transfered to an Eppendorf tube. 10 µl of capture buffer type 3 was added to each 10 mg of gel slice, mixed by inversion and incubation at 60 ºC for 30 min until the agarose gel was completely dissolved. Place the GFX micro spin column into collection tube and capture buffer type 3 by mixing briefly to collect the liquid at the bottom of the tube, up to 800 µl of capture buffer type 3 was transfer to the assembled GFX microspin Column and collection. Incubated at room temperature for 1 min followed by centrifugation at 16000 xg for 30 sec. Flow-through was discarded and repeated till the entire sample was loaded. 500 µl Wash buffer type 1 was added and centrifuged at 16000 xg for 3 sec, followed by 5 min incubation at 65 ºC. The GFX microspin column was transfer to Eppendorf tube, 30 µl ellution buffer type 4 was added and incubated for 1 min at room temperature. Finally, centrifuged at 16000 xg for 1 min for collection and measure the concentration with nanodrop.

Cloning of PCR amplicon was describe in section 2.4.

2.2 RNA Extraction and quality analysis

RNA was extracted from Gonad tissue samples of Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Broun trout (*Salmo trutta*) fish. A Small Gonad tissue sample weighing 25-50 mg was transfered to an Eppendorf tube containing three stainless steel beads. A volume of 1 ml TriZol reagent was added to sample except for Atlantic salmon where 1 ml of QIAZol was added. The tissue was disrupted and homogenized in 10 µl of Dithiothreitol (DTT) and after incubation at room temperature (RT) for 5 min, 200 µl of

Chloroform was added to the sample and vigorously shaken by hand for 15 sec. The samples were incubated for 2-3 min at RT and after centrifuging at 12000 ×g for 15 min at 4 ºC, the aqueous phase (approximately 50 % of total volume) was transferred in to a new Eppendorf tube and homogenized with 500 µl of 100% isopropanol. The mixture was incubated at RT for 10 min and centrifuge at 12000 ×g for 10 min at 4 ºC. After removing the supernatant, the pellet was washed with 1 ml of 75% ethanol and centrifuge at 7500 ×g for 5 min at 4 º C and discards the wash. The RNA pellet was air dry for 5-10 min. The RNA pellet was resuspended with 30 µl RNase-free water. RNA quality was checked using the Nanodrop ND-1000 spectrophotometer (Saveen Werner, v3.8.1).

2.3 cDNA synthesis and 3' RLM-RACE

RNA was processed for the cDNA synthesis according to FirstChoice® RLM-RACE Kit. One µg of total RNA was treated with 4 µl of dNTP mix, 2 µl of 3' RACE Adapter (table 1), 1X Reverse transcriptase buffer, 1 µl RNase Inhibitor, 1 µl M-MLV Reverse Transcriptase in a 20 µl reaction volume. The reaction was mixed gently, spins briefly and finally incubated at 42 ºC for an hour.

Due to non availability of the full-length sdY gene, a 3' RNA Ligase Mediated Rapid Amplification of cDNA Ends (3' RACE) was performed using information from published sdY sequence of Rainbow trout and the partial sdY sequence of Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta*). To amplify the 3' end of the sdY gene, an Outer 3' RACE PCR was performed. Each 50 µl of reaction mixture was set up consisting 1 µl of the reverse transcribed cDNA (as template), 1X of Phusion High-fidelity (HF) buffer, 0.2 mM of dNTPs, 1 unit of Phusion DNA polymerase, 0.5 µM of sdY exon2F (34-51) and 3' RACE Outer primer. The primers are listed in table 1. The PCR was performed in a Veriti® 96-Well Thermal Cycler (Applied Biosystem) with the following cycling parameter: 98 ºC for 28 sec for initial denaturation, followed by 35 cycle of 98 ºC for 10 sec, 60 ºC for 20 sec and 72 ºC for 30 sec per cycle with final elongation step of 72 ºC for 5 min.

Table 1. Primer and Sequence

Name	Sequence (5' - 3')
sdY exon 2F (34-51)	AGTACTGCGAAGAGGAGG
sdY exon 2R (212-230)	GCTTAAAACCACTCCACCC
sdY exon 3F (381-399)	TTCAGCAGAGCAGATGGCT
sdY exon 3R (444-462)	AGATTGGTGCACTGAGTGA
3' RACE Outer Primer	GCGAGCACAGAATTAATACGACT
3' RACE Inner Primer	CGCGGATCCGAATTAATACGACTCACTATAGG
3' RACE Adapter	GCGAGCACAGAATTAATACGACTCACTATAGGT12VN
sdY 5' UTR F	AGAAACACTCCCCTTAACTAAT
sdY 3 UTR DegR	GKYTGAWTGTCTTCTAYTTAGTGTGAG
hiTAIL_ex2R_0	ATCCCATCCATCATCAGGGAGAAG
hiTAIL_ex2R_2	TCCTCTTCGCAGTACTCCACTGAGACAAG
hiTAIL_ex4F_0	ATTAAATTGCCATGGGCTCAGCAGC

Due to the absence of a clear and specific band from the Outer 3' RACE PCR, an inner (nested) 3' RACE PCR was performed. The procedure was same as for the Outer 3' RACE PCR except for template used and extension time. Outer 3' RACE PCR product was used as template (1 µl), 0.5 mM of sdY exon3F (444-462) and 3' RACE inner primers and extension step adjusted to 72 ºC for 15 sec.

2.4 Cloning and Transformation

The blunt ended 3' end of the sdY gene generated from 3' RACE and the PCR product from the hiTAIL PCR was cloned into the pCR® Zero Blunt® (Invitrogen) vector. The vector has

T7 forward and M13 reverse primer (Appendix 1) site for flanking the insertion site in the vector. The vector have several regions necessary to survive on a selection medium for reassuming that the desired insert was present, like for this case kanamycin resistance gene. Map of vector is shown in figure 4.

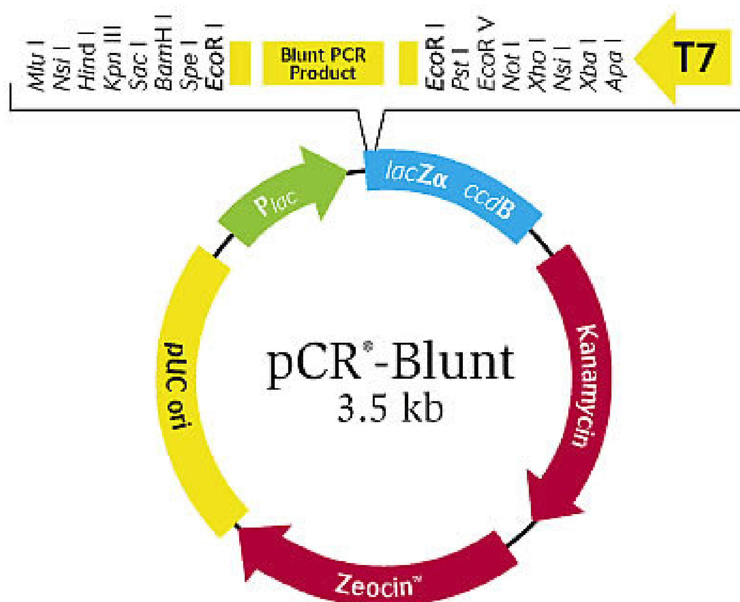


Figure 4: pCR® Zero Blunt® vector

The amplified fragment from the 3' RACE PCR was cloned using the Zero Blunt® PCR Cloning Kit (invitrogen), according to manufacturer's instructions. A total volume of 10 µl of ligation reaction consisting of 0.5 µl of PCR product, 2.5 ng of pCR® Blunt vector, 1X ExpressLink™ T4 DNA Ligase Buffer and 5 units of ExpressLink™ T4 DNA Ligase was setup. The similar setup was done for the hiTAIL PCR amplicon except for the volume of PCR amplicon (i.e. 1µl). The reaction was incubated for 1 hour at room temperature, followed by transformation.

One Shot® TOP10 competent cells (*E.coli*) was transformed with ligation product. 50 µl vial of One Shot® TOP10 cells for each transformation was thawed on ice. 2.5 µl of ligation reaction was added directly into the vial of competent cells and mixed by tapping gently. After incubation on ice for 30 minutes, the cells were incubated at 42 ºC water bath for exactly 30 seconds followed by incubation on ice for 1 minute. A volume of 250 µl of pre-warmed (37 ºC) SOC medium was added to each vial, by incubation at 37 ºC for 1 hour with shaking 225 Rotation per minute (rpm). Transformed cells were further culture at 37 ºC for

16-18 hours on LB agar plates supplemented with 50 µg/ml kanamycin. Finally the plates were wrapped in plastic bags and incubated at 37 ºC for overnight.

2.5 Plasmid isolation and Restriction digestion

The plasmid DNAs were isolated from the overnight cultured cell using PureYield™ Plasmid Miniprep System kit (Promega) according to manufacturer's instructions. 1.5 ml of *E.coli* culture was centrifuge at maximum speed for 30 sec and the cell pellet resuspended with 600 µl of water. The cells were lysed with 100 µl lysis buffer, neutralized with 350 µl neutralization solution and centrifuged for 3 minutes at maximum speed. The supernatant (≈900 µl) was transferred to PureYield™ Minicolumn without disturbing the cell debris pellet and DNA was bound to the column after centrifugation at maximum speed for 15 seconds. The membrane was washed with 200 µl of Endotoxin removal wash (ERB) and 400 µl of Column wash solution. The DNA was eluted with 30 µl of Elution buffer after centrifugation at maximum speed for 15 sec.

Restriction digestion analysis was important for the verification of generated plasmid vectors and constructs. The reaction volume of 10 µl of each reaction mixture consisting 2 µl of plasmid with 0.5 µl of enzymes and 1 µl of buffer according to the restriction map from NEW ENGLAND Biolabs. The reaction was incubated for 1 hour at optimum temperature for enzyme activity.

Generated PCR products and the restriction digest samples were analyzed by gel electrophoresis on 1% agarose in 1X TAE buffer and visualized using ethidium bromide.

2.6 Sequencing and capillary electrophoresis

Constructs were further verified by sequencing using BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). Each reaction of 10 µl volume containing approximately 20 ng of plasmid DNA, 3.2 pmol of primer i.e. T7F and M13R , 1X BigDye sequencing buffer and 0.5 µl of BigDye terminator v3.1 was setup. Cycle sequencing conditions were adjusted according to the stepped elongation time protocol (Platt, Woodhall, & George, 2007). The impurities such as primer residue, dNTP and ddNTP which are

leftover from cycle sequencing were removed by ethanol precipitation method. The extension product was mixed with 2 µl of 3M NaOAc, 2 µl of 125 mM EDTA (pH 8) and 10 µl of H_2O and after incubating for 15 min at room temperature, the reaction was centrifuged at maximum speed for 30 minutes at 4 ºC. The supernatant was carefully removed and DNA pellet was washed with 70 µl of 70 % EtOH, followed by 10 minutes centrifugation at room temperature. Supernatant was removed and the pellet dried at room temperature to get rid of excess ethanol. The purified products were dissolved in 10 µl of deionized formamide and sequenced on an automated ABI prism 3130xL Genetic Analyzer (Applied Biosystems).

The sequencing data was analyzed by the CLC Main Workbench software, and National Center for Biotechnology Information database (NCBI).

2.6.1 Direct Sequencing of PCR products

The direct sequencing of PCR products i.e. genomic DNA sequencing with the target to amplify the 5' end of the sdY gene. PCR reagents sequenced following the same procedure as described in section 2.6, except for the primers i.e. sdY 5' UTR F, sdY 3 UTR DegR, hiTAIL_ex2R_0, hiTAIL_ex2R_2, hiTAIL_ex4F_0 and sdY exon 2R (212,230) and the concentration of primers was 0.5 µM (see table 1). 1 µl of PCR product was treated with 1X sequencing buffer, 2 Units off Exo I and the mixture incubated for 1 hour at 37 ºC then for 15 minutes at 85 ºC. Sequencing reaction of total volume 10 µl contains 5 µl of Exo I treated PCR reaction, 1X sequencing buffer, 0.2 µM of primer (T7F and M13R) and 0.5 µl of BigDye terminator v3.1 was setup and the cycle sequencing and product purification was performed as described in section 2.6. were adjusted according to the steppes elongation time protocol(STeP) (Platt et al., 2007).

2.7 Polymerase Chain Reaction (PCR) analysis

Two seperate uniplex PCR reaction was setup to screen the presence of exon 2 and exon 3 of sdY gene in different species of Salmonidae family i.e. Atlantic salmon (*Salmo salar*), Arctic char (*Salvelinus alpinus*) and Brown trout (*Salmo trutta*). The primers for the PCR amplification are listed in table 1. First uniplex was perform for exon 2, each 15 µl reaction mixture consisting of 11.5 µl PCR Grade water, 15 ng template (DNA), 1X of Hot FIREpol Buffer B2, 1.5 mM of $MgCl_2$, 0.1 mM of dNTPs, 0.9 units of Hot Firepol DNA polymerase, 0.25 µM of exon 2F (34-51) and exon 2R (212-230) primers. Second uniplex for exon 3, 15

µl reaction mixtures consisting of same reagents as for the exon 2 except for the primers and concentration i.e. 1 µM of exon 3F (381-399) and exon 3R (444-462) primers and 9.1 µl of PCR grade water was setup. The reaction mixture was mixed briefly. The PCR was performed in a Veriti® 96-Well Thermal Cycler (Applied Biosystems) with the following parameters: an initial denaturation step at 95 º C for 10 minutes (min) to activate the polymerase, followed by 35 cycles of 95 º C for 30 seconds (sec), 60 ºC for 30 sec, 68 º C for 30sec, followed by a final extension at 68 º C for 10 min. The PCR product was separated on 1 % 1X TBE agarose gel and visualized using ethidium bromide.

Two sets of PCR reaction were successfully developed to differentiate the sex of the selected species of Salmonidae A. Duplex PCR reaction and B. Triplex PCR reaction.

A. Duplex PCR reaction

Each 15 µl of duplex PCR reaction mixture was setup containing 9.9 µl of PCR grade water, 1X of Hot FIREpol buffer B2, 1.5 mM of $MgCl_2$, 0.2 mM of dNTPs, 0.9 units of HOT FIREpol DNA polymerase, 15 ng template (DNA), 0.4 µM of forward and reverse Fabp6b primer and mixture 0.4 µM of hiTAIL_ex3F_2 forward primer and sdY 3UTR DegR reverse primer (table 2). The PCR was performed in a Veriti® 96-Well Thermal Cycler (Applied Biosystems) with the following cycling parameters: an initial denaturation step at 95 ºC for 10 minutes to activate the polymerase, followed by 35 cycles of 95 º C for 30 seconds (sec), 60 º C for 30 sec, 72 º C for 30 sec and a final extension at 72 º C for 10 min.

Table 2. Primer and sequence

Name	Sequence (5'- 3')
Fabp6b F	AATTACGATGAGTTTCTGGAGGCAA
Fabp6b R	CTTTCCGATGGTGAATTTGTTAGTCAA
hiTAIL_ex3F_2	TCAGCAGCTATTCAAGCAAGCTCACGAC
sdY 3 UTR Deg R	GKYTGAWTGTCTTCTAYTTAGTGTGAG
sdY exon 3F(381-399)	TTCAGCAGAGCAGATGGCT
sdY exon 3R (444-462)	AGATTGGTGCACTGAGTGA
sdY ex4 F2	ATTCAAGCAAGCTCACGAC
sdY 3'UTR R2	AAAAGAGGTAAAAAGGTGGAGG

B. Triplex PCR reaction

Each triplex PCR reaction of 15 µl volume containing,15 ng of template (DNA), 0.2 µM of sdY exon 3F (381-399) and sdY exon 3R (444-462) , 0.2 µM of sdY ex4F2 and sdY 3' UTR R2, 0.6 µM of forward and reverse Fabp6b primers and 1X Hot Firepol EvaGreen qPCR Super mix. The cycling parameters were same as the duplex PCR except for annealing temperature was adjusted to 62 ºC.

2.8 High Resolution Melting curve (HRM) analysis

High resolution melting curve (HRM) analysis was perform with a target to see the difference between the melting curve of male and female samples. HRM was performed on two qPCR machines, the Rotor-Gene Q (QIAGEN) with Rotor-Gene Q series software and on the 7500 Fast Real-time PCR system (Applied BioSystem). Each duplex PCR reaction of 15 µl volume containing , 15 ng of template(DNA), 1X Hot Firepol EvaGreen qPCR Super mix, 0.4 µM of Forward and reverse of Fabp6P, 0.4 µM hiTAIL_ex3F_2 and sdY 3 UTR DegR. DNA was amplified using the same cycling parameters as describe in section 2.7 A. Melt Ramp was used from 40 to 95 ºC with 0.2 º C increments. Triplex PCR reaction contain 0.6 µM of forward and reverse fabp6b, 0.2 µM of sdY exon 3F (381-399) and sdY exon 3R (444-462), 0.2 µM of sdY ex4F2 and sdY 3'UTR R2 sets of primers, 1X HOT FIREpol EvaGreen qPCR Super mix and template. The cycle parameter was same as describe in section 2.8A and the result was analyzed using HRM v2.0.1 software.

3. Results

Sequencing of cDNA libraries lead to the description of the complete *Oncorhynchus mykiss* (rainbow trout) sexually dimorphic on the Y chromosome (sdY) mRNA (GenBank accession number: AB626896.1) (Yano et al., 2012). The sdY gene (3271 bp) has a total of 4 exons- exon 1 is 40 bp, exon 2 is 360 bp, exon 3 is 101 bp and exon 4 is 90 bp based on the alignment of sdY genomic sequence (GenBank: EU081756.1) and complete sdY mRNA sequence (alignment not shown).The partial mRNA sequences from S*almo salar* (GenBank accession number: JF826020.1)*, Salmo trutta* (GenBank accession number: JF826019.1) and *Salvelinus alpinus* (GenBank accession number: JF826022.1) are available at NCBI. The alignments of these partial mRNAs with the complete Rainbow trout genomic and mRNA of sdY sequences revealed that these partial mRNA contain the complete exon 2 and exon 3 sequences (alignment appendix 2). The length of partial mRNA of sdY sequences of all three species are exactly of same length (515 bp) and the alignment show mismatches at three positions 16, 53 and 184 bp, where C mismatch with G, G mismatch with A and C mismatch with T respectively. Primers were design to avoid these mismatches, based on the alignment of the partial mRNA two set of primers: sdY exon 2F (34-51) and sdY exon 2R (212-230), and sdY exon 3F (381-399) and sdY exon 3R (444-463), annealing within exon 2 and within exon 3 respectively were design. The alignment of the partial mRNA sequences and the position of the primers are shown in figure 5. sdY exon 2F (34-51) at position 34 to 51 bp and sdY exon 2R (212-230) at 212 to 230 bp, sdY exon 3F (381-399) at position 381 to 399 bp and sdY exon 3R (444-463) at 444 to 463bp. The sequences of the primers are given in table 1.

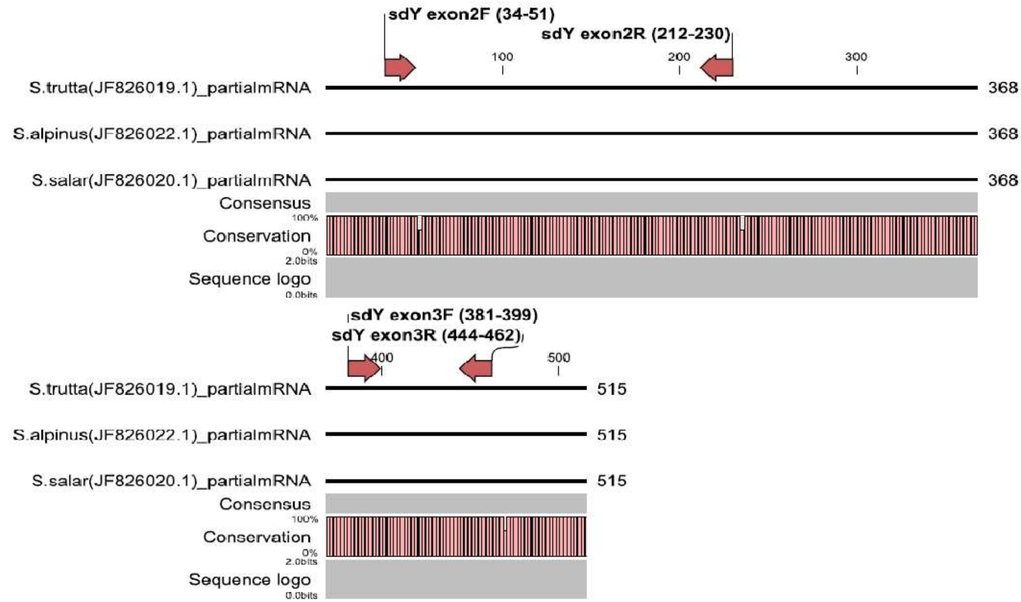


Figure 5. Diagram represents alignment of partial mRNA of sdY sequence of S*almo salar* (JF826020.1)*, Salmo trutta* (JF826019.1) and *Salvelinus alpinus* (JF826022.1) and the position of primers.

3.1 Polymerase chain reaction (PCR) to test presence of Exon 2 and Exon 3 in the sdY gene

Four different species of Salmonids were selected for this study: *Salmo trutta* (Brown trout), *Salar salmo* (Atlantic salmon), *Salvelinus alpinus* (Arctic char) and *Salmo marmoratus* (Italian marble trout). The two sets of primers: sdY exon 2F (34-51), sdY exon 2R (212-230) and sdY exon 3F (381-399), sdY exon 3R (444-462) were used for PCR amplification and to screen for the presence of exon 2 and exon 3 in the different species of genomic (DNA) sample as describe in 2.7. As shown in figure 6 panel A, exon 2 primer yielded a PCR product of approximately 200 bp size which is consistent with the expected size of 197 bp and the panel B for exon 3 primer produced amplicon of approximately 100 bp of size which is consistent with the expected size of 82 bp.

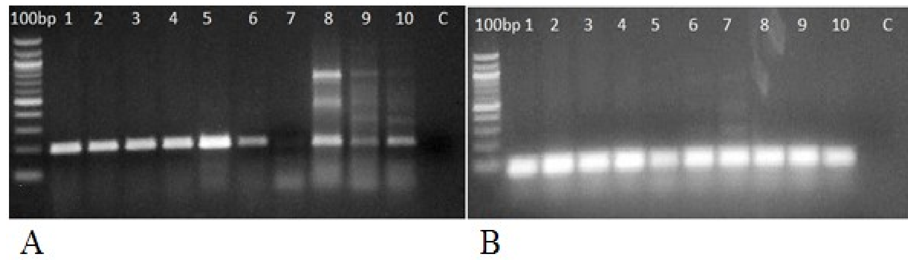


Figure 6. Gel photo showing observed bands generated from uniplex PCR of sdYgene (A and B). A. Uniplex PCR with primers sdY exon2 F (34-51) and sdY exon 2R (212-230), male *Salmo trutta* (A1-A5), *Salmo salar* (A6-A10) and C: negative control. B. Uniplex PCR with primers sdY exon 3F (381-399) and sdY exon 3R (444-463), male *Salmo trutta* (B1-B5), *Salmo salar* (B6-10) and C: negative control.

Since confirmed sex identified samples of *Salvelinus alpinus* was not available at this point of the project, so unable to test these sets of primers on *Salvelinus alpinus.* Even for the *Salmo salar* we did not have conformed sex identified samples at this point of project so these sets of primers were tested on random samples of *Salmo salar.* As able to amplify the sdY gene amplicon with these sets of primer, it is confirmed the presence of sdY gene in these species and decided to carry further work to sequence the sdY gene.

3.2 Sequencing the sdY gene

One aim of the project was to perform 5' RACE and 3' RACE in order to describe the full-length sequences of the *Salmo trutt, Salmo salar* and *Salvelinus alpinus* sdY mRNA (cDNA) gene based on the published partial mRNA of sdY sequences for these three species. 3' RLM- RACE was performed to obtain the 3' end of the sdY mRNAs as described in section 2.3 and 5' RACE was carried out to obtain the 5' end of sdY mRNA with the aim of obtaining the full length sdY mRNA (cDNA) sequences from these species. Unfortunately unable to PCR amplify any 5' RACE product from any species (results are not shown).

RNA was extracted from the gonad tissue (section 2.2) and reverse transcription was performed following the RLM-RACE kit. cDNA obtained from the reverse transcription

was used for the outer and the inner (nested) PCR reaction using Thermo Scientific Phusion High-Fidelity DNA polymerase protocol (section 2.3). The inner PCR amplicon shows distinct band at approximately 500 bp for the *Salmo trutta* and *Salvelinus alpinus* but for the *Salmo salar* no clear band was observed (figure 7).

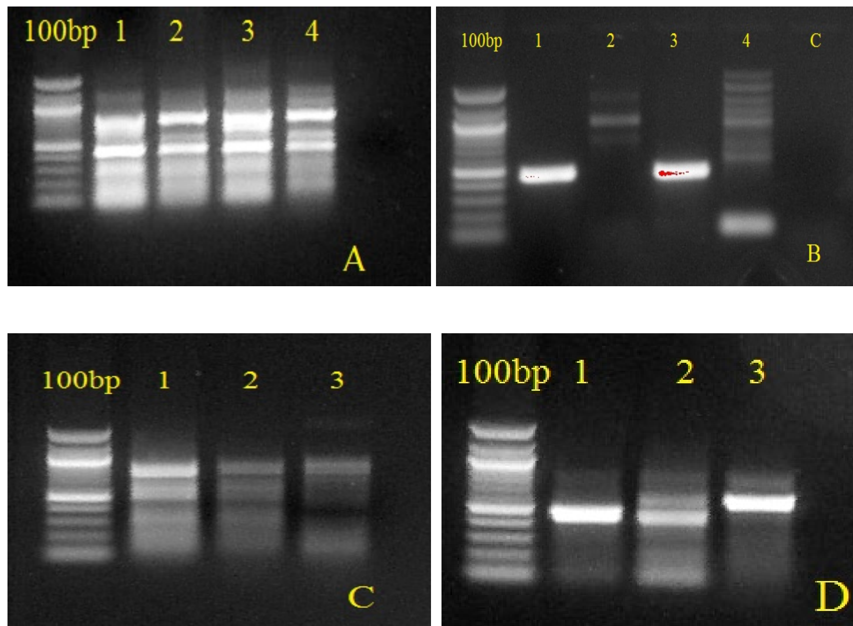


Figure 7. 1 % agrose gel was used to visualise the amplification of cDNA amplicon in different species. A. Outer PCR of *Salmo tuttar* and *Salmo salar,* lane A1: concentrated cDNA, lane A3: 10X diluted cDNA of *Salmo trutta*, lane A2: concentrated cDNA*,* lane A4: 10X diluted cDNA of *Salmo salar.* B. Inner PCR (1000X diluted outer PCR ampicon used) of *Salmo trutta* and *Salmo salar,* lane B1: concentrated cDNA*,* lane B3 10X diluted cDNA of *salmo trutta*, lane B2: concentrated cDNA, lane B4: 10X diluted cDNA *Salmo salar.* C. Outer PCR of *Salvelinus alpinus,* lane C1: concentrated cDNA, lane C2: 10X diluted cDNA, lane C3: 1000X diluted cDNA of *Salvelinus alpinus*. D. Inner PCR, lane D1: concentrated cDNA, lane D2: 10X cDNA and lane D3: 1000X cDNA of *Salvelinus alpinus*.

The 3' RACE products from the nested PCR using cDNA prepared from *Salmo trutta* (figure 7,lane B1) and *Salvelinus alpinus* (figure 7, lane D1) was cloned into the pCR® Zero Blunt® vector and transformed into the TOP 10 *E. coli* competent cells, these cells were cultured overnight in LB medium supplemented with 50 µg/ml kanamycin (section 2.4). As shown in lane B2 and B4, figure, (*Salmo salar)* did not obtain a clear single specific PCR product. We

decided to perform gradient PCR of nested PCR product (lane B2 and B4) with annealing temperatures varying from 60 to 68 ºC (figure 8).

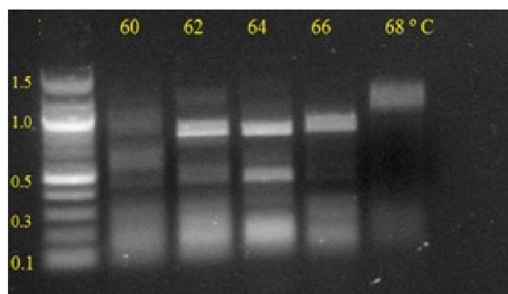


Figure 8. 1% low melting agrose gel was used to visualize the gradient PCR amplicon.

Two clear and separated bands at 64 ºC were observed and these two fragments (one is approximately 1 kb and other is 500 bp) were cloned into the pCR® Zero Blunt® vector and transformed to the TOP 10 *E. coli* competent cells.

Restriction digestion was performed on the plasmid isolated from the overnight cultured cell (isolation of plasmid 2.5), using specific restriction enzymes (EcoRI) in order to verify insertion of the cloned product and for the further sequencing (figure 9).

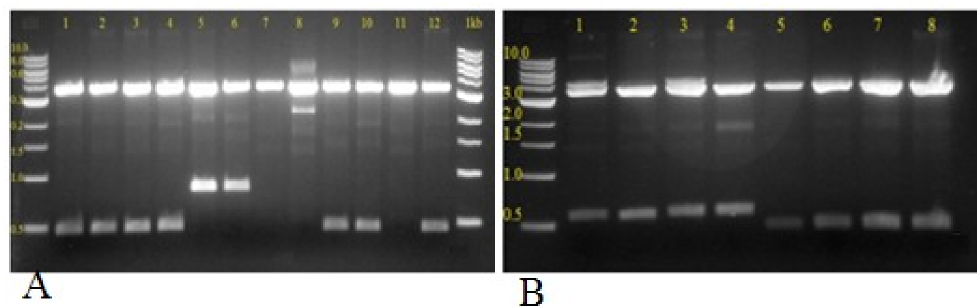
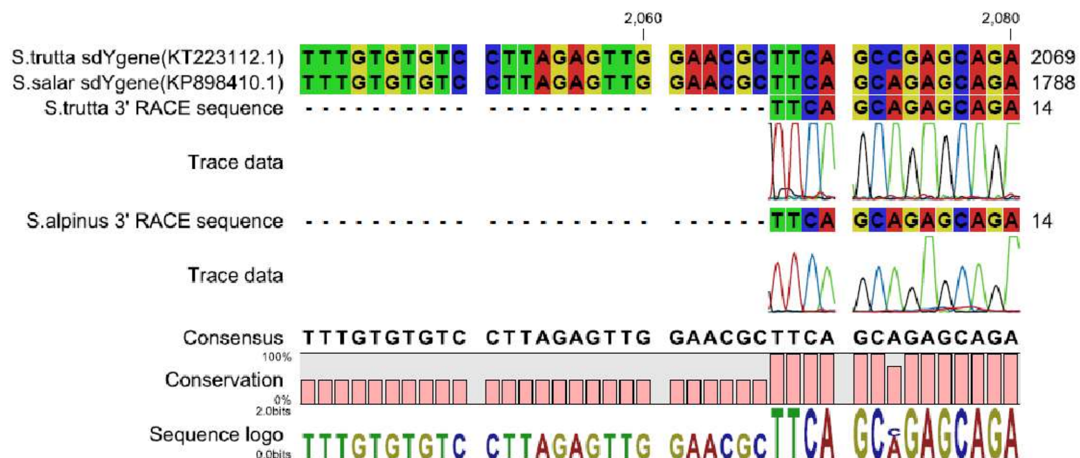


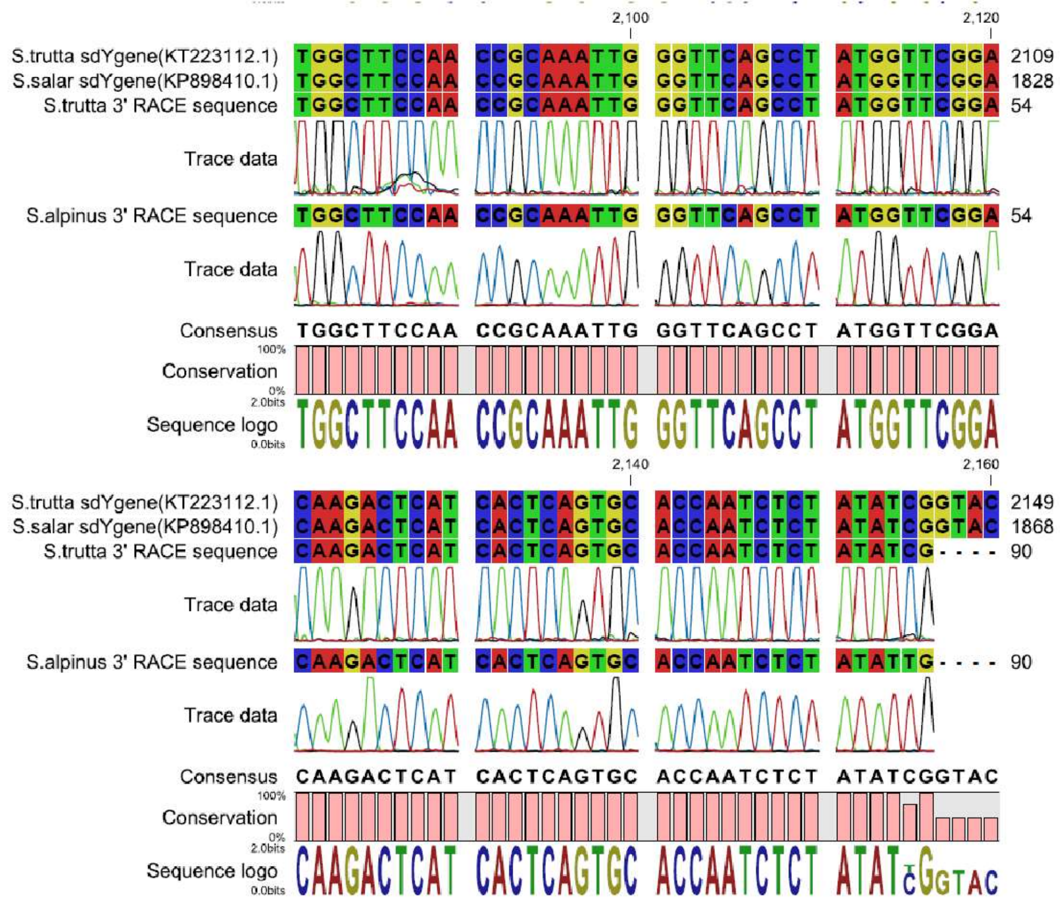
Figure 9. Gel photos showing observed bands generated for cloned 3' RACE product (cDNA gene) from restriction enzyme EcoRI A. *Salmo trutta* (Lane A1-A4) and *Salmo salar* (A5- A12), B. *Salvelinus alpinus* (B1- B8).

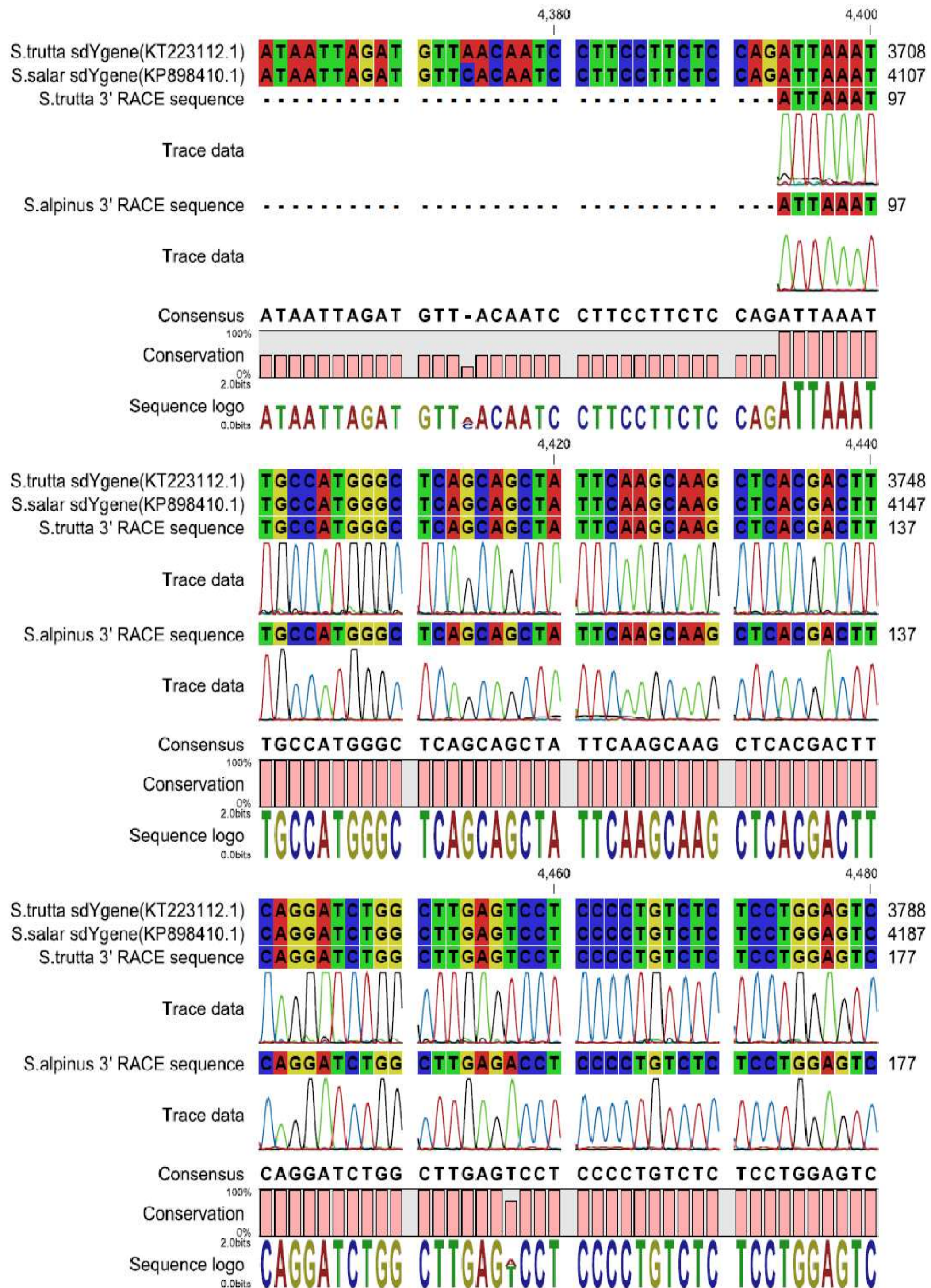
As shown in figure 9, digested with the EcoRI restriction enzymes yield restriction fragment whose size were consistent with those of the expected size of nested PCR product (≈ 500 bp)

for the *Salmo trutta* and restriction fragment for the *Salvelinus alpinus* were also consistent with the size of nested PCR product (i.e.≈ 500 bp and ≈ 600 bp) as seen in figure 7 (line D1 and D3). Size of restriction fragment for *Salmo salar* were also consistent with the size of the gradient PCR as shown in lane A5, A6 is approximately 1 kb and lane A9,A10 and A12 approximately 500 bp. . Lane A1, A2, A5, A6, A9, A10, B2 and B7 were used for the sequencing. These selected plasmid samples were used for further sequencing and analysed. The sequencing results (Appendix 3) showed that the 3' end *sdY* mRNA (cDNA) of *Salmo trutta* and *Salvelinus alpinus* was successfully cloned. The sequencing results for the *Salmo salar* were not the gene of interest sdYmRNA (cDNA) it was PREDICTED: *Salmo salar* RAD54-like 2 (*S. cerevisiae*) (rad54l2), transcript variant X3, mRNA for 500 bp fragment and PREDICTED: *Salmo salar* tubulin polyglutamylase TTLL11-like (LOC106562979), mRNA for the 1 kb fragment.

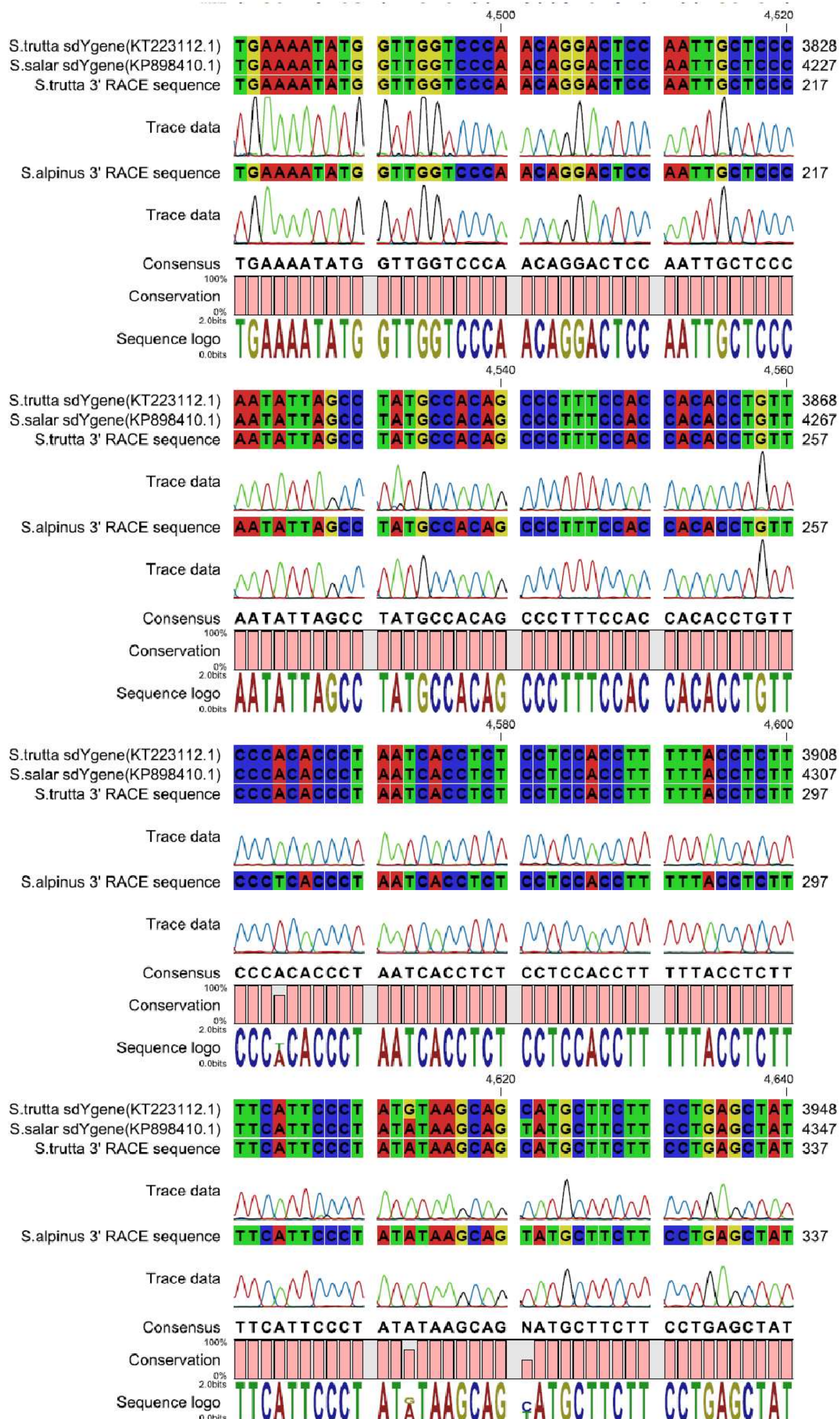
This summer (July-2015) Fraser University of Canada published the complete genomic sdY sequence of *Salmo trutta* (GenBank accession number: KT223112.1) and *Salmo salar* (GenBank accession number: KP898410.1). Based on these published genomic sequence and 3' RACE product of samples further analysis was performed. Alignment of 3' RACE mRNA (cDNA) sequences with the published complete genomic sequence of *Salmo trutta* and *Salmo salar* reveals that 3' RACE sequence (cDNA) is a partial exon 3 (only 1st 10 bp is missing out of 100 bp of exon 3), complete exon 4 and longer 3' UTR of *sdY* gene for both *Salmo trutta* and *Salvelinus alpinus* (alignment figure 10).




2,100
2,120
S.trutta sdYgene(KT223112.1)
S.salar sdYgene(KP898410.1)
S.trutta 3' RACE sequence
TGGCTTCCAA CCGCAAATTG GGTTCAGCCT ATGGTTCGGA
2109
1828
54
Trace data
S.alpinus 3' RACE sequence
54
Trace data
Consensus
TGGCTTCCAA CCGCAAATTG GGTTCAGCCT ATGGTTCGGA
Conservation
Sequence logo
2,140
2,160
S.trutta sdYgene(KT223112.1)
S.salar sdYgene(KP898410.1)
S.trutta 3' RACE sequence
CAAGACTCAT CACTCAGTGC ACCAATCTCT ATATCGGTAC
ATATCG----
2149
1868
90
Trace data
S.alpinus 3' RACE sequence
ATATTG----
90
Trace data
Consensus
CAAGACTCAT CACTCAGTGC ACCAATCTCT ATATCGGTAC
Conservation
Sequence logo



4,380
4,400
S.trutta sdYgene(KT223112.1) ATAATTAGAT GTTAACAATC CTTCCTTCTC CAGATTAAAT 3708
S.salar sdYgene(KP898410.1) ATAATTAGAT GTTCACAATC CTTCCTTCTC CAGATTAAAT 4107
S.trutta 3' RACE sequence ATTAAAT 97
Trace data
S.alpinus 3' RACE sequence ATTAAAT 97
Trace data
Consensus ATAATTAGAT GTT-ACAATC CTTCCTTCTC CAGATTAAAT
Conservation
Sequence logo
4,420
4,440
S.trutta sdYgene(KT223112.1) TGCCATGGGC TCAGCAGCTA TTCAAGCAAG CTCACGACTT 3748
S.salar sdYgene(KP898410.1) TGCCATGGGC TCAGCAGCTA TTCAAGCAAG CTCACGACTT 4147
S.trutta 3' RACE sequence TGCCATGGGC TCAGCAGCTA TTCAAGCAAG CTCACGACTT 137
Trace data
S.alpinus 3' RACE sequence TGCCATGGGC TCAGCAGCTA TTCAAGCAAG CTCACGACTT 137
Trace data
Consensus TGCCATGGGC TCAGCAGCTA TTCAAGCAAG CTCACGACTT
Conservation
Sequence logo
4,460
4,480
S.trutta sdYgene(KT223112.1) CAGGATCTGG CTTGAGTCCT CCCCTGTCTC TCCTGGAGTC 3788
S.salar sdYgene(KP898410.1) CAGGATCTGG CTTGAGTCCT CCCCTGTCTC TCCTGGAGTC 4187
S.trutta 3' RACE sequence CAGGATCTGG CTTGAGTCCT CCCCTGTCTC TCCTGGAGTC 177
Trace data
S.alpinus 3' RACE sequence CAGGATCTGG CTTGAGACCT CCCCTGTCTC TCCTGGAGTC 177
Trace data
Consensus CAGGATCTGG CTTGAGTCCT CCCCTGTCTC TCCTGGAGTC
Conservation
Sequence logo



4,500
4,520
S.trutta sdYgene(KT223112.1) TGAAAATATG GTTGGTCCCA ACAGGACTCC AATTGCTCCC 3828
S.salar sdYgene(KP898410.1) TGAAAATATG GTTGGTCCCA ACAGGACTCC AATTGCTCCC 4227
S.trutta 3' RACE sequence TGAAAATATG GTTGGTCCCA ACAGGACTCC AATTGCTCCC 217
Trace data
S.alpinus 3' RACE sequence TGAAAATATG GTTGGTCCCA ACAGGACTCC AATTGCTCCC 217
Trace data
Consensus TGAAAATATG GTTGGTCCCA ACAGGACTCC AATTGCTCCC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
4,540
4,560
S.trutta sdYgene(KT223112.1) AATATTAGCC TATGCCACAG CCCTTTCCAC CACACCTGTT 3868
S.salar sdYgene(KP898410.1) AATATTAGCC TATGCCACAG CCCTTTCCAC CACACCTGTT 4267
S.trutta 3' RACE sequence AATATTAGCC TATGCCACAG CCCTTTCCAC CACACCTGTT 257
Trace data
S.alpinus 3' RACE sequence AATATTAGCC TATGCCACAG CCCTTTCCAC CACACCTGTT 257
Trace data
Consensus AATATTAGCC TATGCCACAG CCCTTTCCAC CACACCTGTT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
4,580
4,600
S.trutta sdYgene(KT223112.1) CCCACACCCT AATCACCTCT CCTCCACCTT TTTACCTCTT 3908
S.salar sdYgene(KP898410.1) CCCACACCCT AATCACCTCT CCTCCACCTT TTTACCTCTT 4307
S.trutta 3' RACE sequence CCCACACCCT AATCACCTCT CCTCCACCTT TTTACCTCTT 297
Trace data
S.alpinus 3' RACE sequence CCCTCACCCT AATCACCTCT CCTCCACCTT TTTACCTCTT 297
Trace data
Consensus CCCACACCCT AATCACCTCT CCTCCACCTT TTTACCTCTT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
4,620
4,640
S.trutta sdYgene(KT223112.1) TTCATTCCCT ATGTAAGCAG CATGCTTCTT CCTGAGCTAT 3948
S.salar sdYgene(KP898410.1) TTCATTCCCT ATATAAGCAG TATGCTTCTT CCTGAGCTAT 4347
S.trutta 3' RACE sequence TTCATTCCCT ATATAAGCAG CATGCTTCTT CCTGAGCTAT 337
Trace data
S.alpinus 3' RACE sequence TTCATTCCCT ATATAAGCAG TATGCTTCTT CCTGAGCTAT 337
Trace data
Consensus TTCATTCCCT ATATAAGCAG NATGCTTCTT CCTGAGCTAT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

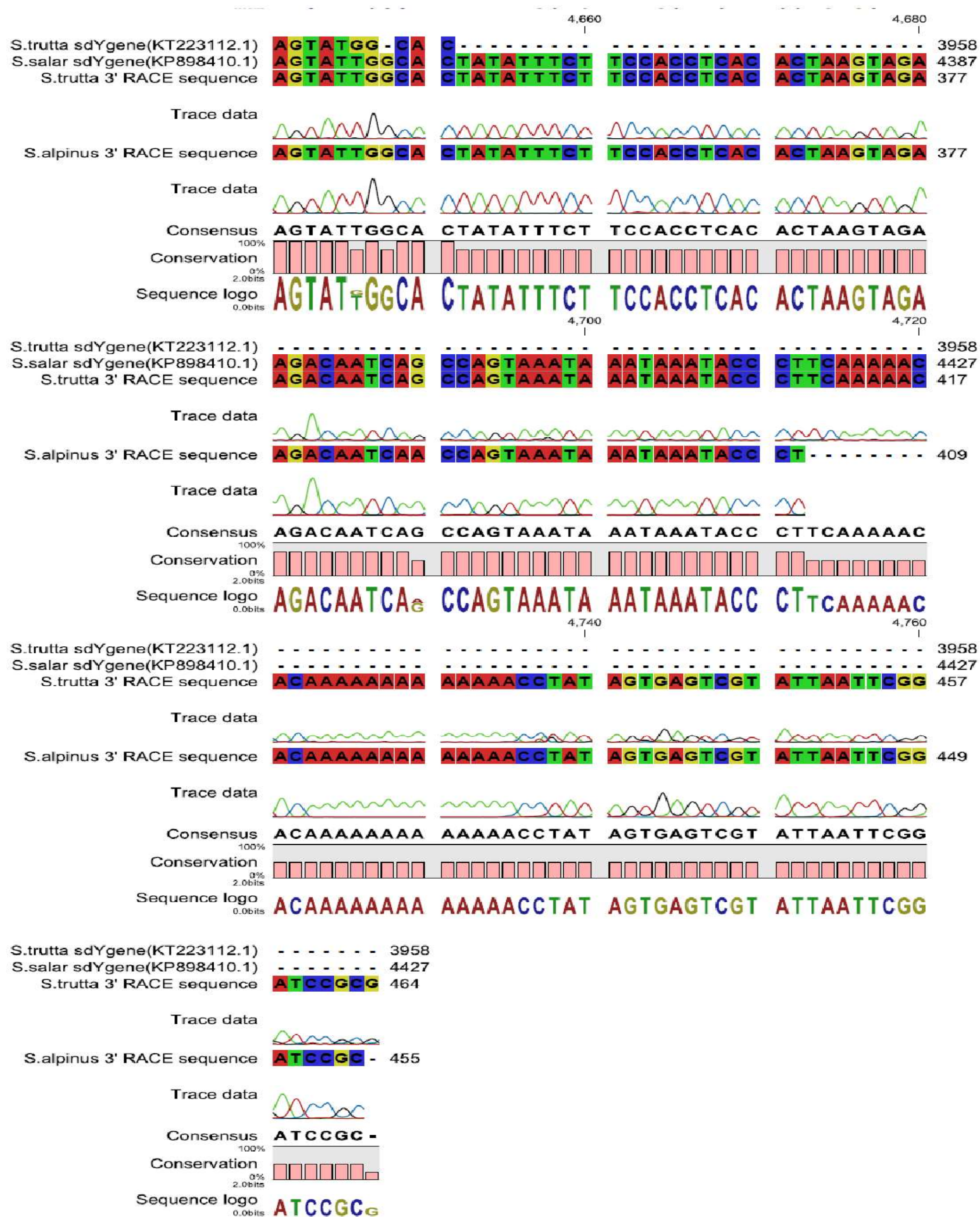


Figure 10. Alignment of published genomic sequence of *Salmo trutta* (KT223112.1), *Salmo salar* (KP898410.1) and 3' RACE sequence of *Salmo trutta, Salvelinus alpinus.*

Based on the published complete genomic sequence of *Salmo trutta* sdY gene, exon 3 is 100 bp long starts from 2045 to 2145 bp, exon 4 is 256 bp long starts from 3702-3958 bp and for *Salmo salar* sdY gene exon 3 is 100 bp long starts from 1764 to 1864 bp, exon 4 is 328 bp long starts from 4101-4429 bp.

The alignment of *Salmo trutta* and *Salmo salar* published genomic sequences with the 3' RACE cDNA sequences of *Salmo trutta* and *Salvelinus alpinus* as shown in figure 10 above reveal that we were able to obtain the partial exon 3 except only first 10 bp was missing out of 100 bp. Alignment for exon 4 start from the exact position as describe in the published genomic sequence of *Salmo trutta* and *Salmo salar* with the 3 ' RACE cDNA of *Salmo trutta* and *Salvelinus alpinus* and it align till the last base pair of the published sequences with just a manual editing at position of 4647 bp where we but a gap. This alignement also reveal that we have longer 3' UTR end sequence than the published sequences in GenBank for *Salmo trutta* and even for the *Salvelinus alpinus* which have only partial mRNA are available in data bank. Successfully able to amplify 115 bp longer 3' UTR end for the *Salmo trutta* and *Salvelinus alpinus* succesfully able to generate exon 4 which is 106 bp longer with compare to published genomic sequence of *Salmo trutta.*

3.3 hiTAIL PCR and cloning

Genomic DNA was used for the hiTAIL PCR (Liu & Chen, 2007) with target to obtain the 5' end of the sdY gene. The result of the secondary PCR is shown in figure 11. The amplicons for *Salmo trutta* and *Salmo salar* was observed after the secondary PCR, for *Salvelinus alpinus* no amplicon was observed.

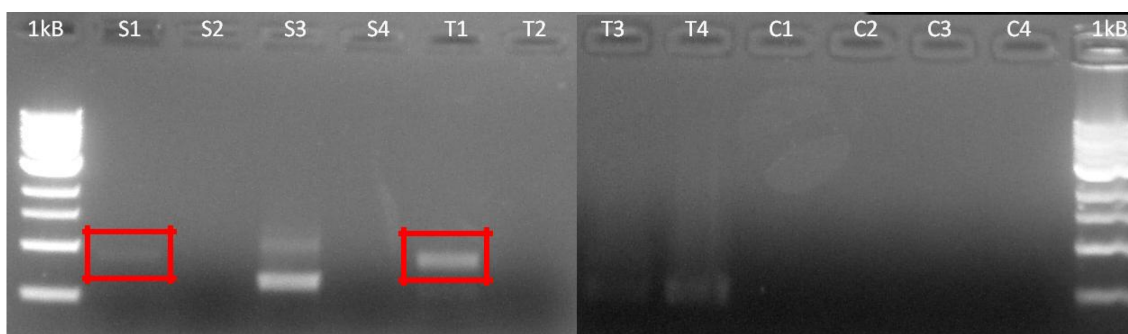


Figure 11. Agarose gel was used to visualise the amplicon of Secondary PCR of the 5' end *of Salmo salar, Salmo trutta* and *Salvelinus alpinus* of hiTAIL PCR of sdY gene. S1-S4: Atlantic salmon with AD1-AD4 primer, T1-T4: *Salmo trutta* with AD1-AD4 primer, C1-C4: *Salvelinus alpinus* with AD1-AD4 primer, S1 and T1 contain amplicon (≈ 900bp); S3 contains two amplicon (≈ 1 kb and 600 bp). T3 and T4 contain amplicon (≈ 500 bp).

From the figure 11, primer AD 1(lane S1) show specific band for the *Salmo salar* and Salmo trutta that it contain an amplicon of ≈ 900 bp. For *Salvelinus alpinus* unable to get any amplicon from secondary PCR. In order to increase the amount of the amplicons, the PCR product from the secondary PCR was used as template in a new PCR using the same primers (figure 12)

As the signal is not as optimal as desired an addition PCR was performed with the fast working polymerase (Phusion^{TM}) which help to obtain more and clear amplicons (figure 12). The largest band was selected as the longest sequence was desired to obtain. These longest bands were cut out and prepared for the sequencing.

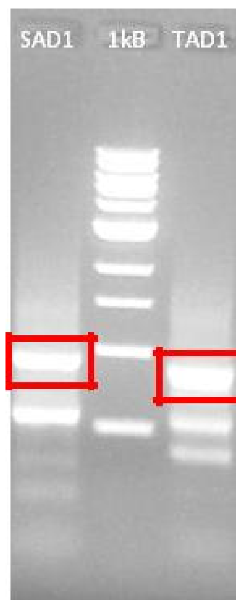


Figure 12. Gel picture picture of addition PCR with AD1 primer and using secondary PCR product as template of *Salmo salar* and *Salmo trutta*

The *E. coli* cells were transformed with the selected amplicon that was cloned in the pCR® Zero Blunt® vector (figure 13). The insert was cut out using the two cleavage sit of EcoRI, giving rise to two fragment of approximately 900 bp which is insert and other was vector (3.5 kb).

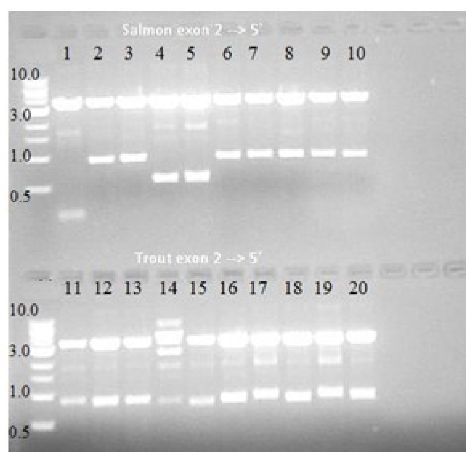


Figure 13. Gel picture showing cloned PCR product from 5' UTR- exon 2 (hiTAIL PCR).

Most of the insert observed on gel was of expected size of approximately 900 bp expect for the lane 1 which is less than 500 bp and for the lane 4 and 5 where we observed approximately of 600 bp of insert. Only the insert of expected size (≈ 900 bp) were used for sequencing.

3.4 Genomic DNA sequencing and assembling sequence

Genomic DNA sequencing of sdY gene (describe in section 2.6.1) was performed with the target to get the 5' end, 3' end of sdY gene, based on the information of 3' RACE product and partial mRNA published at NCBI different sets of primers were design with personal communication with my supervisor (Robert wilson). Analysis and assembling genomic DNA sequence, published partial mRNA and 3' RACE sequence of sdY with target to obtain sdY gene which consists 5' UTR, exon 1, intron 1. exon 2, exon 3, exon 4 and 3' UTR as demonstrated in figure 14.

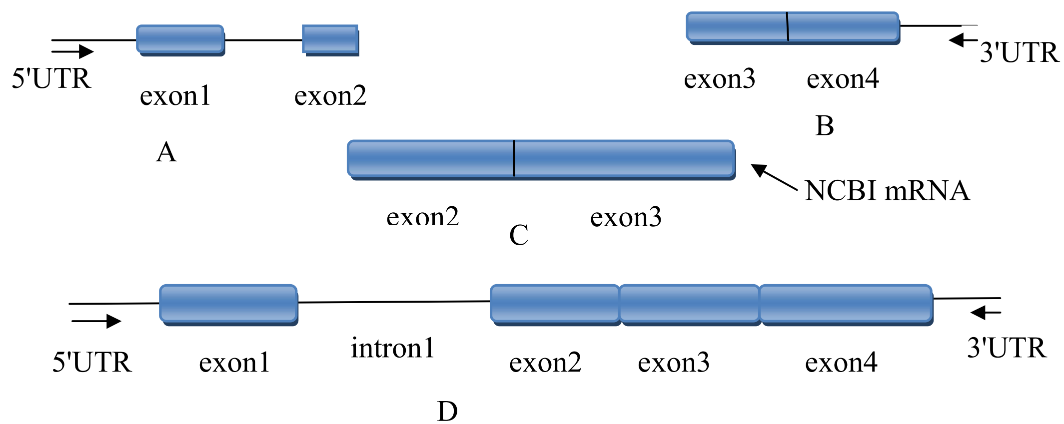


Figure 14. A sketch representing different parts of sdY gene obtained from different experiment and source. A. genomic DNA sequence with exon 1 and exon 2, B. partial Exon 3 and complete exon 4 obtain from 3' RACE, C. partial mRNA published in NCBI and D. Assembled sdY gene.

Successfully able to amplify the longer 5' UTR end of the *Salmo trutta* and *Salmo salar*, the alignment of the published complete genomic sequence of *Salmo trutta* and *Salmo salar* with the assembled 5' UTR end of sdY gene which includes exon 2 of *Salmo trutta* and *Salmo salar* respectively (figure 15).

20 40
Salmo trutta sdY gene 5' end (incl.exon2) CCCTATTCCA TATATAGTGC ACTATTTGTG GTTTACAAAG 40
S.trutta sdY gene(KT223112.1) ---------- ---------- ---------- ---------- -
Consensus CCCTATTCCA TATATAGTGC ACTATTTGTG GTTTACAAAG
100%
Conservation
0%
2.0bits
Sequence logo CCCTATTCCA TATATAGTGC ACTATTTGTG GTTTACAAAG
0.0bits

60 80
Salmo trutta sdY gene 5' end (incl.exon2) AATTATATTG TAGTGTCTTC TGTCTTTATC CTGTCTGGTA 80
S.trutta sdY gene(KT223112.1) ---------- ---------- ---------- ---------- -
Consensus AATTATATTG TAGTGTCTTC TGTCTTTATC CTGTCTGGTA
100%
Conservation
0%
2.0bits
Sequence logo AATTATATTG TAGTGTCTTC TGTCTTTATC CTGTCTGGTA
0.0bits

100 120
Salmo trutta sdY gene 5' end (incl.exon2) CAAGATAAAG TCTGCAAAGA GGAGAAGGGA GAGAGAGTGA 120
S.trutta sdY gene(KT223112.1) ---------- ---------- ---------- ---------- -
Consensus CAAGATAAAG TCTGCAAAGA GGAGAAGGGA GAGAGAGTGA
100%
Conservation
0%
2.0bits
Sequence logo CAAGATAAAG TCTGCAAAGA GGAGAAGGGA GAGAGAGTGA
0.0bits

140 160
Salmo trutta sdY gene 5' end (incl.exon2) GCAAGGCCTT GGCCTTCACC TGACAAGAGA GGATTGAGAG 160
S.trutta sdY gene(KT223112.1) ---------- ---------- ---------- ---------- -
Consensus GCAAGGCCTT GGCCTTCACC TGACAAGAGA GGATTGAGAG
100%
Conservation
0%
2.0bits
Sequence logo GCAAGGCCTT GGCCTTCACC TGACAAGAGA GGATTGAGAG
0.0bits

180 200
Salmo trutta sdY gene 5' end (incl.exon2) AGAAACACTC CCCTTAACTA ATCATTTTAC ATGAAAAAAT 200
S.trutta sdY gene(KT223112.1) -------CT- CCCTT-ACTA ATCATTTTAC ATGAAAAAAT 31
Consensus AGAAACACTC CCCTTAACTA ATCATTTTAC ATGAAAAAAT
100%
Conservation
0%
2.0bits
Sequence logo AGAAACACTc CCCTTaACTA ATCATTTTAC ATGAAAAAAT
0.0bits

220 240
Salmo trutta sdY gene 5' end (incl.exon2) GCATATCAAA AACTCCAGCG CATGTTTAAA AGCAGTCCTT 240
S.trutta sdY gene(KT223112.1) GCATATCAAA AACTCCAGCG CATGTTTAAA AGCAGTCCTT 71
Consensus GCATATCAAA AACTCCAGCG CATGTTTAAA AGCAGTCCTT
100%
Conservation
0%
2.0bits
Sequence logo GCATATCAAA AACTCCAGCG CATGTTTAAA AGCAGTCCTT
0.0bits

260 280
Salmo trutta sdY gene 5' end (incl.exon2) AGAGGTTTTG AATCCACCTG CCCTTCAATG GCTGACAGAG 280
S.trutta sdY gene(KT223112.1) AGAGGTTTTG AATCCACCTG CCCTTCAATG GCTGACAGAG 111
Consensus AGAGGTTTTG AATCCACCTG CCCTTCAATG GCTGACAGAG
100%
Conservation
0%
2.0bits
Sequence logo AGAGGTTTTG AATCCACCTG CCCTTCAATG GCTGACAGAG
0.0bits

300 320
Salmo trutta sdY gene 5' end (incl.exon2) AGGCCAGATT CCAAGGTACA GAACACACAA ATATTCAGTA 320
S.trutta sdY gene(KT223112.1) AGGCCAGATT CCAAGGTACA GAACACACAA ATATTCAGTA 151
Consensus AGGCCAGATT CCAAGGTACA GAACACACAA ATATTCAGTA
100%
Conservation
0%
2.0bits
Sequence logo AGGCCAGATT CCAAGGTACA GAACACACAA ATATTCAGTA
0.0bits

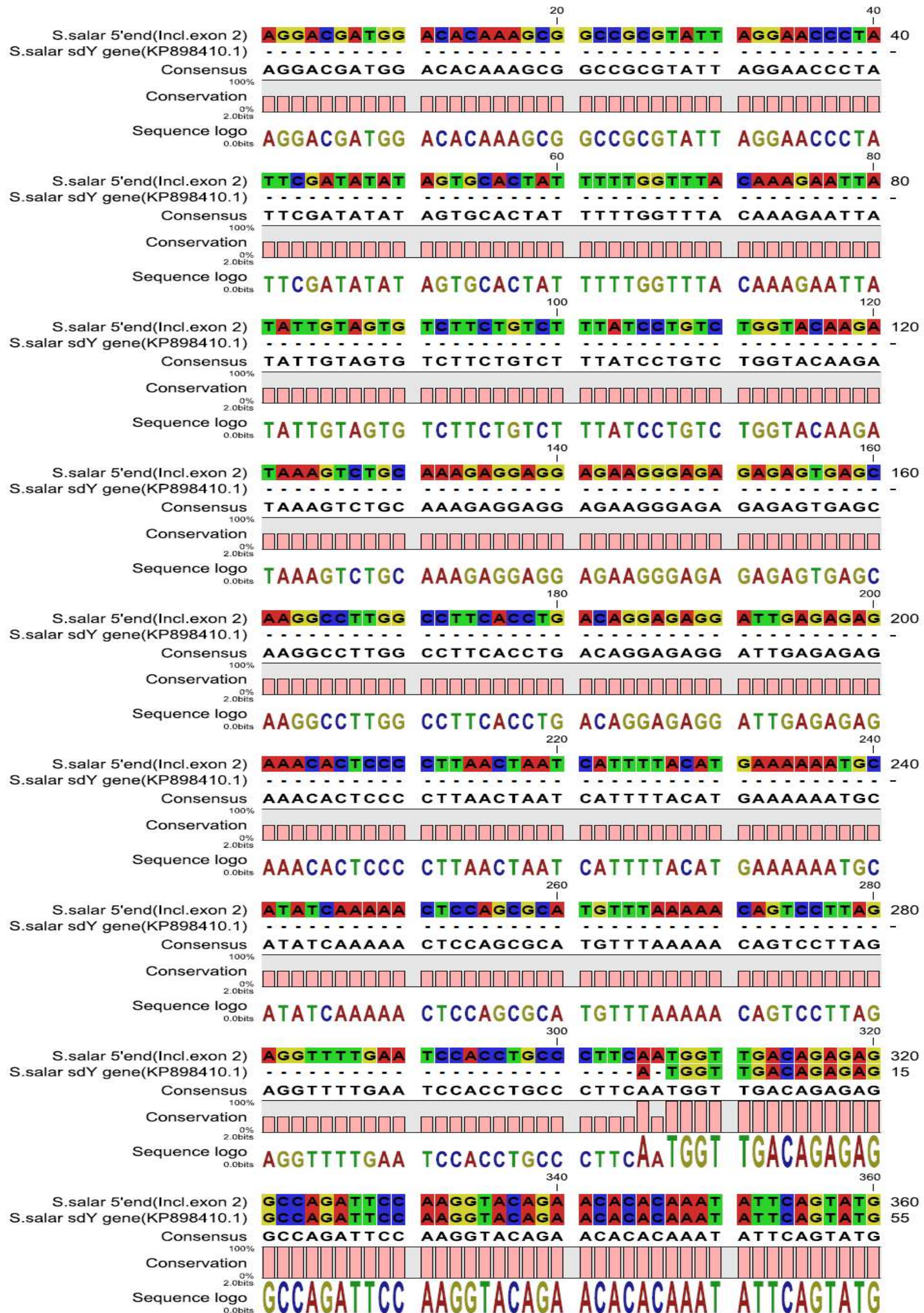


Figure 15. Alignment of published *Salmo trutta* (KT223112.1), *Salmo salar* (KP898410.1) with the assembled sdY gene 5' end including exon 2 of *Salmo trutta* and *Salmo salar* respectively.

Successfully able to obtain longer 5' UTR end of sdY gene of *Salmo trutta* then the published sequence of the *Salmo trutta* (KT223112.1) in GenBank. The manual editing was done at position 170 bp and 176 where we insert a gap to make complete alignment with the published sequence. As shown in figure 15 above 166 bp of long 5' UTR was amplified at the 5' end of the sdY gene for the *Salmo trutta*. In case of *Salmo salar*, able to amplify 305 bp longer 5' URT end of sdY gene than the published genomic sequence of *Salmo salar* (KP898410.1).

(Assembled of sdY gene obtain from different source a describe in figure 14 is presented in appendix 4)

3.5 PCR based genotyping tests were developed to differentiate male and female salmonid species

Based on generated sequence of sdY mRNA (cDNA) gene, published partial mRNA for all three species and published complete genomic DNA sequence of *Salmo trutta* (KT223112.1) and *Salmo salar* (KP898410.1) different primers were designed and different sets of combination of primers were used based on exon 1, exon 2, exon 3 and exon 4, since these exon should be present in a fully functional sdY gene. List of different primers designed are presented in appendix 5. Out of different combinations a duplex and triple PCR based genotyping test developed and further HRM curve analysis was performed to confirm the result.

3.5.1 Duplex PCR genotyping test and High resolution melting curve analysis

The duplex PCR-based genotyping test was developed with the novel sets of primers that were suppose to differentiate male and female of experimental species i.e. *Salmo salar, Salmo trutta, Salvelinus alpinus* and *Salmo marmoratus*. The sdY gene specific primers were designed within exon 3 (hiTAIL_ex3F_2) and 3' UTR end of sdY gene (sdY 3UTR DegR). The Fabp6b Forward and reverse primers were taken from the previous study of (Eisbrenner et al., 2014a), which is supposed to be present in both sex of fish. The Fabp6b amplicon is supposed to be approximately of size 450 bp, and it is used as positive control primer (table

2). The genomic DNA of selected species samples was amplified using sdY gene specific oligonucleotide primers. The amplified fragment using SdY gene specific primers was observed on gel at approximately 300 bp of size which was consistent with the expected size of 282 bp and Fabp6b amplicon was observed at approximately of 500 bp, which was also consistent with the expected size of 450 bp (figure 16)

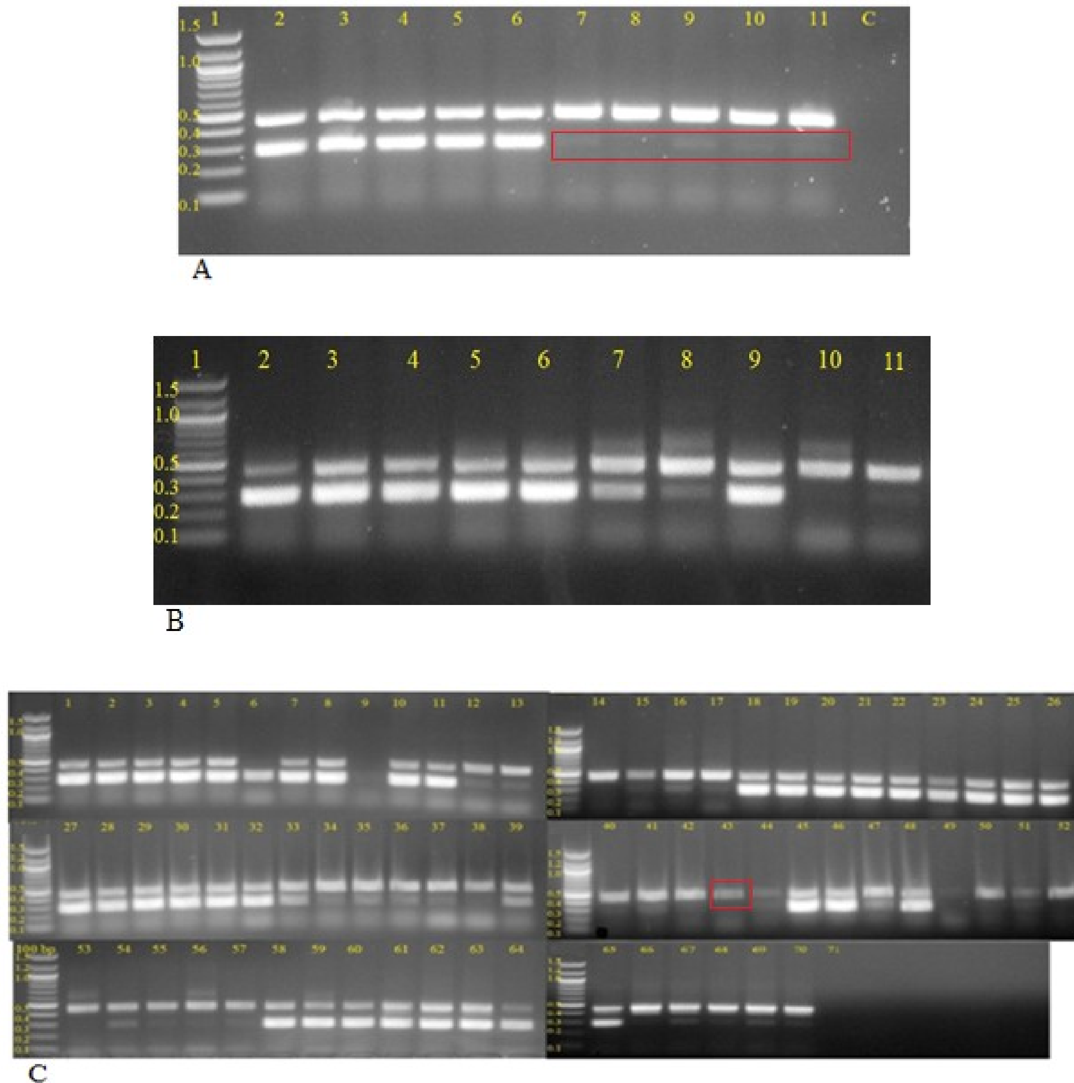


Figure 16. (A-C) Identification of sex of selected fish species by duplex PCR based genotyping test using sdY gene specific primer: hiTAIL_ex3F_2 and sdY 3UTR DegR and positive control primer Fabp6b. A. *Salvelinus alpinus,* lane 1: 100 bp ladder, lane 2-6: male, 7-11: female sample, C: negative control. B. *Salmo trutta*, lane 1: 100 bp ladder, lane 2-6: male sample, lane 7-11: female sample. C. *Salmo marmoratus*, lane: 1-11, 18-32, 41,43-48 and 58-60 are male sample, 12-17,33-40,42 and 50-57 are female sample, 61-65 are *Salmo salar* male , 66-70 are female samples and lane 71: negative control.

These sets of primers were tested on total 90 known samples and as we expect two bands for male i.e. one band for positive control amplicon and another band for the sdY gene specific primers. From the results observed in figure 16, the duplex PCR genotyping test shows strong two bands for all the male samples but faint bands for most of the female samples of all the four selected species. The bands for the *Salvelinus alpinus,* male samples shows two strong bands and the females sample shows very faint bands of the same size of sdY gene (figure 16: panel A). In case of *Salmo trutta*, *Salmo marmoratus* and *Salmo salar,* clear and strong two bands were observed for male samples and clear faint bands of approximate size of sdY amplicon were observed for the most the samples of female. There was some exception for the *Salmo marmoratus* (sample number 9 and 43), which is male samples and supposed to have two amplicon but, in contrast results of sample number 9 does not show any amplicon, as there was no any amplicon even for the positive control.   This may be due to degenerated DNA, for a male (sample number 43 highlighted figure: 16, C), on gel shows a single band of only control amplicon.

Similarly, *Salmo salar* male samples shows two strong bands, one for positive control and other for the sdY specific gene. Whereas the female sample shows strong band for the positive control and faint band for the sdY gene (figure 16. C61 - C70). HRM curve analysis of these samples was done for further analysis.

Melt curve analysis was performed from 40 º C to 95 º C using 0.2 º C increments (as describe in section 2.8). HRM curve analysis shows two different peaks for the male and the female samples (figure 17).

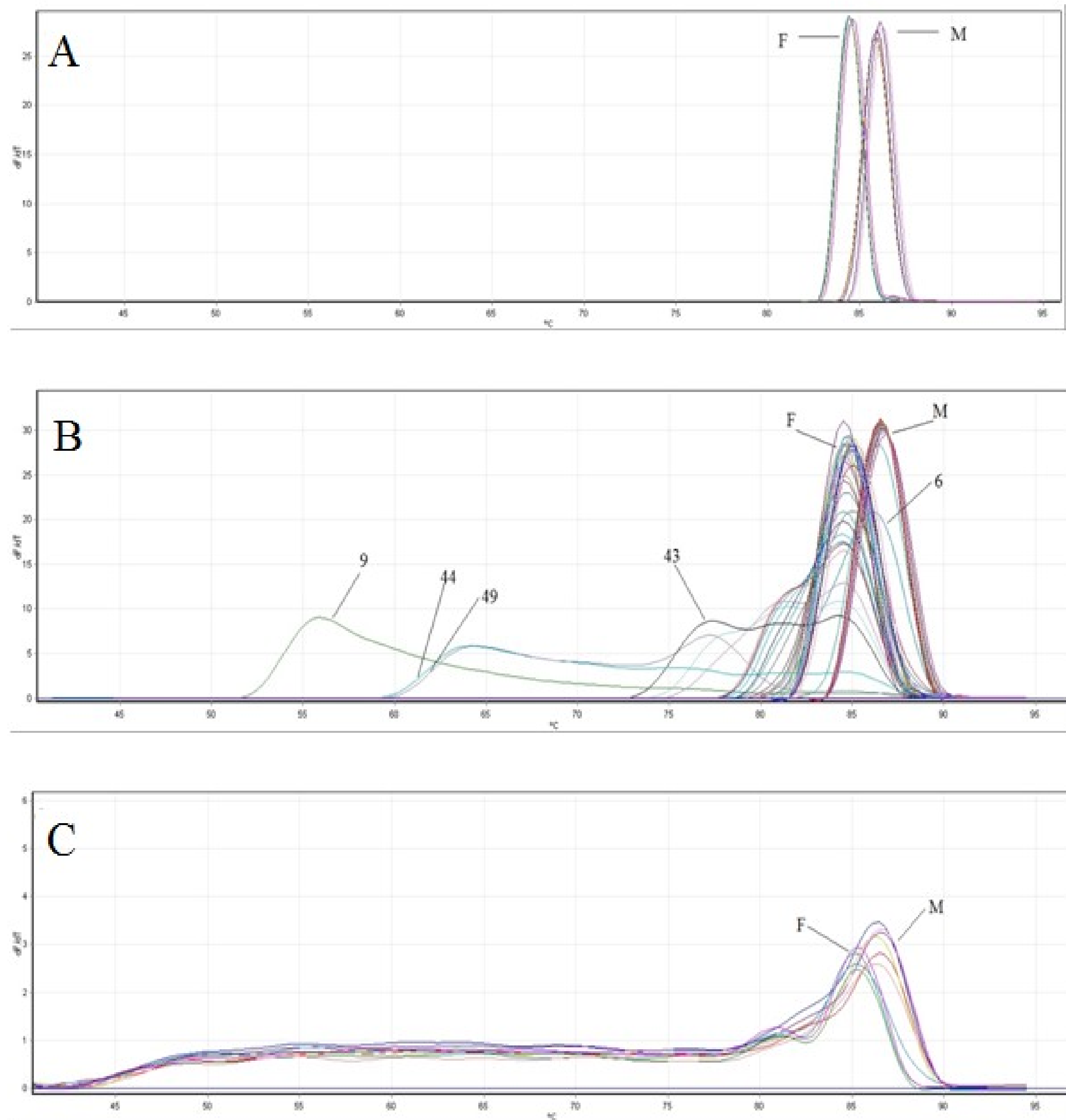


Figure 17. Duplex HRM curve analysis for male and female samples (M: male & F: female). A: *Salvelinus alpinus* B: *Salmo marmoratus* and *Salmo salar* C: *Salmo trutta*

For all the selected species, phenotypic sex identified male samples produce a peak at around 86 ºC and female samples produced a peak at approximately 85 ºC (figure 17) . HRM curve analysis of *Salvelinus alpinus* shows the distinct different peak with different melting temperature for the male and female samples (figure 17 A).In case *Salmo marmoratus* out of 60 sample most of the samples were able to differentiate except for 10% samples (6, 41, 43, 44, 47 and 49), Although sample number 6 was labelled as male sample, the PCR result shows only single band for sdY amplicon, and similarly on HRM the melting peak was observed in between peak of male and female samples. On the other hand samples number

41 and 47 was labelled as male samples, but the PCR and HRM analysis results reveals it as female samples (single band in PCR and Melting peak at 85 ºC). Male labelled samples 43 and 44 shows only single faint band of positive control with PCR based genotyping method, which is irrespective of the double band observed on other male samples, and a different pattern of melting curve was observed on HRM curve analysis as well (Fig. 17 B). Sample number 49 was phenotypically undifferentiated sex from the supplier. Further analysis of this sample on PCR reveals no band both in the positive and the sdY amplicon, similarly different patterns similar to sample no. 44 was observed on HRM analysis.

Among the 10 different samples of *Salmo trutta,* 9 of them follow the above results patterns in PCR-based genotyping experiment. That means strong bands for both positive control and sdY gene in male samples and a faint band of sdY amplicon on female samples (figure16. B). Similarly, for these 9 samples from PCR shows distinct peak in HRM curve analysis for male and female samples. But in case of 9th sample (figure16. B9), although phenotypically it was labelled as female, it shows strong bands for both the positive and sdY gene amplicon on PCR and even in HRM shows melting peak at approximately 86 ºC, which suggest it as a male sample and there may be human error while the sampling. For *salmo salar*, all the samples were able to differentiate with HRM analysis.

HRM curve analysis genotyping test with the identified *Salmo marmoratus* samples that was collected from the three different places in different year (table 3), out of which 10 % does not match with the recorded samples.

Table 3. List of *Salmo marmoratus*.

Serial No.	Length (cm)	Weight (kg)	Approved Sex	HRM confirmed
		Year 2010- Adige		
1	47	1.290	M	+
2	51	1.605	M	+
3	40	0.755	M	+
4	53	2.065	M	+
5	50.5	1.830	M	+
6	45	1.030	M	-
7	48	1.275	M	+

8	41	0.870	M	+
9	47	1.20	M	-
10	47	1.345	M	+
11	56	2.395	M	+
12	43.5	1.140	F	+
13	42.5	1.075	F	+
14	50.5	1.500	F	+
15	44	1.095	F	+
16	51	1.730	F	+
17	47	1.385	F	+
18	45	1.045	M	+
19	49	1.520	M	+
20	48	1.240	M	+
		Year 2009- Brenca		
21	48	1.520	M	+
22	44	0.910	M	+
23	47	1.270	M	+
24	49	1.475	M	+
25	47.5	1.525	M	+
26	44	1.055	M	+
27	38	0.980	M	+
29	50.5	1.650	M	+
30	51	1.675	M	+
31	46	1.340	M	+
32	47	1.280	M	+
33	45	1.375	F	+
34	47	1.580	F	+
35	42.5	1.170	F	+
36	44	1.205	F	+
37	41	1.060	F	+

38	44	1.105	F	+
39	39	0.785	F	+
40	44	1.275	F	+
		Year 2008- Piave		
41	55	1.500	M	-
42	47	1.230	F	+
43	49	1.310	M	-
44	57	1.935	M	-
45	50	1.580	M	+
46	49	1.480	M	+
47	51	1.530	M	-
48	48	1.205	M	+
49	47	1.100	?	?
50	44	1.150	F	+
51	41	0.895	F	+
52	49	1.365	F	+
53	44	1.100	F	+
54	44.5	0.955	F	+
55	48	1.260	F	+
56	44	0.930	F	+
57	60	2.600	F	+
58	47.5	1.445	M	+
59	52	1.755	M	+
60	57.5	1.005	M	+

M = Male, F = Female, + = HRM results consistent with the labelled record of samples, **-** = HRM results against the labelled records of samples and ? = Undifferentiated sample

Few samples do not match with the labelled records that are highlighted in table 3. For the samples number 41 and 47, both the PCR-based genotyping and HRM curve analysis show the same result i.e. strong band of only control amplicon on the gel and the melting temperature of approximately 85 ºC. Therefore for these two samples it can be suggested

that there may be the human error while sampling. For other samples (6, 43, 44, 49) unable to differentiate.

B. Triplex PCR genotyping test and High resolution melting (HRM) curve analysis

Triplex PCR genotyping test was also developed to differentiate the male and female with the three sets of specific sdY gene primers i.e. sdY exon 3F (381-399) and sdY exon 3R (444-462), sdY exon 4 F2 and sdY 3'UTR R2 and a set of positive quality control Fabp6b primers (table 2.). Since both sets of primers are sdY gene specific, a fully functional sdY gene most contains these amplicon and amplified during PCR reaction. Any individual in which both of these amplicon amplified were designated as predicted male and Fabp6b as the positive control should present in both sex of the samples. The DNA fragment amplified using sdY gene specific oligonucleotide primers: sdY exon 3F (381-399) and sdY exon 3R (444-462) was observed on gel with approximate size of 100 bp which is consistent with expected size 82 bp and for primers sdY exon 4 F2 and sdY 3'UTR R2 bands observed was approximately of 200 bp which is consistent with the expected size 183 bp (figure 18).

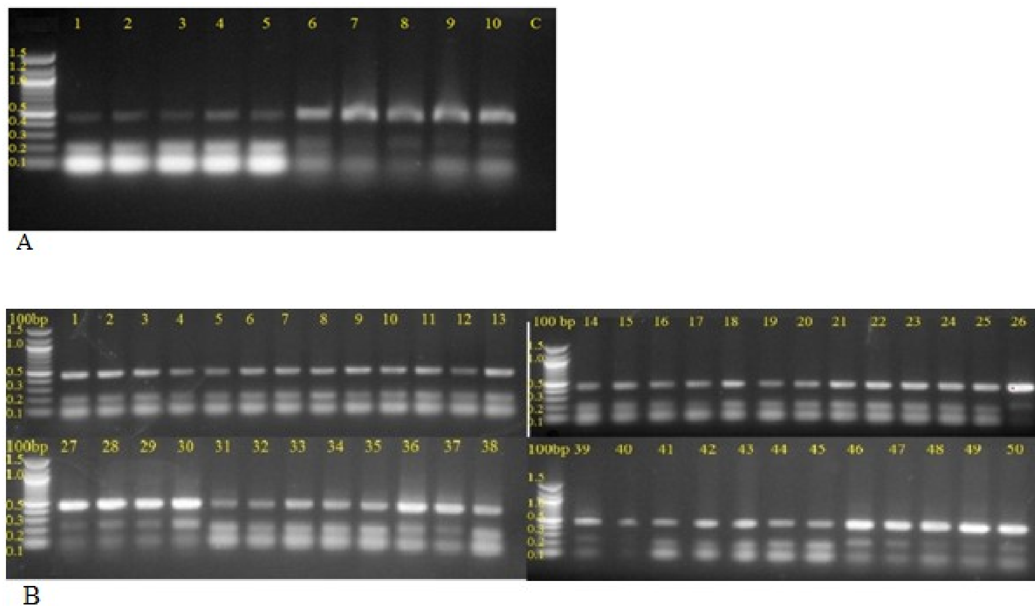


Figure 18. (A&B) Identification of male and female of selected fish by PCR based genotyping tests. A. Lane A1-A5: male, lane A6-A10 female *Salmo marmoratus* and lane C: negative control. B. lane B1-B20 & B41-B45: male and B46-B50 female *Salmo salar*, lane B21-B25: male *Salvelinus alpinus* and B26-B30 female *Salvelinus alpinus* samples, lane B31-B35: male and lane B36-B40: female *Salmo trutta* samples.

This triplex PCR genotyping test was tested on all these four species of samples. We expect three band for male samples and single band for the female samples but in figure 18 we observed 3 bands for both male and female samples, only the difference is that male has strong sdY gene specific bands and female have faint bands for sdY specific amplicon. Based on these finding it can be concluded that the triplex PCR genotyping does not work for any of the four species (*Salmo marmoratus, Salmo trutta, Salmo salar* and *Salvelinus alpinus*). Further analysis was done by the HRM curve analysis.

Melting curve analysis was performed with the objective to get different melting peak for different amplicons. Melt curve analysis was performed from 40 to 95 ºC using 0.2 increments (as described in section 2.8). All phenotypic sex identified males produced two peaks and all phenotypic sex identified female samples produced one peak (figure 19).

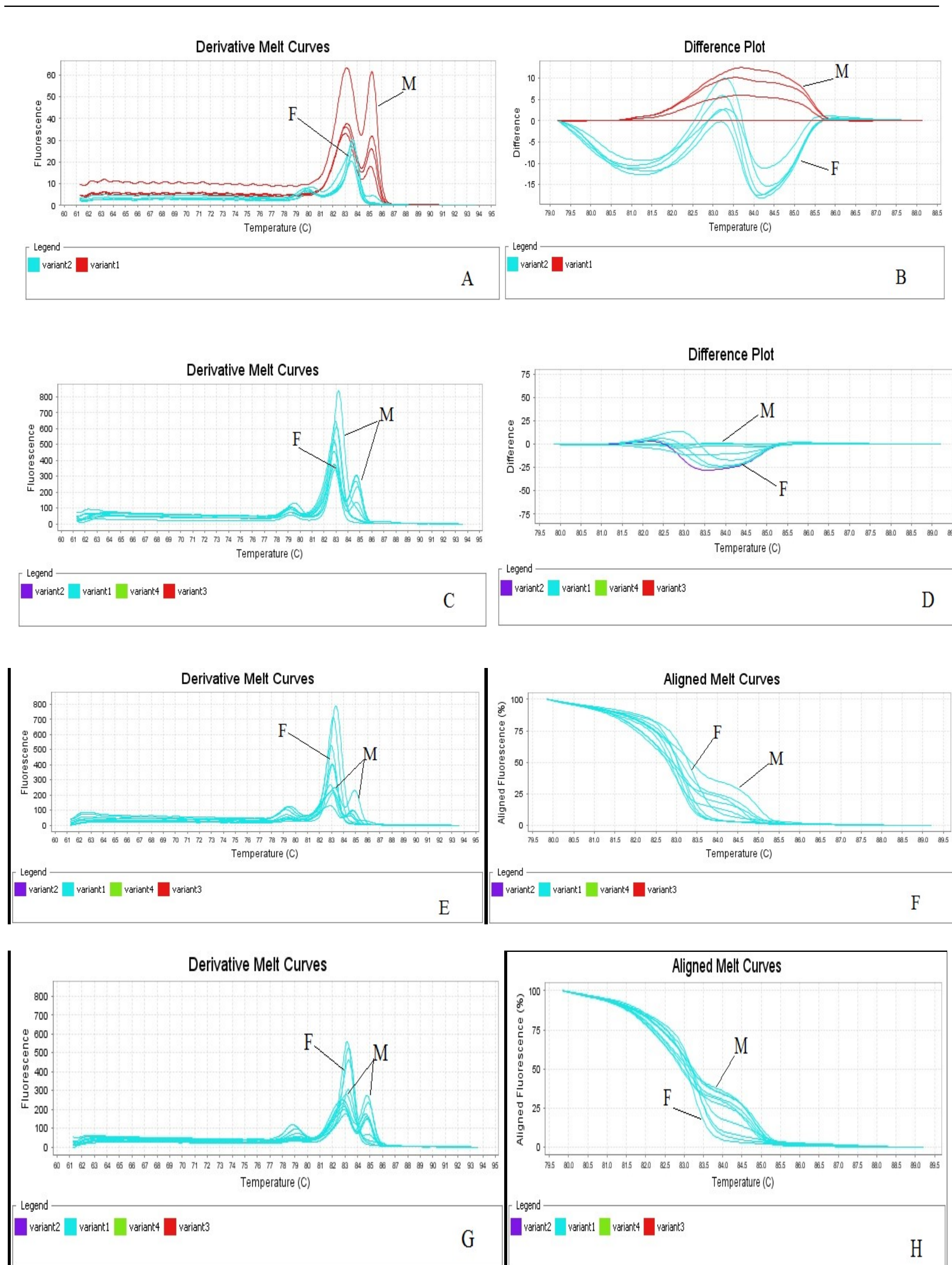


Figure 19. Triplex quantitative polymerase chain reaction (qPCR) melt curve analysis for male (red) and female (green) fish samples. A. Derivative melt curves, B. difference plot for *Salmo marmoratus*, C. derivative melt curves, D. difference plot for *Salvelinus alpinus.* E. derivative melt curves F. aligned melt curves for *Salmo salar* and G. derivative melt curves, H. aligned mellt curves *for Salmo trutta*. M: male & F: female

Result of the HRM curve analysis show distinct difference between the male and female samples, clear two peak for the male fish samples and single peak for the female samples was observed on the derivative melt curves for all the selected species. Melting temperature for male samples was observed from the HRM melting curve was approximately 85 ºC and for the female samples was approximately 83 ºC. The difference plot also show distinct difference between male and female samples, for *Salmo marmoratus* as shown in figure 19 (B) wave like structure is female samples (green) and red line represents male samples. The difference plot for *Salvelinus alpinus* male samples was identified as the straight line and the S-shape structure was identified as female samples figure 19 (D). Aligned melt curves for the *Salmo salar* and *Salmo trutta* show different pattern for the male and female samples as it is clearly shown in figure 19 (F and H). Even the triplex PCR genotyping tests were unable to differentiate sex of all the selected species the HRM curve analysis clearly able to differentiate the sex of all the selected species.

4. Discussion

The purpose for selecting the following species of family Salmonid for this study: *Salmo trutta*, *Salmo salar*, *Salvelinus alpinus* and *Salmo marmoratus* was that these species are very important in wild fisheries and in aquaculture industries. The recent discovery of the sdY gene was identified as male specific linked gene on Y chromosome in rainbow trout and was considered as male specific master sex-determining gene (Yano et al., 2012). Most of the salmonid family (Salmoninae, Coregoninae and Thymallinae subfamilies) share a conserved sex-determining locus and also suggests that sdY may be a conserved master sex-determining gene (Yano et al., 2013).

Two popular hypotheses have been proposed for the salmonid sex determination. The first hypothesis suggests that salmonids utilize the same sex determining gene that is jumping from the end of one chromosome to the end of different chromosome mediated either by transposons or by homologous recombination during meiosis (RB Phillips, Konkol, Reed, & Stein, 2001). The second hypothesis suggest that different species among the salmonids have different sex determining genes that evolved independently and may be related to speciation (Davidson et al., 2009).

The sdY gene is not located at homologous genomic positions among different salmonid species, its jumping has been associated to mobile elements and supports the first hypothesis of salmonid sex determination (Yano et al., 2012). The sdY gene has evolved through neofunctionalization from the Irf9 gene, which lost its role in IFN signaling pathway and acquired new function in sex determination (Yano et al., 2012). Despite of its position in the genome, sequence of sdY gene may useful in sex determination among salmonids.

4.1 Primer development

One of the aims of this project was to determine the full-length sequence of sdY mRNA (cDNA) of *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*. From the previous study, sdY (mRNA) gene was identified as master sex-determining gene in Rainbow trout and a fully functional sdY gene consist of exon 1, exon 2, exon 3 and exon 4 (Yano et al., 2012). The partial mRNA of *Salmo trutta* (JF826019.1), *Salmo salar* (JF826020.1) and *Salvelinus alpinus* (JF826022.1) are available at NCBI. Based on sequence alignments of sdY gene of

Rainbow trout and the published partial mRNA of *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*, we were able to determine that these partial mRNA sequences consist of the complete exon 2 and exon 3. A fully fuctional sdY gene contain exon 1, exon 2, exon 3 and exon 4 so we decided to design primers targeting the conserved region of exon 2 and exon 3 of published partial mRNA (table 1), which can be used for 5' and 3' RACE to amplify 5' and 3' end of the *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*. On other hand, for the hiTAIL PCR primer were designed following (Liu & Chen, 2007).

Other objective of this project was to develop PCR-based genotyping method that can differentiate male and female samples, that can confirmed by HRM analysis. For the PCR-based method, challenge was to identify a conserved region of sdY sequence to design primers that can be used to amplify sdY gene for all three selected species.

4.2 RLM-RACE

The primers design on the conserved region of the sdY gene was used for the 3' RACE to amplify the 3' end of the *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*. Primer based on exon 2 and exon 4 of the sdY gene was published (Eisbrenner et al., 2014a), although these primers shows alignment with our selected species, they were not useful because published the exon 2 forward primer starts from the position 112 to 146 bp of exon 2. As we desired to cover a maximum region of exon 2 so we decided to design a new exon 2 forward primer that starts from the position 34-51 of the exon 2 of the sdY gene. Other sets of primers were design against exon 3, and they were used for the 3' RACE to successfully amplify the 3' end of mRNA (cDNA) of sdY gene of *Salmo trutta* and *Salvelinus alpinus* (section 3.2).

The project was not carried out as it was planned because last summer (July-2015) a research group at Simon Fraser University in Canada published the complete genomic sdY sequence of *Salmo trutta* (GenBank accession number: KT223112.1) and *Salmo salar* (GenBank accession number: KP898410.1). After discussion with my supervisors it was decided to perform hiTAIL PCR to amplify the 5' and 3' end of sdY gene to the extend the sequence information for the sdY gene as a supplement to the 3' end sequence already elucidated by 3' RACE.

4.3 Sequencing and assembling

hiTAIL PCR was performed to determine the 5' flanking region of the sdY gene of *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*. The resulted in the successfully determination of the 5 flanking region of the *Salmo trutta* and *Salmo salar* sdY gene. The published sequence of the *Salmo trutta* (GenBank accession nummber: KT223112.1) provides only 21 nucleotides upstream to the start codon, while we were able to extend this to 190 nucleotides. Also, the published sequence of *Salmo salar* (GenBank accession number: KP898410.1) does not provide any 5' UTR sequence information as the sequence starts from the start codon (ATG); we are able to determine 304 nucleotides of genomic sequence at 5' end, which can play important role to determine the location of the sdY gene in the genome and the 5' UTR sequence of the encoded transcript. Unfortunately, we were unable to obtain 5' flanking region of the *Salvelinus alpinus* gene. This might be due to lower similarity between genera, since *Salmo trutta* and *Salmo salar* belongs to same genus.

The alignment of 3' RACE of *Salmo trutta* and published genomic sequence of *Salmo trutta* (KT223112.1) revealed that the sequence of mRNA (cDNA) obtain from the 3' RACE consist of longer (116 bp) 3' UTR of the sdY (mRNA) than the published genomic sequence of *Salmo trutta*. This 3' UTR end of the sdY (mRNA) can play an important role to design primer and able to amplify the 3' UTR region of the sdY gene.

GenBank containing only a partial mRNA sequence for *Salvelinus alpinus* (i.e. exon 2 and exon 3) but has no information about the exon 4 sequence in this species. Alignment of the 3' RACE product generated in this study from *Salvelinus alpinus* with the published genomic sequence of *Salmo trutta* and *Salmo salar* shows that the 3' RACE of *Salvelinus alpinus* align completely with exon 4 of these two published genomic sequence. Therefore, it can be concluded that the complete sequence information about exon 4 of *Salvelinus alpinus* sdY is now available and this sequence information will play important for the amplification of full-length sequence of sdY gene in *Salvelinus alpinus.*

4.4 Duplex PCR and HRM analysis

The sdY gene was identifed as the master sex-determining gene in Rainbow trout (Yano et al., 2012) and the sdY gene are consider as conserved master sex-determining in most of the salmonid species (Yano et al., 2013). The different sets of primers were designed within the exons of the sdY gene with the target to develop combinations of primers that can be used to identify the male specific gene (sdY) in different species of salmonids. Recently, PCR-based/agarose gel genotyping method was developed to identify the sex of Tasmanian Atlantic salmon, based on the primers design on the exon 2 and exon 4 of the sdY gene (Eisbrenner et al., 2014a). According to the published information the primers were designed against exon 2 and exon 4 of the sdY gene but when we aligned the exon 4 primers with the published partial mRNA sequence of *Salmo trutta*, *Salmo salar* and *Salvelinus alpinus*. They aligned with exon 3 of all three species so it can be concluded that the published exon 4 primers were actually exon 3 primers.

A duplex PCR-based/agarose gel genotyping test was developed with the combination of Fabp6b forward and reverse primer as a control taken from the previous study (Eisbrenner et al., 2014a) and the hiTAIL_ex3F_2 as forward and sdY 3UTR DegR reverse primer. Results of PCR using these sets of primers shows apparent amplifiction of sdY-like bands for the female samples which could argue against the hypothesis that the sdY gene is conserved master sex determining gene in most of the salmonid (Yano et al., 2012).

High resolution melting (HRM) curve analysis were performed with the EvaGreen qPCR super mix and with the same sets of primers (hiTAIL_ex3F_2 and sdY 3UTR DegR). The HRM analysis result showed two clear peaks with two different melting temperatures approximately 86 ºC for male and 85 ºC for female (figure 17). There are some samples that do not match with the label sampling record of the *Salmo marmoratus* (10 %) and *Salmo trutta* with the PCR-based/agarose gel genotyping method. For the further confirmation of those sample run with the HRM analysis and the result of PCR-based/agarose gel genotyping and HRM analysis seems to match with each other, so it can be suggested that there may be human error while sampling the sex. Further conclusion can be done by testing these methods with more samples of known sex.

4.5 Triplex PCR and HRM analysis

A triplex PCR-based/agarose gel genotyping test was developed with the combination of one set of control Fabp6b forward and reverse primers and two sets of sdY specific primers, namely- sdY ex 3 F (381-399) and sdY ex 3 R(444-462), sdY exon 4 F2 and sdY 3'UTR R2. Result of PCR using these sets of primers also shows apparent amplification of sdY-liked bands for the female samples which could argue against the hypothesis that suggest that the sdY gene is the conserved master sex determining gene in most of the salmonid (Yano et al., 2012). Alternatively, the primer sets may not be specific to sdY and may also amplify other, homologous sequences in female salmonids.

HRM analysis was performed with the same sets of primers as in triplex PCR with the EvaGreen qPCR super mix. The result of HRM analysis shows clear differences between the patterns of melting curves of male and female samples. When we compare the results of agarose gels and the HRM analysis, two questions can be raised; the first question is why the pattern which looks almost the same on agarose gels shows significant difference with the use of HRM analysis? The other question is why we obtained amplification using the sdY specific gene primers in female sample? These question may be explained following the finding of (Yano et al., 2012) who suggest that the sdY gene has evolved through neofunctionalization from the Irf9 gene, so it can be suggested that the amplicon obtained using primer sets- sdY exon 3F (381-399) and sdY exon 3R (444-462), sdY exon 4 F2 and sdY 3' UTR R2 contain a mixture of true the sdY amplicon in addition to an Ifr9 amplicon. The Ifr9 sequences are present in both male and female samples; for male samples strong PCR amplicon was observed suggesting that these amplicons are the mixture of both sdY and Ifr9 sequence. For the female samples; only faint PCR bands were observed which suggest that these are only Ifr9 sequence. For the HRM genotyping analysis only the nature of sequence are enough to identify male and female.

Further analysis could have done to know more about the sdY-liked amplicon that was observed in the female samples. As we have limited time so unable to carry out further analysis. This could have done by performing cloning and sequencing of these amplicons. In case of duplex PCR method, we could have done uniplex PCR for female samples with the combination of hiTAIL_ex3F and sdY 3UTR DegR primers. Either the PCR product has done direct sequencing of PCR product or by cloning in the vector and then sequencing to identify homologous sequences of sdY-liked sequence in BLAST. After sequencing if sdY-

liked amplicon shows homologous with the Irf9 gene then we could have able to suggest more accurately that the sdY gene was evolved from the Irf9 gene. Similarly, for the sdY-liked amplicons from the triplex PCR method that was observed in the female samples can be analysed following similar methods.

5. Conclusion

Both the duplex PCR-based/agarose gel and triplex PCR-based/agarose gel genotyping test cannot able to differentiate the sex of the selected species (*Salmo trutta, Salmo salar Salvelinus alpinu* and *Salmo marmoratus*). HRM analysis with the same sets of primers as in duplex PCR and triplex PCR can easily differentiate the sex of selected species. These HRM analyses can be classified as quick diagnostic sex tests as the data analysis can be accomplished with in few hours and can be performed using crude tissue (fins) as a source of DNA.

These methods should be tested on more samples to be certain about the method and should be also tested on other species of salmonids to check whether these methods are able to differentiate the sex of other species of salmonids.

6. References

Allendorf, F. W., & Thorgaard, G. H. (1984). Tetraploidy and the evolution of salmonid fishes *Evolutionary genetics of fishes* (pp. 1-53): Springer.

Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., . . . Wang, Y.-L. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science, 282*(5394), 1711-1714.

Baroiller, J.-F., Guiguen, Y., & Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *Cellular and Molecular Life Sciences CMLS, 55*(6-7), 910-931.

Baroiller, J., d'Cotta, H., & Saillant, E. (2008). Environmental effects on fish sex determination and differentiation. *Sexual development: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation, 3*(2-3), 118-135.

Bernard, P., Ryan, J., Sim, H., Czech, D. P., Sinclair, A. H., Koopman, P., & Harley, V. R. (2011). Wnt signaling in ovarian development inhibits Sf1 activation of Sox9 via the Tesco enhancer. *Endocrinology, 153*(2), 901-912.

Christoffels, A., Koh, E. G., Chia, J.-m., Brenner, S., Aparicio, S., & Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Molecular biology and evolution, 21*(6), 1146-1151.

Chue, J., & Smith, C. A. (2011). Sex determination and sexual differentiation in the avian model. *Febs Journal, 278*(7), 1027-1034.

Cutting, A., Chue, J., & Smith, C. A. (2013). Just how conserved is vertebrate sex determination? *Developmental Dynamics, 242*(4), 380-387.

Davidson, W., Huang, T.-K., Fujiki, K., Von Schalburg, K., & Koop, B. (2009). The sex determining loci and sex chromosomes in the family Salmonidae. *Sexual Development, 3*(2-3), 78-87.

Davidson, W., Huang, T., Fujiki, K., Von Schalburg, K., & Koop, B. (2008). The sex determining loci and sex chromosomes in the family salmonidae. *Sexual development: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation, 3*(2-3), 78-87.

Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture, 208*(3), 191-364.

di Clemente, N., Ghaffari, S., Pepinsky, R., Pieau, C., Josso, N., Cate, R., & Vigier, B. (1992). A quantitative and interspecific test for biological activity of anti-mullerian hormone: the fetal ovary aromatase assay. *Development, 114*(3), 721-727.

Eggers, S., & Sinclair, A. (2012). Mammalian sex determination—insights from humans and mice. *Chromosome Research, 20*(1), 215-238.

Eisbrenner, W., Botwright, N., Cook, M., Davidson, E., Dominik, S., Elliott, N., . . . Lubieniecki, K. (2014a). Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (Salmo salar). *Heredity, 113*(1).

Eisbrenner, W., Botwright, N., Cook, M., Davidson, E., Dominik, S., Elliott, N., . . . Lubieniecki, K. (2014b). Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (Salmo salar). *Heredity, 113*(1), 86-92.

Fishelson, L. (1970). Protogynous sex reversal in the fish Anthias squamipinnis (Teleostei, Anthiidae) regulated by the presence or absence of a male fish.

Freeman, S., Herron, J. C., HHodin, J. A., Miner, B., & Sidor, C. (2007). *Evolutionary analysis*: Pearson Prentice Hall Upper Saddle River, NJ.

Graves, J. A. M. (2008). Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annual review of genetics, 42*, 565-586.

Graves, J. A. M., Ferguson-Smith, M., McLaren, A., Mittwoch, U., Renfree, M., & Burgoyne, P. (1995). The Evolution of Mammalian Sex Chromosomes and the Origin of Sex Determining Genes [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences, 350*(1333), 305-312.

Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., . . . Strüssmann, C. A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences, 109*(8), 2955-2959.

Hayes, T. B. (1998). Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. *The Journal of experimental zoology, 281*(5), 373-399.

Heiligenberg, W. (1965). *Colour polymorphism in the males of an African cichlid fish.* Paper presented at the Proceedings of the Zoological Society of London.

Hiramatsu, R., Matoba, S., Kanai-Azuma, M., Tsunekawa, N., Katoh-Fukui, Y., Kurohmaru, M., . . . Kanai, Y. (2009). A critical time window of Sry action in gonadal sex determination in mice. *Development, 136*(1), 129-138.

Hoegg, S., Boore, J. L., Kuehl, J. V., & Meyer, A. (2007). Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni. *BMC Genomics, 8*(1), 317.

Hudsonb, C. S. K. R. Q., & Sinclaira, A. (2007). Avian sex determination: what, when and where? *Cytogenet Genome Res, 117*, 165-173.

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics, 11*(2), 97-108.

Jameson, J. L., Achermann, J. C., Ozisik, G., & Meeks, J. J. (2003). Battle of the sexes: new insights into genetic pathways of gonadal development. *Transactions of the American Clinical and Climatological Association, 114*, 51.

Jameson, S. A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D. M., Mork, L., . . . Capel, B. (2012). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet, 8*(3), e1002575.

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., . . . Hosoya, S. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu). *PLoS Genet, 8*(7), e1002798.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., . . . Kasai, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature, 447*(7145), 714-719.

Kikuchi, K., & Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. *Developmental Dynamics, 242*(4), 339-353.

Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.-C., . . . Capel, B. (2006). Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. *PLoS biology, 4*(6), e187.

Kobayashi, T., & Nagahama, Y. (2008). Molecular aspects of gonadal differentiation in a teleost fish, the Nile tilapia. *Sexual development: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation, 3*(2-3), 108-117.

Kondo, M., Nanda, I., Hornung, U., Asakawa, S., Shimizu, N., Mitani, H., . . . Schartl, M. (2003). Absence of the candidate male sex-determining gene dmrt1b (Y) of medaka from other fish species. *Current Biology, 13*(5), 416-420.

Koopman, P., Münsterberg, A., Capel, B., Vivian, N., & Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature, 348*(6300), 450-452.

Kumar, S., & Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature, 392*(6679), 917-920.

Li, J. (2010). *Sex determination in the genus Salmo.* Science: Department of Molecular Biology and Biochemistry.

Liu, Y. G., & Chen, Y. (2007). High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *BioTechniques, 43*, 649-656.

Lynch, M., & Force, A. G. (2000). The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist, 156*(6), 590-605.

Mank, J. E., & Avise, J. C. (2006). Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica, 127*(1-3), 321-327.

Matson, C. K., Murphy, M. W., Sarver, A. L., Griswold, M. D., Bardwell, V. J., & Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature, 476*(7358), 101-104.

Matsuda, M. (2003). Sex determination in fish: Lessons from the sex-determining gene of the teleost medaka, Oryzias latipes. *Development, growth & differentiation, 45*(5-6), 397-403.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., . . . Shimizu, N. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature, 417*(6888), 559-563.

Meyer, A., & Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays, 27*(9), 937-945.

Moghadam, H. K., Ferguson, M. M., & Danzmann, R. G. (2005a). Evidence for Hox gene duplication in rainbow trout (Oncorhynchus mykiss): a tetraploid model species. *Journal of molecular evolution, 61*(6), 804-818.

Moghadam, H. K., Ferguson, M. M., & Danzmann, R. G. (2005b). Evolution of Hox clusters in Salmonidae: a comparative analysis between Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss). *Journal of molecular evolution, 61*(5), 636-649.

Munger, S. C., & Capel, B. (2012). Sex and the circuitry: progress toward a systems-level understanding of vertebrate sex determination. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 4*(4), 401-412.

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., . . . Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, Oryzias luzonensis. *Genetics, 191*(1), 163-170.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., . . . Nishigaki, R. (2000). A detailed linkage map of medaka, Oryzias latipes: comparative genomics and genome evolution. *Genetics, 154*(4), 1773-1784.

Nelson, J. (1994). Fishes of the world. New York: J: Wiley.

Ohno, S. (1970). *Evolution by gene duplication*: Springer-Verlag.

Ospina-Alvarez, N., & Piferrer, F. (2008). Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS One, 3*(7), e2837.

Otake, H., Shinomiya, A., Matsuda, M., Hamaguchi, S., & Sakaizumi, M. (2006). Wild-derived XY sex-reversal mutants in the medaka, Oryzias latipes. *Genetics, 173*(4), 2083-2090.

Patiño, R., Davis, K. B., Schoore, J. E., Uguz, C., Strüssmann, C. A., Parker, N. C., . . . Goudie, C. A. (1996). Sex differentiation of channel catfish gonads: normal development and effects of temperature. *Journal of Experimental Zoology, 276*(3), 209-218.

Phillips, R., Konkol, N., Reed, K., & Stein, J. (2001). Chromosome painting supports lack of homology among sex chromosomes in Oncorhynchus, Salmo, and Salvelinus (Salmonidae). *Genetica, 111*(1-3), 119-123.

Phillips, R., & Rab, P. (2001). Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews of the Cambridge Philosophical Society, 76*(01), 1-25.

Phillips, R. B., Keatley, K. A., Morasch, M. R., Ventura, A. B., Lubieniecki, K. P., Koop, B. F., . . . Davidson, W. S. (2009). Assignment of Atlantic salmon (Salmo salar) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (Oncorhynchus mykiss). *Bmc Genetics, 10*(1), 46.

Platt, A. R., Woodhall, R. W., & George, A. L. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques, 43*(1), 58.

Raymond, C. S., Parker, E. D., Kettlewell, J. R., Brown, L. G., Page, D. C., Kusz, K., . . . Bardwell, V. J. (1999). A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. *Human Molecular Genetics, 8*(6), 989-996.

Rodríguez-Marí, A., Yan, Y.-L., BreMiller, R. A., Wilson, C., Cañestro, C., & Postlethwait, J. H. (2005). Characterization and expression pattern of zebrafish Anti-Müllerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. *Gene Expression Patterns, 5*(5), 655-667.

Rubin, D. A. (1985). Effect of pH on sex ratio in cichlids and a poecilliid (Teleostei). *Copeia*, 233-235.

Saito, D., & Tanaka, M. (2008). Comparative aspects of gonadal sex differentiation in medaka: a conserved role of developing oocytes in sexual canalization. *Sexual development: genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation, 3*(2-3), 99-107.

Shoemaker, C. M., & Crews, D. (2009). *Analyzing the coordinated gene network underlying temperature-dependent sex determination in reptiles.* Paper presented at the Seminars in cell & developmental biology.

Sidow, A. (1996). Gen (om) e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev, 6*(6), 715-722.

Sinclair, A., Berta, P., Palmer, M., Hawkins, J., Griffiths, B., & Palmer, M. (1990). A gene from the human sex-determining region encodes a protein with. *Nature, 346*(6281), 240-244.

Sinisi, A., Pasquali, D., Notaro, A., & Bellastella, A. (2002). Sexual differentiation. *Journal of endocrinological investigation, 26*(3 Suppl), 23-28.

Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., & Sinclair, A. H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature, 461*(7261), 267-271.

Sunobe, T., & Nakazono, A. (1993). Sex change in both directions by alteration of social dominance in Trimma okinawae (Pisces: Gobiidae). *Ethology, 94*(4), 339-345.

Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, Oryzias javanicus and O. hubbsi. *Chromosome Research, 16*(5), 801-811.

Takehana, Y., Naruse, K., & Sakaizumi, M. (2005). Molecular phylogeny of the medaka fishes genus Oryzias (Beloniformes: Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. *Molecular phylogenetics and evolution, 36*(2), 417-428.

Thorgaard, G. H. (1977). Heteromorphic sex chromosomes in male rainbow trout. *Science, 196*(4292), 900-902.

Uchida, D., Yamashita, M., Kitano, T., & Iguchi, T. (2004). An aromatase inhibitor or high water temperature induce oocyte apoptosis and depletion of P450 aromatase activity in the gonads of genetic female zebrafish during sex-reversal. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 137*(1), 11-20.

Uhlenhaut, N. H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., . . . Holter, N. I. (2009). Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell, 139*(6), 1130-1142.

Urton, J., McCann, S., & Peichel, C. (2011). Karyotype differentiation between two stickleback species (Gasterosteidae). *Cytogenet Genome Res, 135*(2), 150-159.

Volff, J. (2005). Genome evolution and biodiversity in teleost fish. *Heredity, 94*(3), 280-294.

von Hofsten, J., & Olsson, P.-E. (2005). Zebrafish sex determination and differentiation: involvement of FTZ-F1 genes. *Reproductive Biology and Endocrinology, 3*(1), 63.

Wallace, B., & Wallace, H. (2003). Synaptonemal complex karyotype of zebrafish. *Heredity, 90*(2), 136-140.

Waters, P. D., Wallis, M. C., & Graves, J. A. M. (2007). *Mammalian sex—origin and evolution of the Y chromosome and SRY.* Paper presented at the Seminars in cell & developmental biology.

Woram, R. A., Gharbi, K., Sakamoto, T., Hoyheim, B., Holm, L.-E., Naish, K., . . . Stein, J. (2003). Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res, 13*(2), 272-280.

Yamamoto, E. (1999). Studies on sex-manipulation and production of cloned populations in hirame, Paralichthys olivaceus (Temminck et Schlegel). *Aquaculture, 173*(1), 235-246.

Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., . . . Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, Oncorhynchus mykiss. *Current Biology, 22*(15), 1423-1428.

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary applications, 6*(3), 486-496.

Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., . . . Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. *Proceedings of the National Academy of Sciences, 105*(7), 2469-2474.

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution, 18*(6), 292-298.

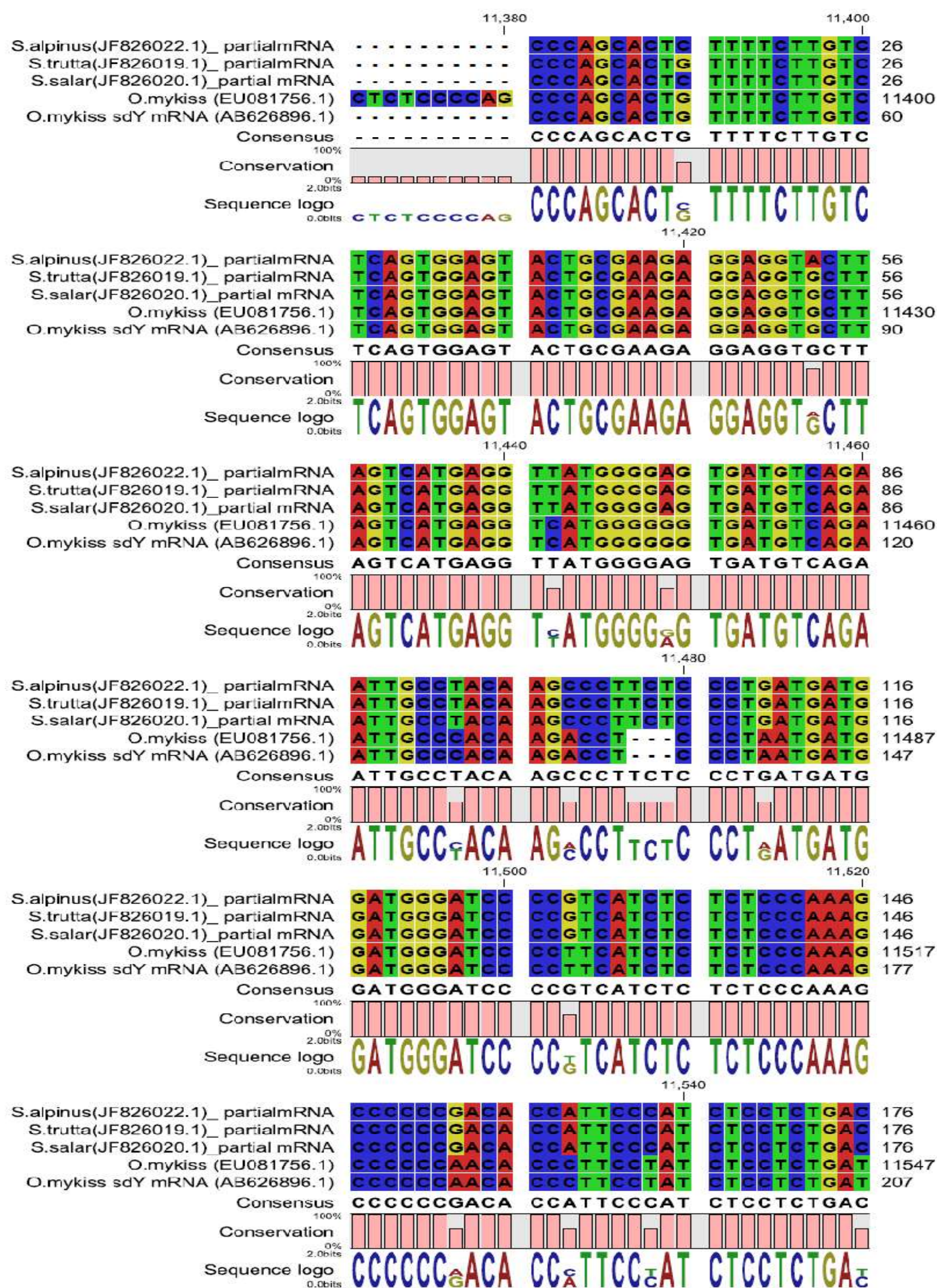
Appendices

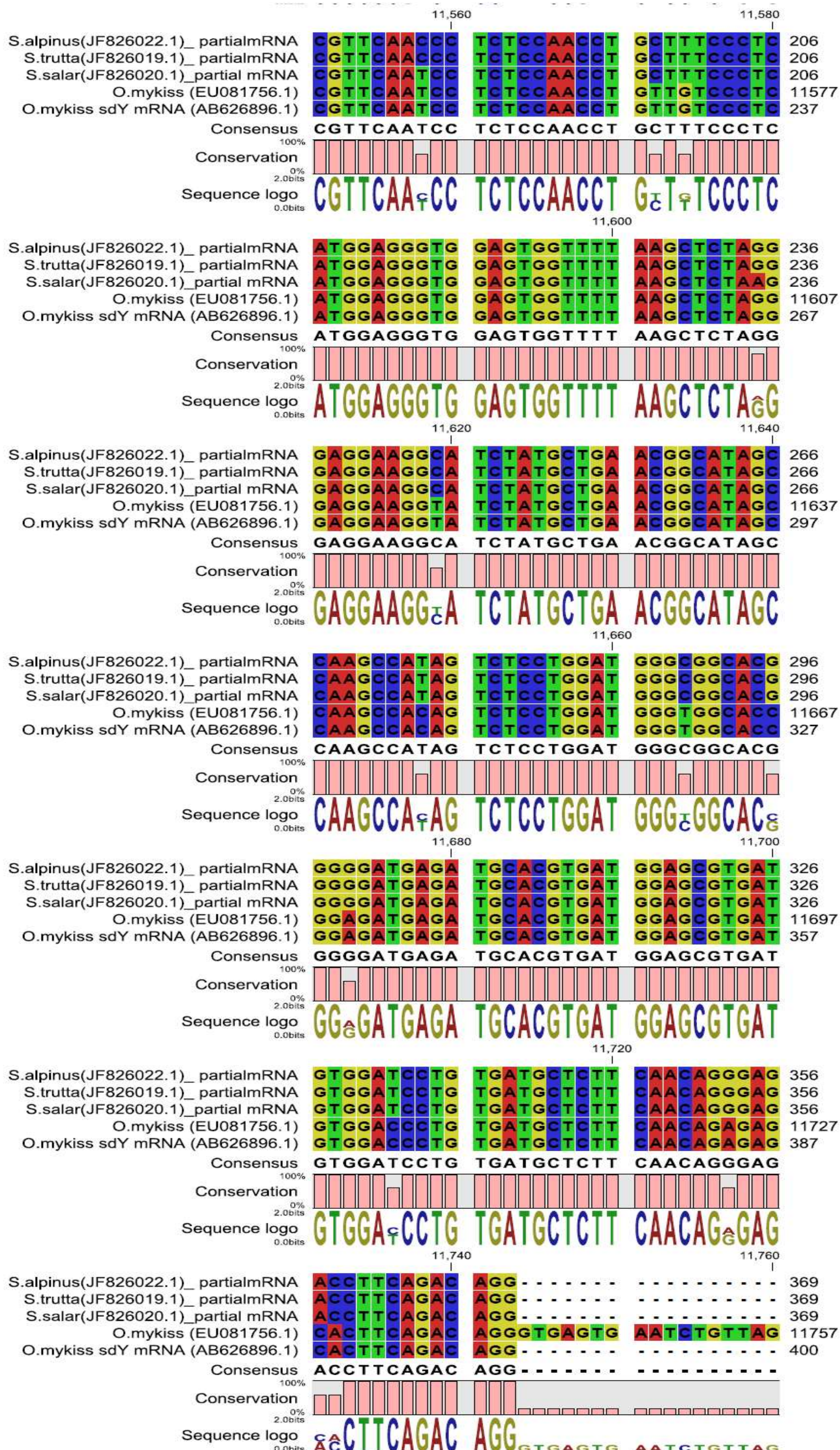
Appendix 1. primers for the direct sequencing of PCR products.

Name	Sequence (5' ⟶ 3')
M13 Reverse	CAGGAAACAGCTATGAC
T7 promoter	CCCTATAGTGAGTCGTATT
sdY 5' UTR F	AGAAACACTCCCCTTAACTAAT
sdY 3 UTR DegR	GKYTGAWTGTCTTCTAYTTAGTGTGAG
hiTAIL_ex2R_0	ATCCCATCCATCATCAGGGAGAAG
hiTAIL_ex2R_2	TCCTCTTCGCAGTACTCCACTGAGACAAG
sdY ex2R (212,230)	GCTTAAAACCACTCCACCC
hiTAIL_ex4F_0	ATTAAATTGCCATGGGCTCAGCAGC

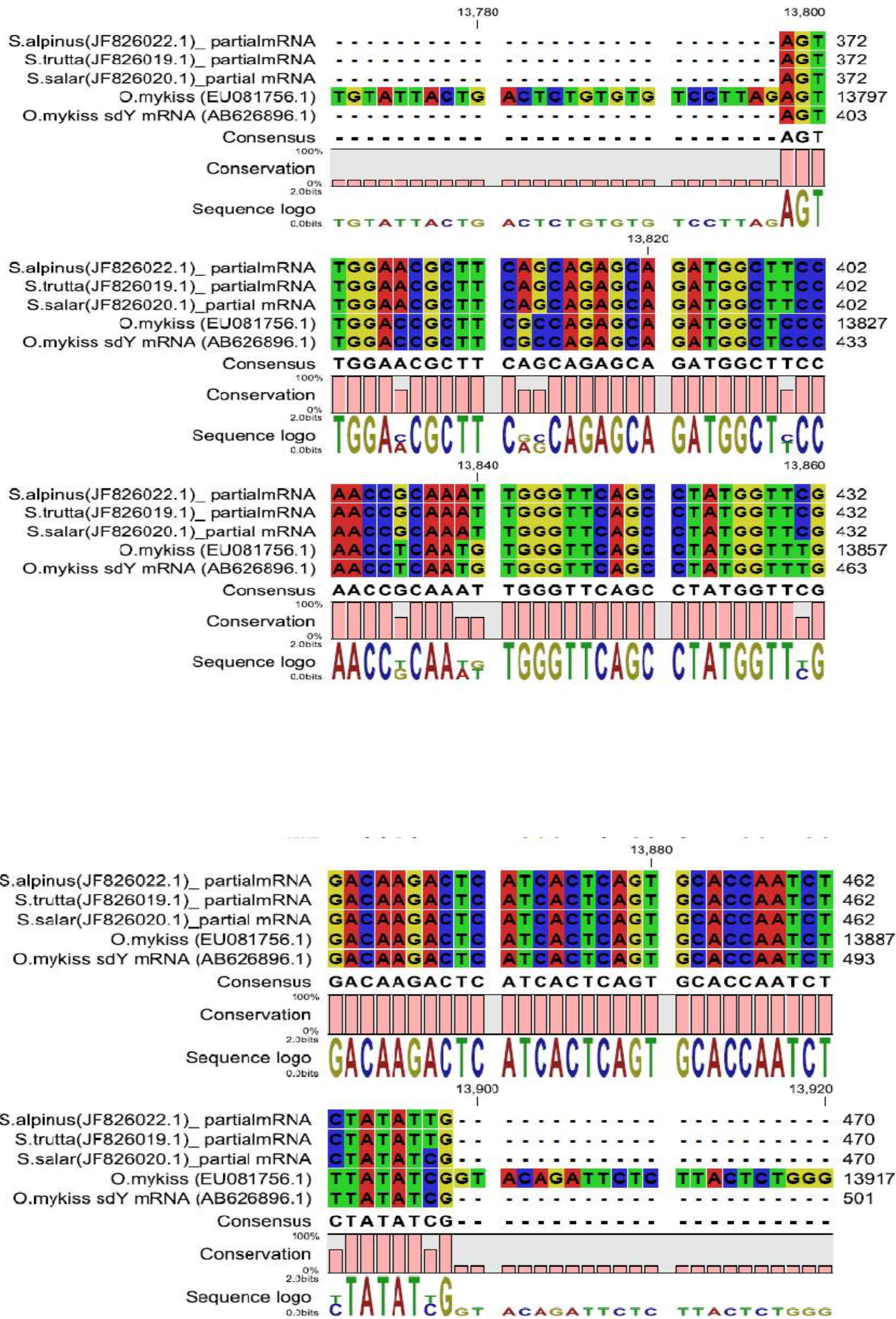
Appendix 2. Alignments of all partial mRNA with the complete genomic and mRNA *sdY* sequence of Rainbow trout revealed that these partial mRNA

Appendix 2.1. Exon 2 (starts from 11381 to 11740 of genomic *sdY* sequence of Rainbow trout)



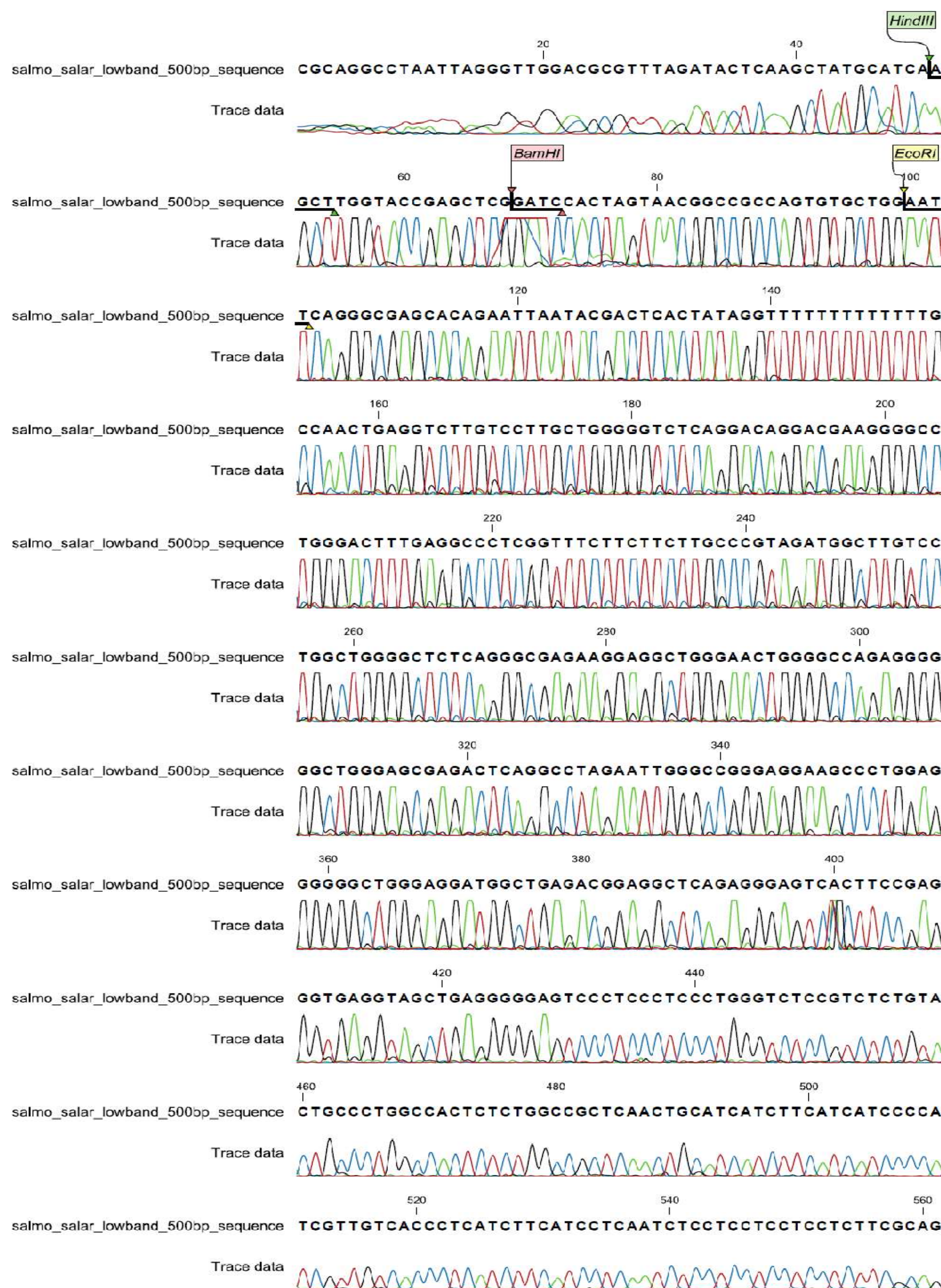

11,560
11,580
S.alpinus(JF826022.1)_ partialmRNA CGTTCAACCC TCTCCAACCT GCTTTCCCTC 206
S.trutta(JF826019.1)_ partialmRNA CGTTCAACCC TCTCCAACCT GCTTTCCCTC 206
S.salar(JF826020.1)_partial mRNA CGTTCAATCC TCTCCAACCT GCTTTCCCTC 206
O.mykiss (EU081756.1) CGTTCAATCC TCTCCAACCT GTTGTCCCTC 11577
O.mykiss sdY mRNA (AB626896.1) CGTTCAATCC TCTCCAACCT GTTGTCCCTC 237
Consensus CGTTCAATCC TCTCCAACCT GCTTTCCCTC
Conservation
Sequence logo
11,600
S.alpinus(JF826022.1)_ partialmRNA ATGGAGGGTG GAGTGGTTTT AAGCTCTAGG 236
S.trutta(JF826019.1)_ partialmRNA ATGGAGGGTG GAGTGGTTTT AAGCTCTAGG 236
S.salar(JF826020.1)_partial mRNA ATGGAGGGTG GAGTGGTTTT AAGCTCTAAG 236
O.mykiss (EU081756.1) ATGGAGGGTG GAGTGGTTTT AAGCTCTAGG 11607
O.mykiss sdY mRNA (AB626896.1) ATGGAGGGTG GAGTGGTTTT AAGCTCTAGG 267
Consensus ATGGAGGGTG GAGTGGTTTT AAGCTCTAGG
Conservation
Sequence logo
11,620
11,640
S.alpinus(JF826022.1)_ partialmRNA GAGGAAGGCA TCTATGCTGA ACGGCATAGC 266
S.trutta(JF826019.1)_ partialmRNA GAGGAAGGCA TCTATGCTGA ACGGCATAGC 266
S.salar(JF826020.1)_partial mRNA GAGGAAGGCA TCTATGCTGA ACGGCATAGC 266
O.mykiss (EU081756.1) GAGGAAGGTA TCTATGCTGA ACGGCATAGC 11637
O.mykiss sdY mRNA (AB626896.1) GAGGAAGGTA TCTATGCTGA ACGGCATAGC 297
Consensus GAGGAAGGCA TCTATGCTGA ACGGCATAGC
Conservation
Sequence logo
11,660
S.alpinus(JF826022.1)_ partialmRNA CAAGCCATAG TCTCCTGGAT GGGCGGCACG 296
S.trutta(JF826019.1)_ partialmRNA CAAGCCATAG TCTCCTGGAT GGGCGGCACG 296
S.salar(JF826020.1)_partial mRNA CAAGCCATAG TCTCCTGGAT GGGCGGCACG 296
O.mykiss (EU081756.1) CAAGCCACAG TCTCCTGGAT GGGTGGCACC 11667
O.mykiss sdY mRNA (AB626896.1) CAAGCCACAG TCTCCTGGAT GGGTGGCACC 327
Consensus CAAGCCATAG TCTCCTGGAT GGGCGGCACG
Conservation
Sequence logo
11,680
11,700
S.alpinus(JF826022.1)_ partialmRNA GGGGATGAGA TGCACGTGAT GGAGCGTGAT 326
S.trutta(JF826019.1)_ partialmRNA GGGGATGAGA TGCACGTGAT GGAGCGTGAT 326
S.salar(JF826020.1)_partial mRNA GGGGATGAGA TGCACGTGAT GGAGCGTGAT 326
O.mykiss (EU081756.1) GGAGATGAGA TGCACGTGAT GGAGCGTGAT 11697
O.mykiss sdY mRNA (AB626896.1) GGAGATGAGA TGCACGTGAT GGAGCGTGAT 357
Consensus GGGGATGAGA TGCACGTGAT GGAGCGTGAT
Conservation
Sequence logo
11,720
S.alpinus(JF826022.1)_ partialmRNA GTGGATCCTG TGATGCTCTT CAACAGGGAG 356
S.trutta(JF826019.1)_ partialmRNA GTGGATCCTG TGATGCTCTT CAACAGGGAG 356
S.salar(JF826020.1)_partial mRNA GTGGATCCTG TGATGCTCTT CAACAGGGAG 356
O.mykiss (EU081756.1) GTGGACCCTG TGATGCTCTT CAACAGAGAG 11727
O.mykiss sdY mRNA (AB626896.1) GTGGACCCTG TGATGCTCTT CAACAGAGAG 387
Consensus GTGGATCCTG TGATGCTCTT CAACAGGGAG
Conservation
Sequence logo
11,740
11,760
S.alpinus(JF826022.1)_ partialmRNA ACCTTCAGAC AGG------- ---------- 369
S.trutta(JF826019.1)_ partialmRNA ACCTTCAGAC AGG------- ---------- 369
S.salar(JF826020.1)_partial mRNA ACCTTCAGAC AGG------- ---------- 369
O.mykiss (EU081756.1) CACTTCAGAC AGGGTGAGTG AATCTGTTAG 11757
O.mykiss sdY mRNA (AB626896.1) CACTTCAGAC AGG------- ---------- 400
Consensus ACCTTCAGAC AGG------- ----------
Conservation
Sequence logo

Appendix 2.2 Exon 3(starts from 13795 to 13895 of genomic *sdY* sequence)

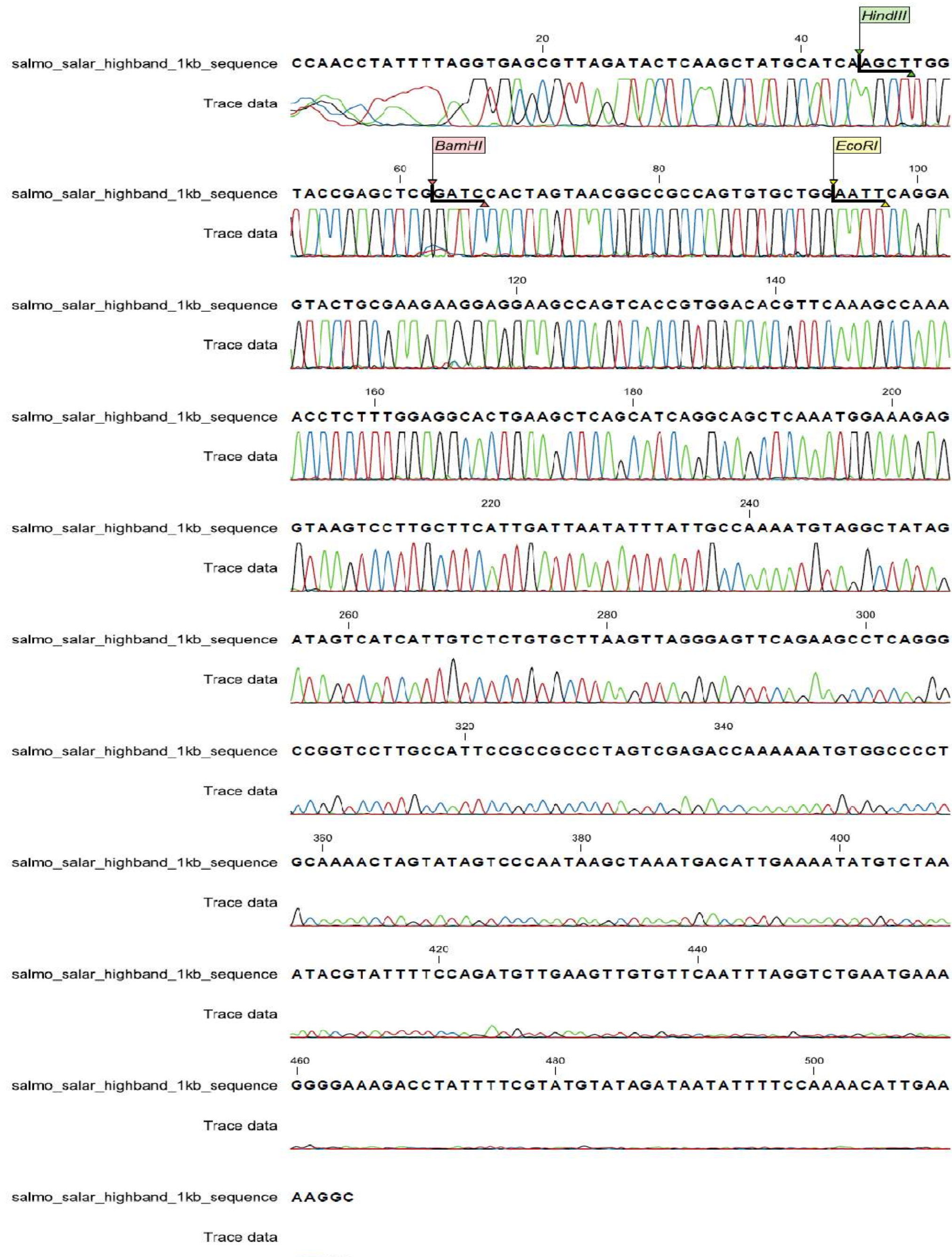


Appendix 3. Sequence of *salmo salar, Salmo trutta,*and *Salvelinus alpinus*.

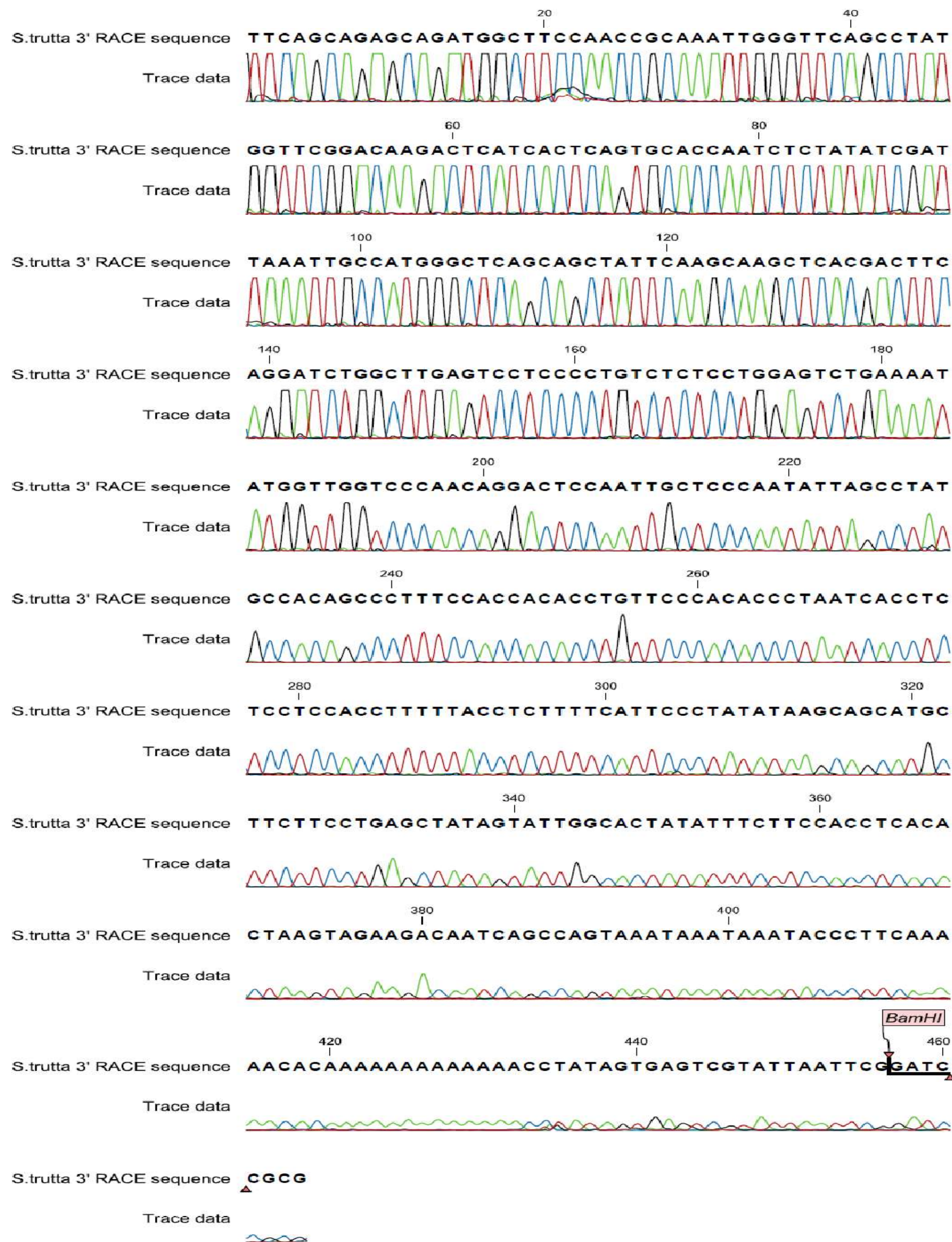
Appendix 3.1 Sequence obtained for *Salmo salar*, the smaller (500bp) bands is PREDICTED: Salmo salar RAD54-like 2 (S. cerevisiae) (rad54l2), transcript variant X3, mRNA



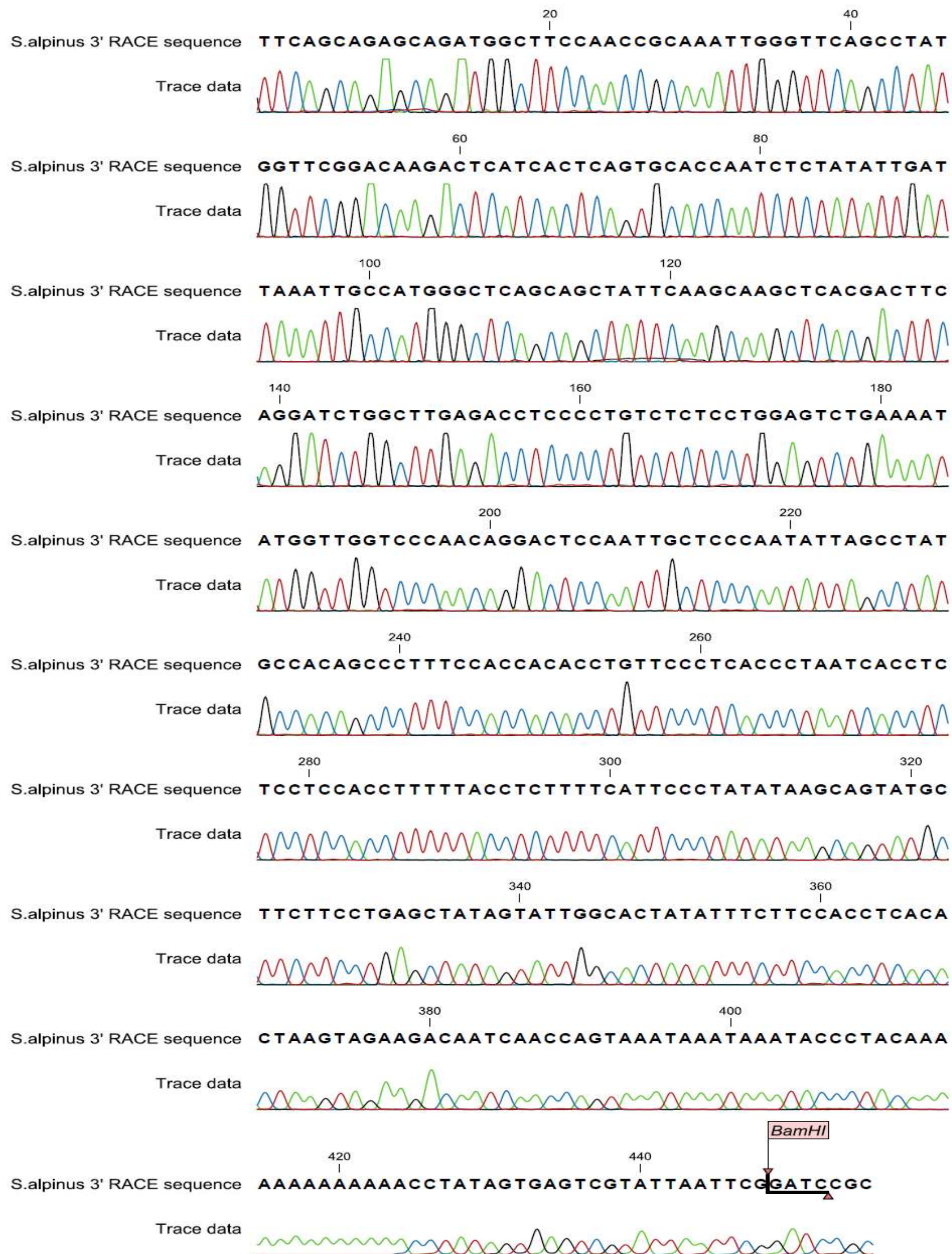
Appendix 3.2 Sequence obtained for *Salmo salar* longer (1 kb) bands is PREDICTED: Salmo salar tubulin polyglutamylase TTLL11-like (LOC106562979), mRNA



Appendix 3.3 3' RACE sequence *Salmo trutta* (464 bp)

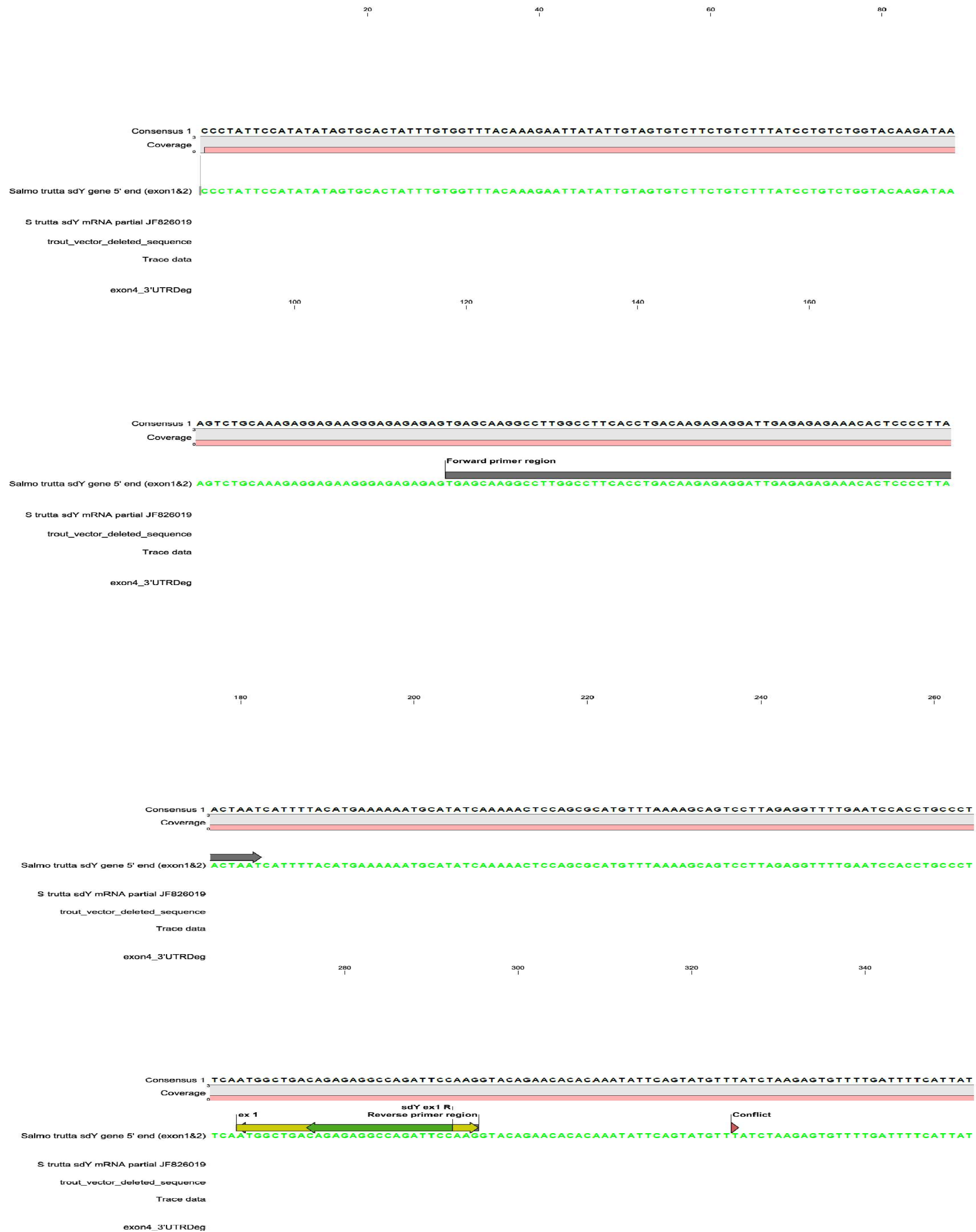


Appendix 3.4 3' RACE sequence *Salvelinus alpinus* (455 bp)

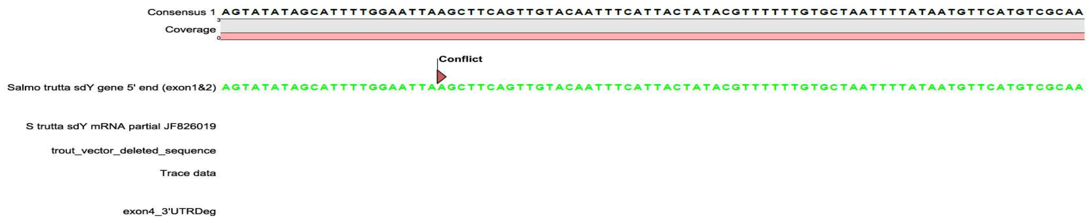


Appendix 4. Assemble results

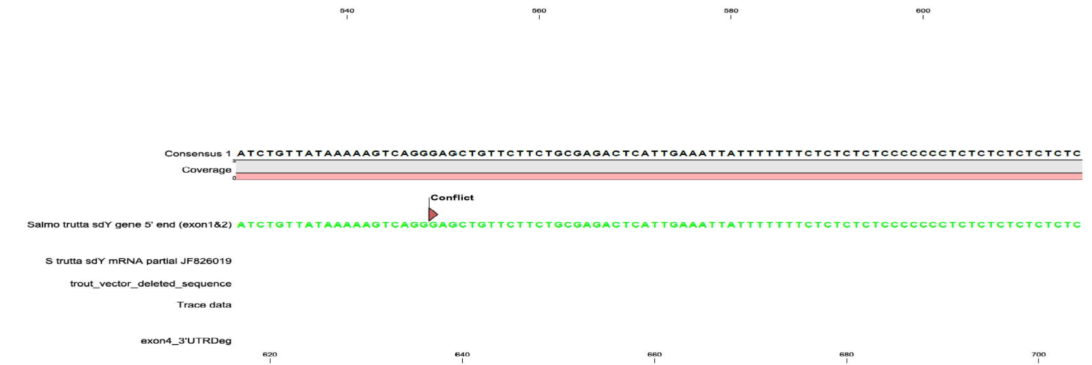
Appendix 4.1. 5' end, exon 1, intron 1. exon 2 ,exon 3, exon 4 and 3' UTR for *Salmo trutta*



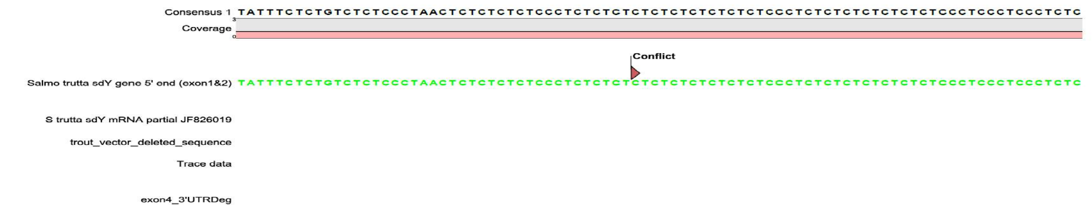
360
380
400
420
440
Consensus 1 TTTTGGGCCCTATGAATTTCTGATGTTAAGGACTGTGTCTATGTGGTTATTTCCCATGTCTGATCTTTTGAGGGAAATGCTCAAAATG
Coverage
Conflict
Conflict
Conflict
Salmo trutta sdY gene 5' end (exon1&2) TTTTGGGCCCTATGAATTTCTGATGTTAAGGACTGTGTCTATGTGGTTATTTCCCATGTCTGATCTTTTGAGGGAAATGCTCAAAATG
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
460
480
500
520



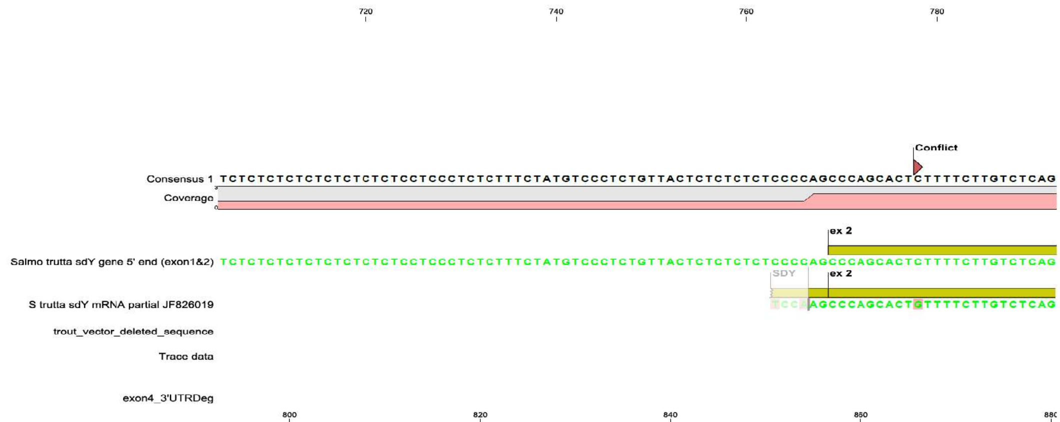
Consensus 1 AGTATATAGCATTTTGGAATTAAGCTTCAGTTGTACAATTTCATTACTATACGTTTTTTGTGCTAATTTTATAATGTTCATGTCGCAA
Coverage
Conflict
Salmo trutta sdY gene 5' end (exon1&2) AGTATATAGCATTTTGGAATTAAGCTTCAGTTGTACAATTTCATTACTATACGTTTTTTGTGCTAATTTTATAATGTTCATGTCGCAA
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg

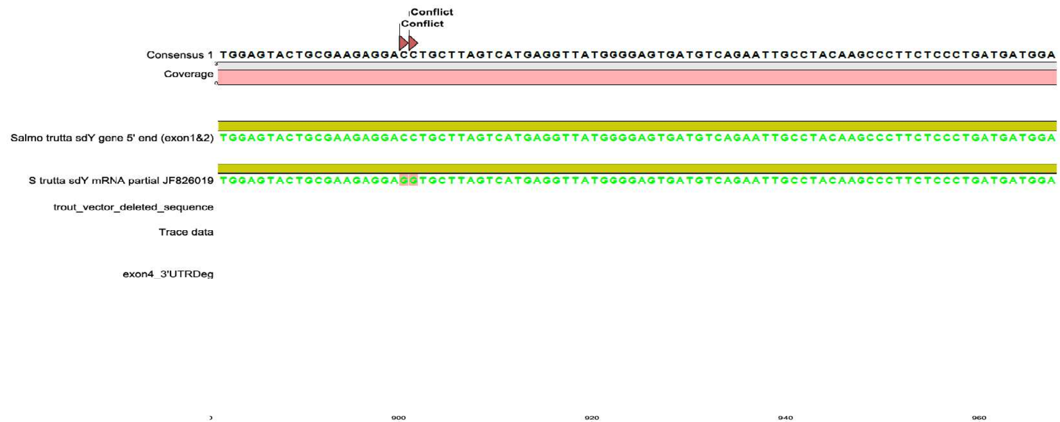


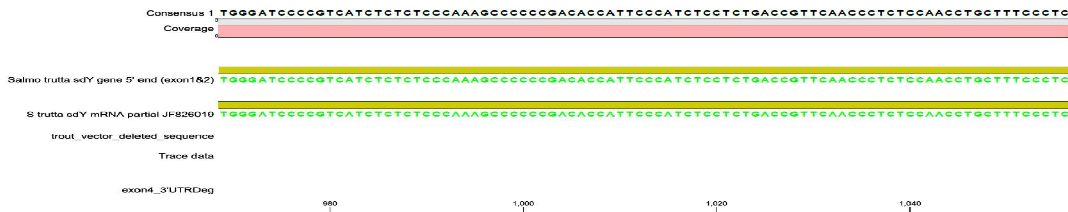
540
560
580
600
Consensus 1 ATCTGTTATAAAAAGTCAGGGAGCTGTTCTTCTGCGAGACTCATTGAAATTATTTTTTTCTCTCTCTCCCCCCCTCTCTCTCTCTCTC
Coverage
Conflict
Salmo trutta sdY gene 5' end (exon1&2) ATCTGTTATAAAAAGTCAGGGAGCTGTTCTTCTGCGAGACTCATTGAAATTATTTTTTTCTCTCTCTCCCCCCCTCTCTCTCTCTCTC
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
620
640
660
680
700

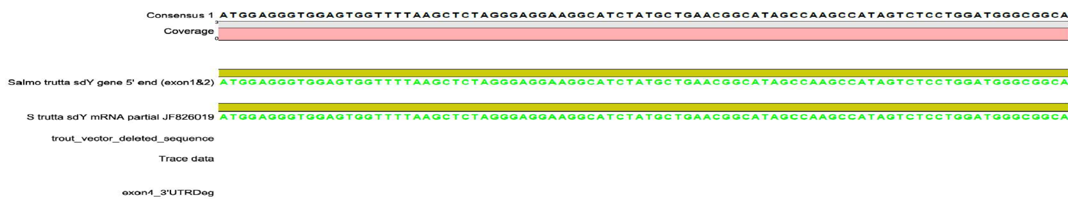


Consensus 1 TATTTCTCTGTCTCTCCCTAACTCTCTCTCTCCCTCTCTCTCTCTCTCTCTCTCTCCCTCTCTCTCTCTCTCTCCCTCCCTCCCTCTC
Coverage
Conflict
Salmo trutta sdY gene 5' end (exon1&2) TATTTCTCTGTCTCTCCCTAACTCTCTCTCTCCCTCTCTCTCTCTCTCTCTCTCTCCCTCTCTCTCTCTCTCTCCCTCCCTCCCTCTC
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg

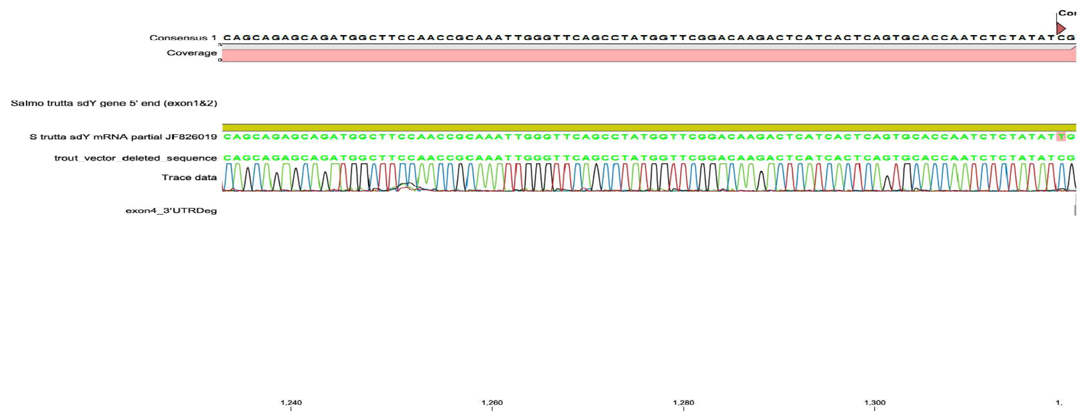

720
740
760
780
Conflict
Consensus 1 TCTCTCTCTCTCTCTCTCTCCTCCCTCTCTTTCTATGTCCCTCTGTTACTCTCTCTCTCCCCAGCCCAGCACTCTTTTCTTGTCTCAG
Coverage
ex 2
Salmo trutta sdY gene 5' end (exon1&2) TCTCTCTCTCTCTCTCTCTCCTCCCTCTCTTTCTATGTCCCTCTGTTACTCTCTCTCTCCCCAGCCCAGCACTCTTTTCTTGTCTCAG
SDY
ex 2
S trutta sdY mRNA partial JF826019 TCCAAGCCCAGCACTGTTTTCTTGTCTCAG
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg


800
820
840
860
880
Conflict
Conflict
Consensus 1 TGGAGTACTGCGAAGAGGACCTGCTTAGTCATGAGGTTATGGGGAGTGATGTCAGAATTGCCTACAAGCCCTTCTCCCTGATGATGGA
Coverage
Salmo trutta sdY gene 5' end (exon1&2) TGGAGTACTGCGAAGAGGACCTGCTTAGTCATGAGGTTATGGGGAGTGATGTCAGAATTGCCTACAAGCCCTTCTCCCTGATGATGGA
S trutta sdY mRNA partial JF826019 TGGAGTACTGCGAAGAGGAGGTGCTTAGTCATGAGGTTATGGGGAGTGATGTCAGAATTGCCTACAAGCCCTTCTCCCTGATGATGGA
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg

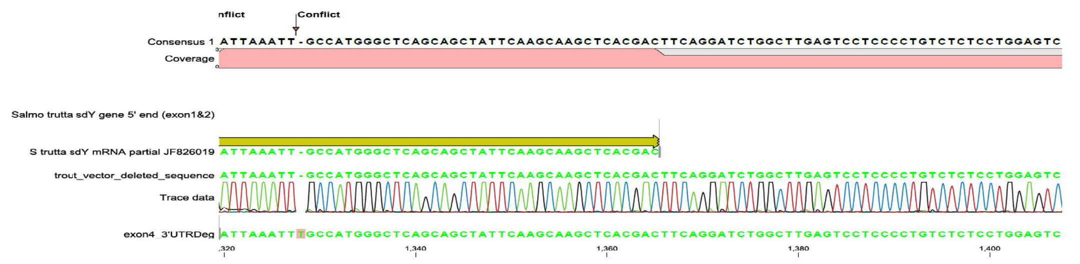

900
920
940
960
Consensus 1 TGGGATCCCCGTCATCTCTCTCCCAAAGCCCCCCGACACCATTCCCATCTCCTCTGACCGTTCAACCCTCTCCAACCTGCTTTCCCTC
Coverage
Salmo trutta sdY gene 5' end (exon1&2) TGGGATCCCCGTCATCTCTCTCCCAAAGCCCCCCGACACCATTCCCATCTCCTCTGACCGTTCAACCCTCTCCAACCTGCTTTCCCTC
S trutta sdY mRNA partial JF826019 TGGGATCCCCGTCATCTCTCTCCCAAAGCCCCCCGACACCATTCCCATCTCCTCTGACCGTTCAACCCTCTCCAACCTGCTTTCCCTC
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
980
1,000
1,020
1,040


Consensus 1 ATGGAGGGTGGAGTGGTTTTAAGCTCTAGGGAGGAAGGCATCTATGCTGAACGGCATAGCCAAGCCATAGTCTCCTGGATGGGCGGCA
Coverage
Salmo trutta sdY gene 5' end (exon1&2) ATGGAGGGTGGAGTGGTTTTAAGCTCTAGGGAGGAAGGCATCTATGCTGAACGGCATAGCCAAGCCATAGTCTCCTGGATGGGCGGCA
S trutta sdY mRNA partial JF826019 ATGGAGGGTGGAGTGGTTTTAAGCTCTAGGGAGGAAGGCATCTATGCTGAACGGCATAGCCAAGCCATAGTCTCCTGGATGGGCGGCA
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg

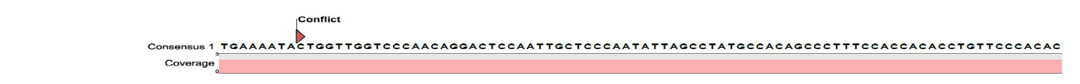
1,060
1,080
1,100
1,120
1,140
Consensus 1
Coverage
Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
1,160
1,180
1,200
1,220



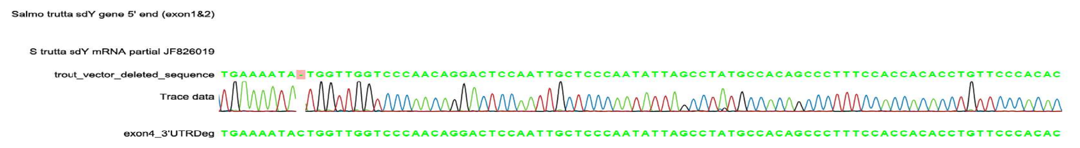
Consensus 1
Coverage
Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
1,240
1,260
1,280
1,300



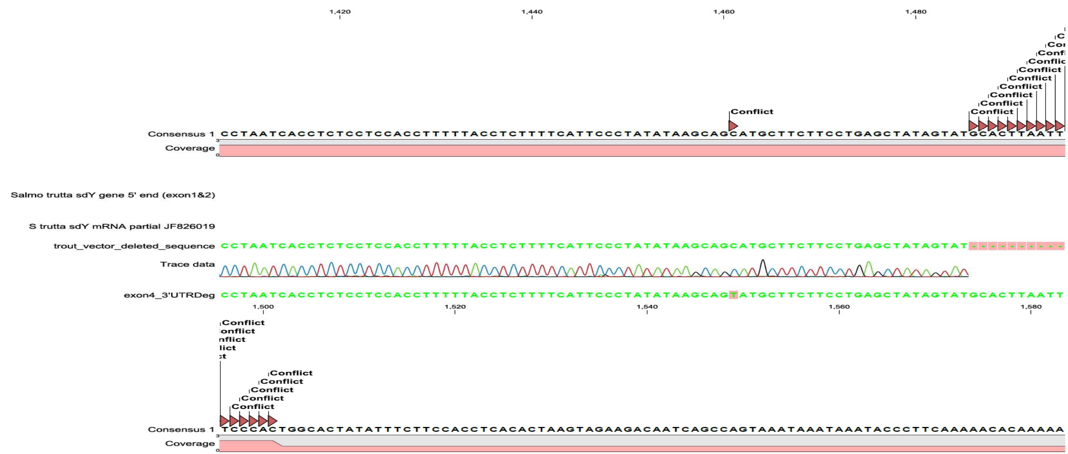
Conflict
Consensus 1
Coverage
Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg
1,340
1,360
1,380
1,400

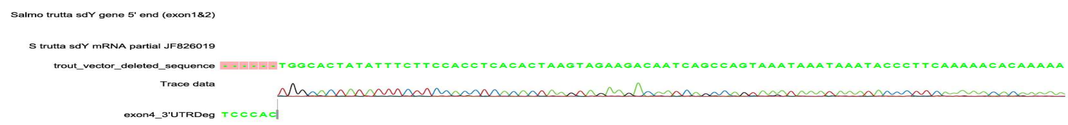


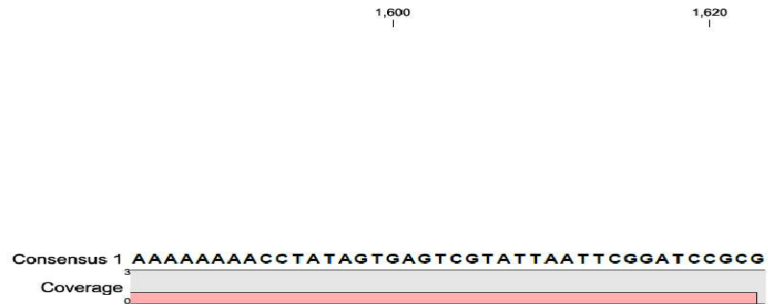
Conflict
Consensus 1
Coverage



Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence
Trace data
exon4_3'UTRDeg

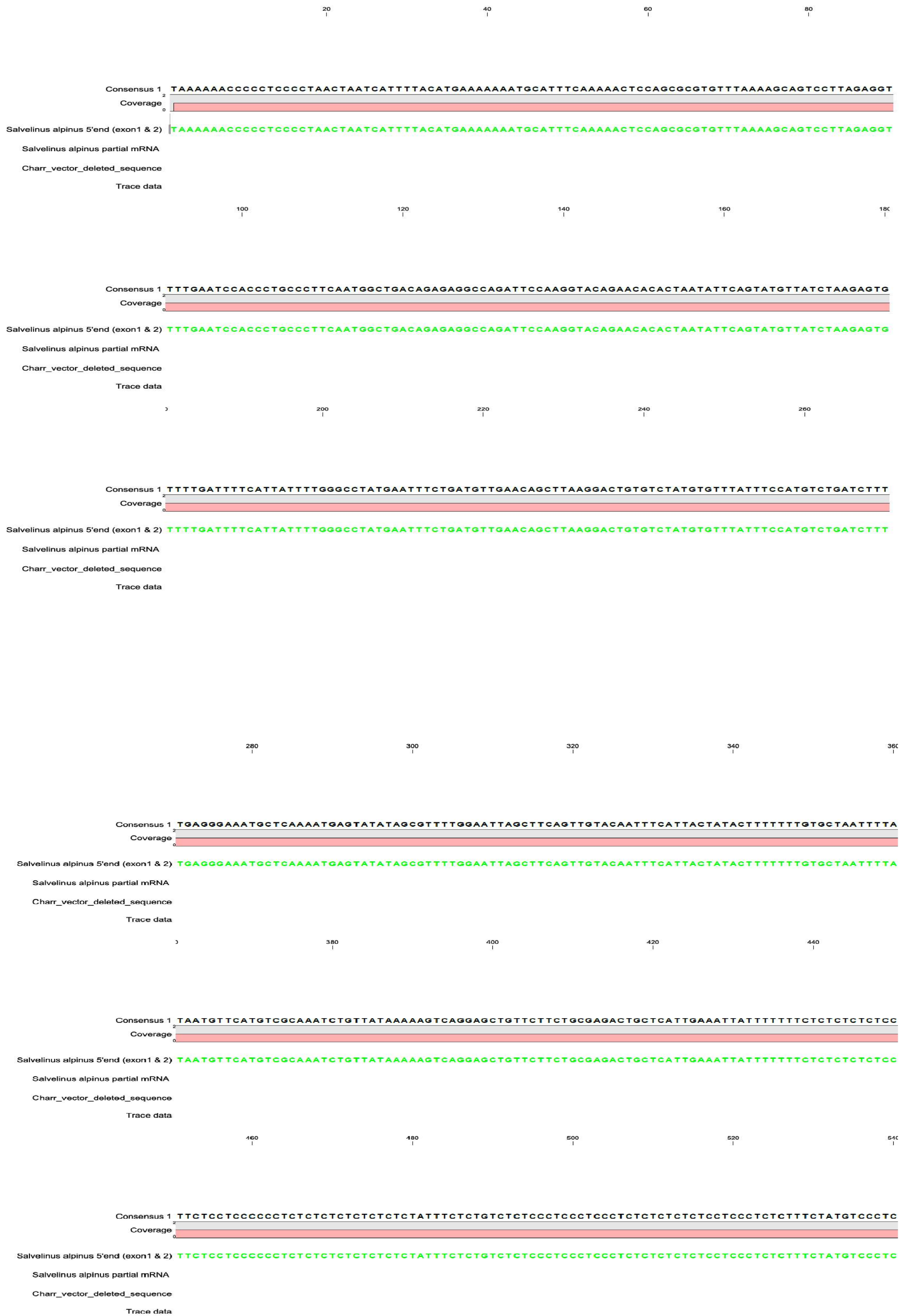

1,420
1,440
1,460
1,480
Conflict
Consensus 1 CCTAATCACCTCTCCTCCACCTTTTTACCTCTTTTCATTCCCTATATAAGCAGCATGCTTCTTCCTGAGCTATAGTATGCACTTAATT
Coverage
Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence CCTAATCACCTCTCCTCCACCTTTTTACCTCTTTTCATTCCCTATATAAGCAGCATGCTTCTTCCTGAGCTATAGTAT
Trace data
exon4_3'UTRDeg CCTAATCACCTCTCCTCCACCTTTTTACCTCTTTTCATTCCCTATATAAGCAGTATGCTTCTTCCTGAGCTATAGTATGCACTTAATT
1,500
1,520
1,540
1,560
1,580
Conflict
Consensus 1 TCCCACTGGCACTATATTTCTTCCACCTCACACTAAGTAGAAGACAATCAGCCAGTAAATAAATAAATACCCTTCAAAAACACAAAAA
Coverage

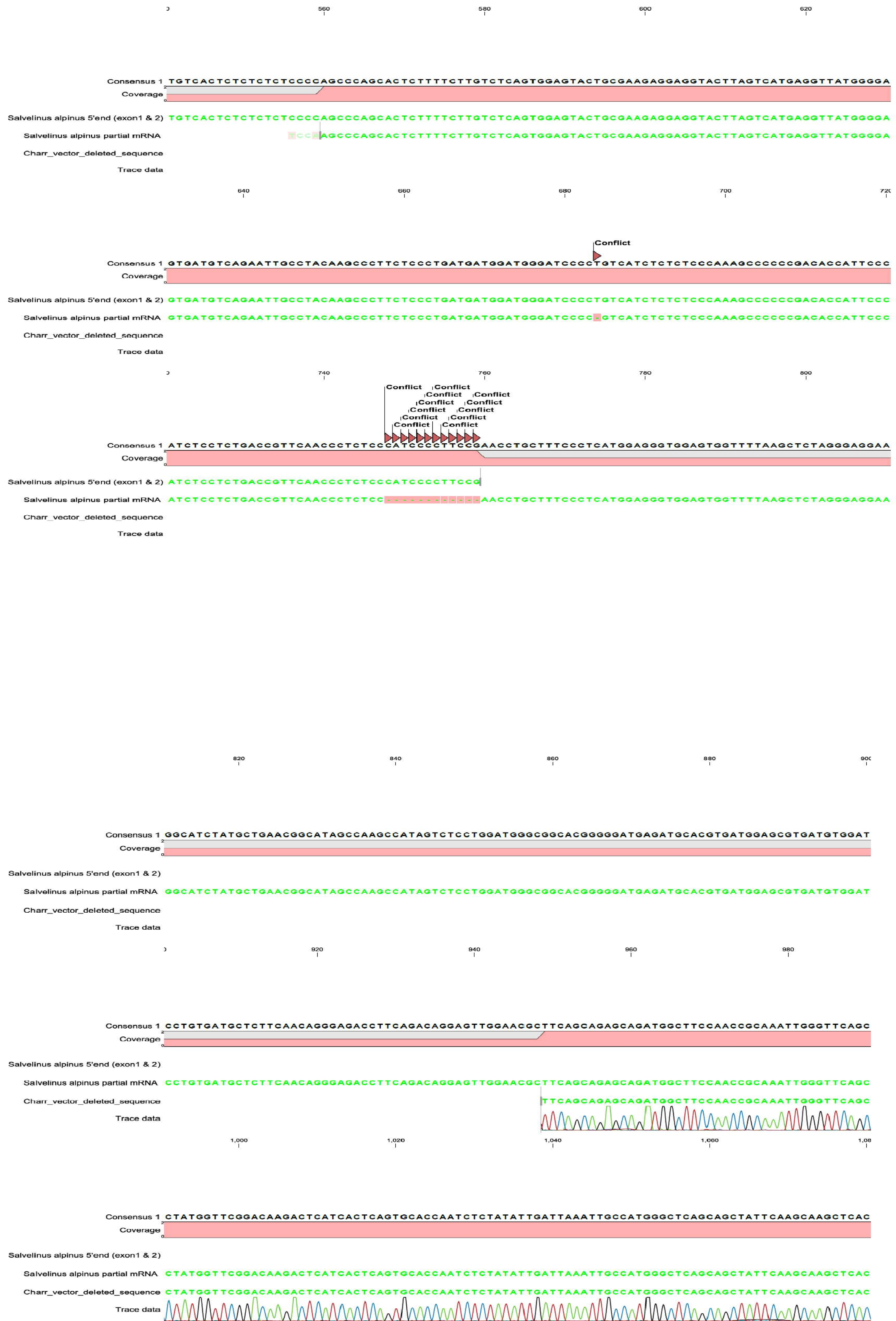

Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence TGGCACTATATTTCTTCCACCTCACACTAAGTAGAAGACAATCAGCCAGTAAATAAATAAATACCCTTCAAAAACACAAAAA
Trace data
exon4_3'UTRDeg TCCCAC

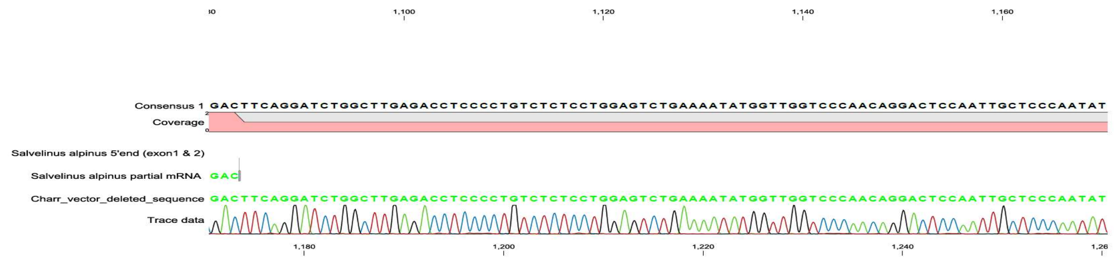

1,600
1,620
Consensus 1 AAAAAAAACCTATAGTGAGTCGTATTAATTCGGATCCGCG
Coverage

Salmo trutta sdY gene 5' end (exon1&2)
S trutta sdY mRNA partial JF826019
trout_vector_deleted_sequence AAAAAAAACCTATAGTGAGTCGTATTAATTCGGATCCGCG
Trace data
exon4_3'UTRDeg

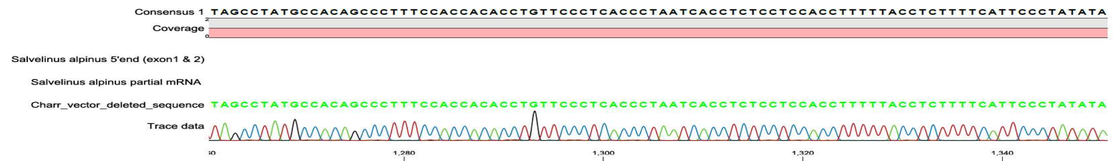
Appendix 4.2. 5' end, exon 1, intron 1, exon 2, exon 3 and exon 4 for *Salvelinus alpinus*



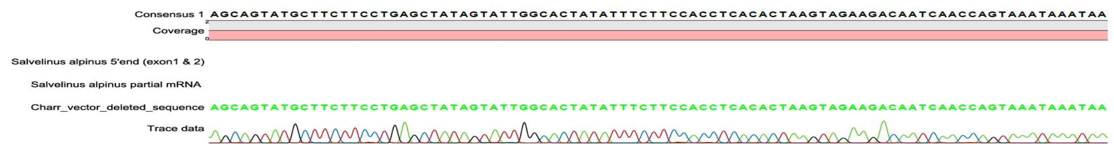

Consensus 1
Coverage
Salvelinus alpinus 5'end (exon1 & 2)
Salvelinus alpinus partial mRNA
Charr_vector_deleted_sequence
Trace data
Conflict



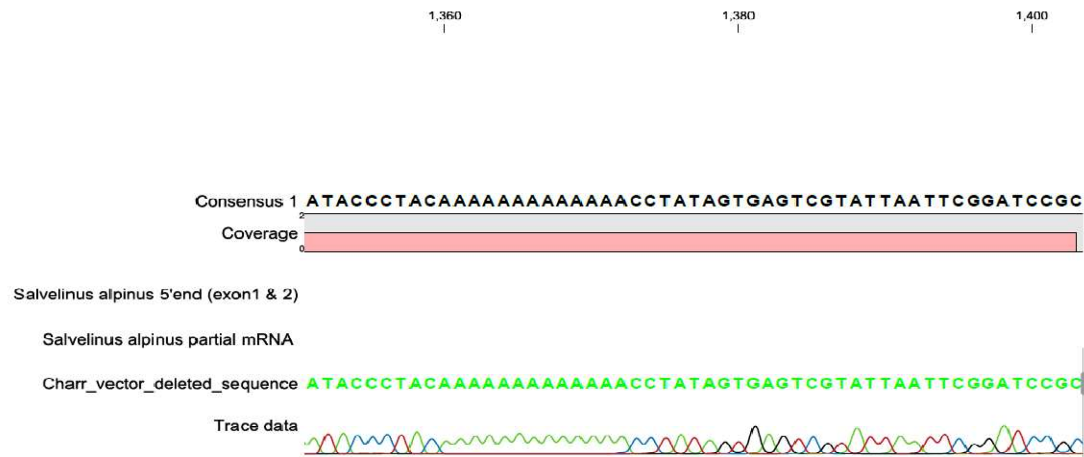
1,100
1,120
1,140
1,160
Consensus 1 GACTTCAGGATCTGGCTTGAGACCTCCCCTGTCTCTCCTGGAGTCTGAAAATATGGTTGGTCCCAACAGGACTCCAATTGCTCCCAATAT
Coverage
Salvelinus alpinus 5'end (exon1 & 2)
Salvelinus alpinus partial mRNA GAC
Charr_vector_deleted_sequence GACTTCAGGATCTGGCTTGAGACCTCCCCTGTCTCTCCTGGAGTCTGAAAATATGGTTGGTCCCAACAGGACTCCAATTGCTCCCAATAT
Trace data
1,180
1,200
1,220
1,240



Consensus 1 TAGCCTATGCCACAGCCCTTTCCACCACACCTGTTCCCTCACCCTAATCACCTCTCCTCCACCTTTTTACCTCTTTTCATTCCCTATATA
Coverage
Salvelinus alpinus 5'end (exon1 & 2)
Salvelinus alpinus partial mRNA
Charr_vector_deleted_sequence TAGCCTATGCCACAGCCCTTTCCACCACACCTGTTCCCTCACCCTAATCACCTCTCCTCCACCTTTTTACCTCTTTTCATTCCCTATATA
Trace data
1,280
1,300
1,320
1,340

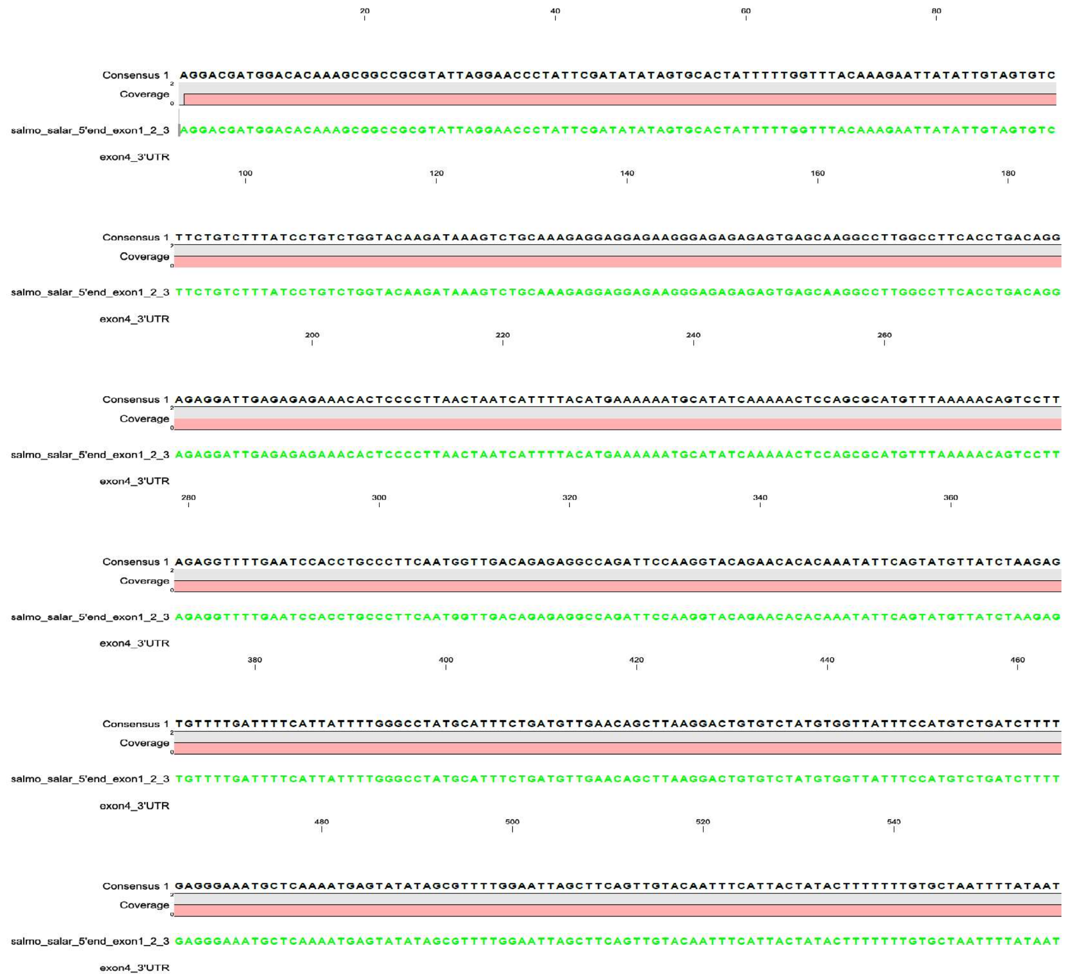


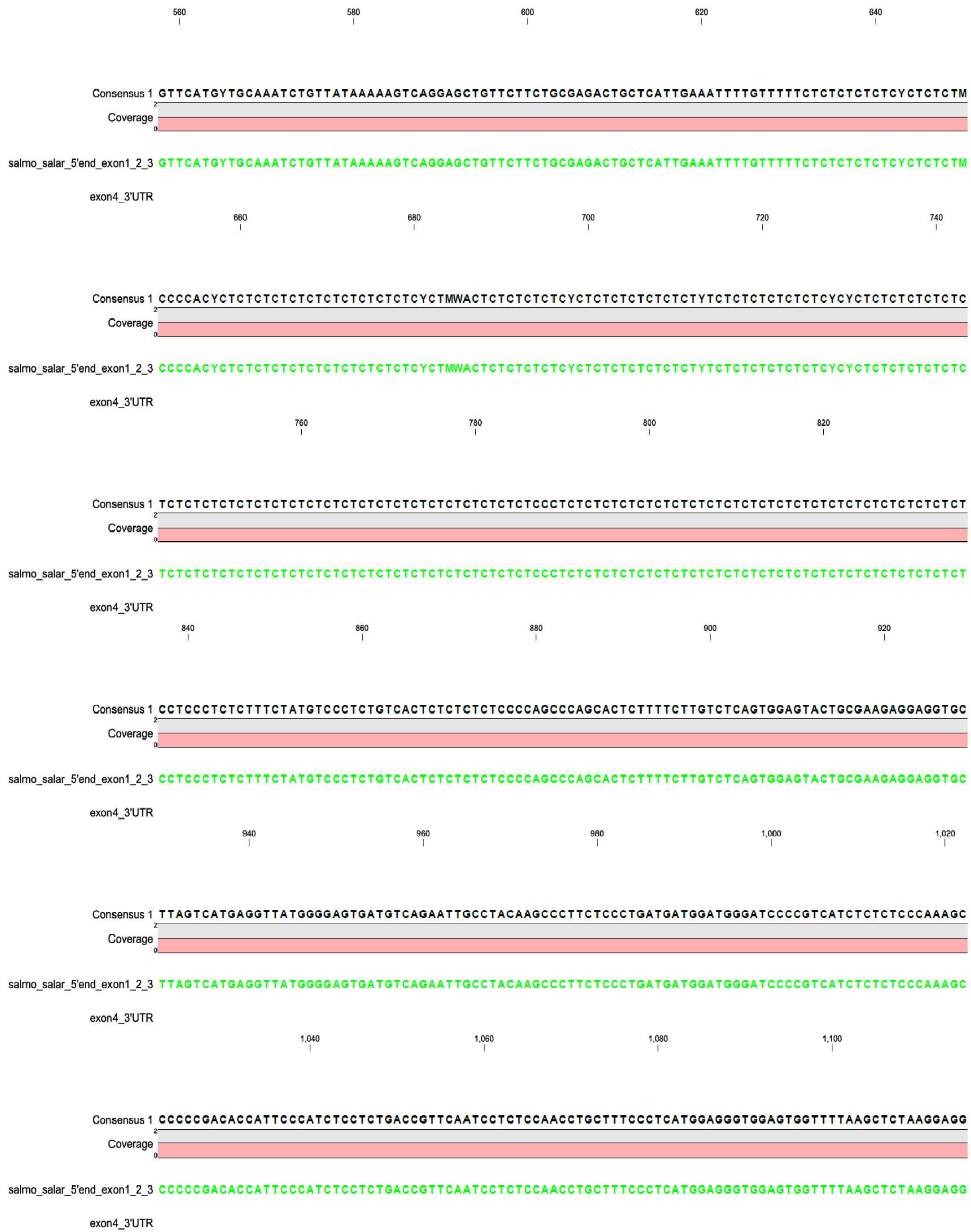
Consensus 1 AGCAGTATGCTTCTTCCTGAGCTATAGTATTGGCACTATATTTCTTCCACCTCACACTAAGTAGAAGACAATCAACCAGTAAATAAATAA
Coverage
Salvelinus alpinus 5'end (exon1 & 2)
Salvelinus alpinus partial mRNA
Charr_vector_deleted_sequence AGCAGTATGCTTCTTCCTGAGCTATAGTATTGGCACTATATTTCTTCCACCTCACACTAAGTAGAAGACAATCAACCAGTAAATAAATAA
Trace data



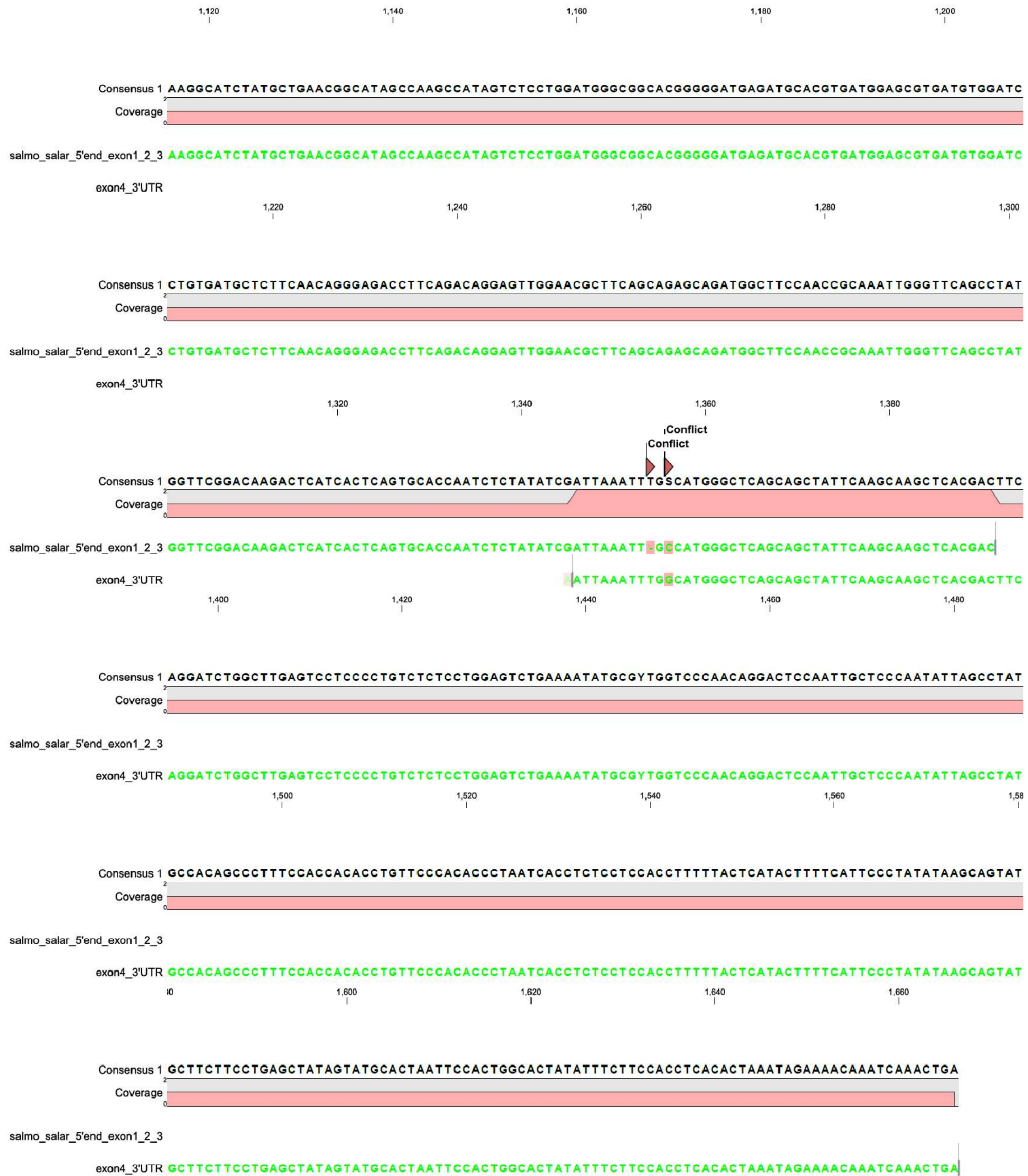
1,360
1,380
1,400
Consensus 1 ATACCCTACAAAAAAAAAAAAAACCTATAGTGAGTCGTATTAATTCGGATCCGC
Coverage
Salvelinus alpinus 5'end (exon1 & 2)
Salvelinus alpinus partial mRNA
Charr_vector_deleted_sequence ATACCCTACAAAAAAAAAAAAAACCTATAGTGAGTCGTATTAATTCGGATCCGC
Trace data

Appendix 4.3. 5' end with exon1, intron1, exon 2 ,exon 3, exon 4 and 3' UTR for *Salmo salar*





Consensus 1
Coverage
salmo_salar_5'end_exon1_2_3
exon4_3'UTR
560
580
600
620
640
660
680
700
720
740
760
780
800
820
840
860
880
900
920
940
960
980
1,000
1,020
1,040
1,060
1,080
1,100


1,120
1,140
1,160
1,180
1,200
Consensus 1 AAGGCATCTATGCTGAACGGCATAGCCAAGCCATAGTCTCCTGGATGGGCGGCACGGGGGATGAGATGCACGTGATGGAGCGTGATGTGGATC
Coverage
salmo_salar_5'end_exon1_2_3 AAGGCATCTATGCTGAACGGCATAGCCAAGCCATAGTCTCCTGGATGGGCGGCACGGGGGATGAGATGCACGTGATGGAGCGTGATGTGGATC
exon4_3'UTR
1,220
1,240
1,260
1,280
1,300
Consensus 1 CTGTGATGCTCTTCAACAGGGAGACCTTCAGACAGGAGTTGGAACGCTTCAGCAGAGCAGATGGCTTCCAACCGCAAATTGGGTTCAGCCTAT
Coverage
salmo_salar_5'end_exon1_2_3 CTGTGATGCTCTTCAACAGGGAGACCTTCAGACAGGAGTTGGAACGCTTCAGCAGAGCAGATGGCTTCCAACCGCAAATTGGGTTCAGCCTAT
exon4_3'UTR
1,320
1,340
1,360
1,380
Conflict
Conflict
Consensus 1 GGTTCGGACAAGACTCATCACTCAGTGCACCAATCTCTATATCGATTAAATTTGSCATGGGCTCAGCAGCTATTCAAGCAAGCTCACGACTTC
Coverage
salmo_salar_5'end_exon1_2_3 GGTTCGGACAAGACTCATCACTCAGTGCACCAATCTCTATATCGATTAAATT-GCCATGGGCTCAGCAGCTATTCAAGCAAGCTCACGAC
exon4_3'UTR ATTAAATTTGGCATGGGCTCAGCAGCTATTCAAGCAAGCTCACGACTTC
1,400
1,420
1,440
1,460
1,480
Consensus 1 AGGATCTGGCTTGAGTCCTCCCCTGTCTCTCCTGGAGTCTGAAAATATGCGYTGGTCCCAACAGGACTCCAATTGCTCCCAATATTAGCCTAT
Coverage
salmo_salar_5'end_exon1_2_3
exon4_3'UTR AGGATCTGGCTTGAGTCCTCCCCTGTCTCTCCTGGAGTCTGAAAATATGCGYTGGTCCCAACAGGACTCCAATTGCTCCCAATATTAGCCTAT
1,500
1,520
1,540
1,560
Consensus 1 GCCACAGCCCTTTCCACCACACCTGTTCCCACACCCTAATCACCTCTCCTCCACCTTTTTACTCATACTTTTCATTCCCTATATAAGCAGTAT
Coverage
salmo_salar_5'end_exon1_2_3
exon4_3'UTR GCCACAGCCCTTTCCACCACACCTGTTCCCACACCCTAATCACCTCTCCTCCACCTTTTTACTCATACTTTTCATTCCCTATATAAGCAGTAT
1,600
1,620
1,640
1,660
Consensus 1 GCTTCTTCCTGAGCTATAGTATGCACTAATTCCACTGGCACTATATTTCTTCCACCTCACACTAAATAGAAAACAAATCAAACTGA
Coverage
salmo_salar_5'end_exon1_2_3
exon4_3'UTR GCTTCTTCCTGAGCTATAGTATGCACTAATTCCACTGGCACTATATTTCTTCCACCTCACACTAAATAGAAAACAAATCAAACTGA

Appendix 5. list of primers that was design to identify male samples.

Name	Sequence (5' - 3')
hiTAIL_ex3F_0	GCAGAGCAGATGGCTTCCAAC
hiTAIL_ex3F_1	ACGATGGACTCCAGTCCGGCCCCTATGGTTCGGACAAGACTCATCA CTC
hiTAIL_ex3F_2	TCAGCAGCTATTCAAGCAAGCTCACGAC
hiTAIL_ex4F_0	ATTAAATTGCCATGGGCTCAGCAGC
hiTAIL_ex4F_1	ACGATGGACTCCAGTCCGGCCACTTCAGGATCTGGCTTGAGACCTC
hiTAIL_ex4F_2	CTCCCCTGTCTCTCCCGGA
hiTAIL_ex4F2	ATTCAAGCAAGCTCACGAC
sdY ex4F	AGTTGGAACGCTTCAGCAGAGCAGATGG
sdY in1F	GCTGTTCTTCTGCGAGACT
sdY5' UTR F	AGAAACACTCCCCTTAACTAAT
sdY ex4 F	AGTTGGAACGCTTCAGCAGAGCAGATGG
sdY ex4 R	GACAAGACTCATCACTCAGTGCACCAATCT
sdY ex4 R (corrected)	AGATTGGTGCACTGAGTGATGAGTCTTGTCC
sdY ex2 R2	ATCACTCCCCATAACCTCATGACTAAG
sdY ex1R	GGAATCTGGCCTCTCTG
sdY ex4 F2	ATTCAAGCAAGCTCACGAC
sdY 3'UTR R2	AAAAGAGGTAAAAAGGTGGAGG
sdY 3UTR Deg R	GKYTGAWTGTCTTCTAYTTAGTGTGAG

hiTAIL_ex2R_0	ATCCCATCCATCATCAGGGAGAAG
hiTAIL_ex2R_1	ACGATGGACTCCAGTCCGGCCGGGCTTGTAGGCAATTCTGACATCA CTC
hiTAIL_ex2R_2	TCCTCTTCGCAGTACTCCACTGAGACAAG
sdY ex2 R (correct)	TAGAGCTTAAAACCACTCCACCCTCCATGAGGGA