Onchorhynchus mykiss pax7 sequence variations with comparative analyses against other teleost species

Kalyan C Chapalamadugu, Brenda M Murdoch, Barrie D Robison, Rodney A Hill and Gordon K Murdoch

Abstract
The paired box-7 (pax7) transcription factor expressed in satellite cells (SCs) is an essential regulator of skeletal muscle growth and regeneration in vertebrates including fish. Characterization of rainbow trout (Onchorhynchus mykiss) pax7 gene/s may offer novel insights into skeletal myogenesis by SCs in this indeterminate growth species. Further, evaluation of promoters for cis-regulatory regions may shed light on the evolutionary fate of the duplicated genes. Employing standard PCR, cloning and computational approach, we identified and report complete coding sequences of two pax7 paralogs of rainbow trout (rt); rt\textit{pax7\textalpha} and rt\textit{pax7\textbeta}. Both genes show significant identity in the nucleotide (97%) and the predicted amino acid (98%) sequences, and bear the characteristic paired domain (PD), octapeptide (OP) and homeodomain (HD) motifs. We further report several splice variants of each gene and nucleotide differences in coding sequence that predicts six putative amino acid changes between the two genes. Additionally, we noted a trinucleotide deletion in rt\textit{pax7\textbeta} that results in putative serine elimination at the N-terminus and a single nucleotide polymorphism (SNP) in majority of the rt\textit{pax7\textbeta} variants (6/10) that predicts an arginine substitution for a lysine. We also deciphered the genomic organization up to the first three exons and the upstream putative promoter regions of both genes. Comparative in silico analysis of both the trout pax7 promoters with that of zebrafish pax7 duplicates; zf\textit{pax7\textalpha} and zf\textit{pax7\textbeta}; predicts several important cis-elements/transcription factor binding sites (TFBS) in these teleost pax7 promoter regions.

Keywords: Gene duplication, Paired box transcription factor-7, Rainbow trout, Satellite cells, Skeletal muscle

Background
Growth and regeneration of skeletal muscle in vertebrates is mainly attributed to mitotically proficient adult muscle stem cells termed satellite cells (SCs) (Lepper et al. 2011; Moss and Leblond 1971; Motohashi and Asakura 2014; White et al. 2010). Activation of SCs by intrinsic cues such as hepatocyte growth factor (HGF) (Tatsumi et al. 1998) result in the generation of new myoblasts that terminally differentiate to myocytes and either fuse with the existing myofibers (hypertrophy) or among themselves to generate new myofibers (hyperplasia) (Collins et al. 2005; Mozdziak et al. 1997; Rowlerson et al. 2001). Alternately, the myogenic precursor cells can self-renew to replenish the intramuscular pool of SCs (Collins et al. 2005; Olguin et al. 2007). Depletion of pax7 expressing SCs in skeletal muscle lead to compromised growth and skeletal muscle regeneration (Pascoal et al. 2013; Sambasivan et al. 2011).

While the myogenic program of SCs is primarily driven by myogenic regulatory factors (Megeney et al. 1996; Montarras et al. 2000; Smith et al. 1994), their maintenance, propagation and self-renewal in growing muscle have been attributed to the expression of pax7 (Oustanina et al. 2004; Seale et al. 2004). The functional significance of pax7 in the physiology of SCs is primarily understood through studies conducted in knock-out mice (Kuang et al. 2006; Oustanina et al. 2004; Relaix et al. 2006; Seale et al. 2004). Indeed, pax7 is a widely accepted marker of SCs in vertebrates (Seale et al. 2000). Homozygous pax7 null mice either suffer early postnatal
lethality or grow to a small size, and show defective development of central nervous system and craniofacial muscles. Additionally, these mice suffer from defective postnatal skeletal muscle growth and regeneration due to a deficiency in the number of SCs (Kuang et al. 2006; Oustanina et al. 2004; Seale et al. 2000), suggesting that pax7 deletion affects skeletal muscle development. Most recent studies using inducible knockout mouse models have further shown that expression of pax7 is essential in mature skeletal muscle for effective regeneration and repair after injury ( Günther et al. 2013; von Malzahn et al. 2013). Together, pax7 can be advanced as a key player in SCs biology with significant roles in skeletal muscle plasticity during both development and adult stages of higher vertebrates.

The role of SCs in fish skeletal muscle growth is well recognized (Koumans and Akster 1995; Marschallinger et al. 2009; Pascoal et al. 2013; Rossi and Messina 2014; Seger et al. 2011). Similar to mammals, SCs in various fish species express pax7 (Devoto et al. 2006; Froehlich et al. 2013; Gotensparre et al. 2006; Marschallinger et al. 2009; Sibthorpe et al. 2006), and contribute to growth and regeneration of skeletal muscle ( Seger et al. 2011) suggesting an evolutionarily important role of this transcription factor in vertebrate skeletal myogenesis. However, unlike mammals, teleost fish genomes contain more than one pax7 gene. At least two pax7 genes exist in zebrafish (Minchin and Hughes 2008), which has been attributed to the whole genome duplication early in the teleost lineage after divergence from their common mammalian ancestor (Jaillon et al. 2004). The salmonid genome may contain more copies of pax7 (Gotensparre et al. 2006; Sibthorpe et al. 2006) due to another round of whole genome duplication around 88–103 Mya (Macqueen and Johnston 2014). Recent evidence from genomic sequencing studies in rainbow trout indicate that nearly half of the duplicated paralogs from this event are retained in the genome (Berthelot et al. 2014). Further evidence suggests that gene duplication in salmonids may also arise from localized gene duplication (Macqueen and Johnston 2006). Because of the importance of pax7 in mediation of skeletal myogenesis by SCs, and its genetic complexity in teleost, an improved characterization of gene/s and promoter would add to a comprehensive understanding of the regulation and function of pax7 in these species.

While growth of postnatal skeletal muscle in amniotes is primarily through hypertrophy, post-larval muscle accretion in salmonids is accomplished through both hyperplasia as well as hypertrophy (Mommsen 2001; Valente et al. 1998). Rainbow trout are an important global aquaculture species and an excellent animal model to study skeletal muscle growth that is mediated by SCs. However the structure and function of the pax7 gene/s and the corresponding promoter/s is not well understood. In this study, we isolated multiple transcript variants of two rainbow trout pax7 paralog genes (rt pax7α and rt pax7β), using skeletal muscle mRNA as the source of nucleotide sequences. Additionally, we deciphered the genomic organization of the first three exons and the associated 5′-flanking regions of both genes. Finally, an in silico analysis was performed to identify the potential cis-regulatory elements/TFBS in the putative promoter regions of each gene as compared to that of the zf pax7a and zf pax7b genes.

Methods
RNA isolation and RT-PCR
Skeletal muscle tissue from the hypaxial and epaxial regions of adult rainbow trout was collected following euthanization induced by 100 ppm of tricaine methanesulfonate (MS-222). Fish rearing, experimental sampling and handling procedures were approved by the University of Idaho Animal Care and Use Committee. All tissues were snap frozen in liquid N2 and stored at −80°C until RNA isolation. RNA was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer’s recommendations. Briefly, 1 ml of TRIzol was added to ~50 mg of ground muscle tissue and homogenized using a bead homogenizer at a frequency of 25 Hz for up to 1 min 30 s. RNA was separated by adding 0.2 ml of chloroform and centrifuged at 12,000×g for 10 min, at 4°C. The aqueous phase was collected and RNA was precipitated by adding 0.5 ml isopropanol alcohol and centrifuged at 12,000×g for 30 min at 4°C. RNA samples were finally washed twice in 1 ml of 75% ethanol and quantified on Nanodrop® ND-1000 UV–Vis Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) following manufacturer’s recommendations. Template quality was verified by visualization on a 1.5% formaldehyde agarose gel. A 2 µg of DNase I (Ambion, Foster City, CA, USA) treated RNA was then reverse transcribed using superscript III (Invitrogen) according to the manufacturer’s instructions, and the resultant cDNA served as a template for subsequent PCR amplification of pax7 cDNA sequences.

Isolation of pax7 cDNA clones
Multiple nucleotide comparisons of the Atlantic salmon (Salmo salar) and Arctic char (Salvelinus alpinus) pax7 cDNA variants (NCBI database) showed ~100% identity in the first exon and downstream of the initiator codon. Further, a rainbow trout pax7-like 3′-end enriched EST sequence (CB493668) also showed significant identity (94–98%) to various pax7 variants of the above two species. Using the above information, a set of gene specific primers (GP7F, GP7R) was designed to amplify the
putative complete protein coding sequences of trout pax7 (Table 1). A touch down PCR was performed using Platinum® Tag DNA Polymerase (Invitrogen) on the cDNA prepared from total RNA of skeletal muscle as a template. Thermal cycler parameters comprised of an initial denaturation at 94°C for 3 min. Next six cycles each had a denaturation step at 94°C for 15 s, annealing step for 30 s where temperature was dropped by 1°C/cycle from 60 to 55°C and extension at 72°C for 2 min. Subsequent amplification was conducted for 29 cycles that comprised denaturation at 94°C for 15 s, annealing at 54°C for 30 s and extension at 72°C for 2 min. The reactions were completed with final extension performed at 72°C for 10 min. The PCR products were resolved on a 1% agarose gel and the appropriate amplicon was eluted from the gel using PureLink Quick gel extraction kit (Invitrogen). Plasmid DNA was extracted using a GenElute™ HP Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) following manufacturer’s recommendations (Applied Biosystems, Foster City, CA, USA).

5′-rapid amplification of cDNA end to determine transcription start site (5′-RACE)

To identify the transcription start site (TSS), a rapid amplification of cDNA end to determine transcription start site (5′-RACE) was performed by employing either random hexamers or oligo(dt) primers. Subsequently, a touch-down nested PCR amplification was performed using reverse gene specific primers, GP5R1 (outer) and GP5R2 (inner) that were designed to anneal to a region around the start of PD region. Manufacturer supplied 5′-RACE adaptor primers served as forward primers. Thermal cycler parameters were as follows for the outer PCR; initial denaturation at 94°C–3 min. The next eight cycles each had a denaturation step at 94°C–30 s, annealing step for 1 min where temperature was dropped by 2°C for every two cycles from 64 to 61°C and extension at 72°C–1 min. The last 27 cycles had a denaturation step at 94°C–30 s, annealing step at 60°C–1 min and extension at 72°C–1 min. The reactions were completed with final extension performed at 72°C for 7 min. Subsequently performed inner PCR parameters were as follows; initial denaturation at 94°C–3 min. The next four cycles each had a denaturation step at 94°C–30 s, annealing step for 1 min where the temperature was dropped by 1°C for every two cycles from 66 to 65°C and extension at 72°C–1 min. The last 31 cycles had a denaturation step at 94°C–30 s, annealing step at 64°C–1 min and extension at 72°C–1 min. Final extension was performed for 7 min at 72°C.

**Isolation of genomic DNA and identification of pax7 gene and promoter sequences**

DNA was isolated from the skeletal muscle of adult rainbow trout. Briefly, ~100 mg of pulverized tissue was incubated at 37°C for 3 h in ten volumes of lysis buffer (10 mM Tris–Cl, pH 8.0, 0.1 M EDTA, pH 8.0, 0.5% SDS and 20 µg/ml DNase-free RNase A). Finally, proteinase K at a final concentration of 100 µg/ml was added to the lysate and incubated at 55°C overnight. DNA was then extracted by phenol–chloroform extractions followed by two washes in 75% ethanol. The quality of DNA was evaluated on a 1% agarose gel. Genome walker libraries were constructed using a GenomeWalker™ Universal kit (Clontech, Palo Alto, CA, USA) following the manufacturer’s instructions. Concisely, the genomic DNA was digested with blunt end restriction enzymes; EcoR V, Dra I, Pvu II or Stu I. The resultant DNA fragments were ligated to the Genome Walker adaptor supplied by the manufacturer (Clontech). Using a set of gene specific primers, GP7a1R, GP7a2R, gp7a3R and gp7a4R (Table 1) in combination with the manufacturer supplied adaptor primers, the rtpax7β genomic DNA (gDNA) from the start of the 4th exon and up to 980 bp upstream of the initiator codon was isolated. Similarly, we isolated rtpax7β gDNA of 2,624 bp.

### Table 1 List of primers used in PCR amplification reactions

| Primer       | Sequence                        | Length (bp) |
|--------------|---------------------------------|-------------|
| GP7F         | 5′-ATG ACT CCT TTA CCA GCA ACA GT-3′ | 23          |
| GP7R         | 5′-TCA GTA GCC CTG TCC CGT CTC-3′ | 21          |
| GP7a1        | 5′-GAA TGC CAT CGA TGC TAT CCT TTT-3′ | 27          |
| GP7a2        | 5′-CAC TAT CGT CGT CAT CTT TCT TGC-3′ | 27          |
| GP7a3        | 5′-TCT TAG CAA CAA TGT CAC CAT TGG TTT G-3′ | 29          |
| GP7a4        | 5′-GCA ATA ATG TCA CCA TGG GTT TGG TAA CTA-3′ | 29          |
| GP7b1        | 5′-GCT AAA GGG GTC TTC TTT TAC CCC ACA AA-3′ | 29          |
| GP7b2        | 5′-TTT TGA GAG GAC AAT CTA CTA GAT TCT-3′ | 29          |
| GP7b3        | 5′-AGA CCT AAG CAA ATG CTC GGC GGA AAA ATA C-3′ | 28          |
| GP7b4        | 5′-CTT TTT ATG CTA GCC CCT GGG ACA GTC T C-3′ | 28          |
| GP7b5        | 5′-GGG GGT TGA TAC TGT TCC ACA ATA AAG ATA CT-3′ | 30          |
| GP7b6        | 5′-GGG GCA TCA TCG TAC GGC CGA GTG TCT CTC CTG G-3′ | 30          |
| GP5R1        | 5′-TGC GGC ATC TCT ACT ATG TGT TGT CTG A-3′ | 28          |
| GP5R2        | 5′-GGG GGT TGA TAC TGT TCC ACA ATA AAG ATA-3′ | 30          |
upstream of the initiator codon (gp7b5R and gp7b6R) (Table 1), and up to around the first three exons (using forward primers; gp7b1F, gp7b2F, gp7b3F and gp7b4F) (Table 1). All PCR were performed using Advantage 2 Polymerase Mix (Clontech) and thermal cycler parameters for outer PCR included 5 cycles at 94°C–25 s, denaturation; 72°C for 3 min, annealing and extension; followed by 32 cycles at 94°C–25 s, 67°C–3 min. A final extension at 67°C was performed for 7 min. All PCR products were analyzed on a 1.5% agarose gel, subcloned and sequenced as reported above.

**Phylogenetic analysis**

Phylogenetic was performed using phylogeny.fr pipeline (http://www.phylogeny.fr) with default parameters using rainbow trout putative Pax7 sequences (rtPax7A and rtPax7B) along with those previously reported for other species; Atlantic salmon Pax7α (CAF02090), Atlantic salmon Pax7β (CAHO4385), zebrafish Pax7a (NP_571400.1), zebrafish Pax7b (NP_001139621), tilapia Pax7 (LOC100708659, XP_003454575), tilapia Pax7 (LOC100696153, XP_003459869), stickleback Pax7 (1/2) (ENSGACP0000017071), stickleback Pax7 (2/2) (ENSGACP000002231), medaka Pax7 (2/2) (ENSO1P0000005345), human Pax7 (DQ322591.1) and mouse Pax7 (NP_035169.1). Briefly, multiple alignments of amino acid sequences were accomplished using MUSCLE alignment program. Alignment curation was performed by Gblocks and phylogenetic analysis was performed by PhyML using a default substitution model (Dereeper et al. 2008).

**In silico analysis**

Sequence analyses were performed using Vector NTI 11.5 advance (Invitrogen) at default parameters. Contiguous sequence (contig) alignment of nucleotide sequences was performed using the Contig assembly feature in Vector NTI 11.5 advance (Invitrogen). Gene homology searches were performed using the blast resources of NCBI database. Multiple comparisons of nucleotide and deduced amino acid sequences were performed using clustalW set at default parameters. Exon/intron boundaries were delineated using Spidey that is available at NCBI public domain (http://www.ncbi.nlm.nih.gov/spidey/). Prosite was used to perform pattern searches in the deduced amino acid sequences (Hulo et al. 2006). Putative cis-regulatory elements/transcription factor binding sites (TFBs) in gene regulatory/promoter regions were predicted using MatInspector software (http://www.genomatix.de), set at a matrix similarity threshold of 0.75.

**Results**

**Isolation and characterization of two trout pax7 paralogs**

Performing RT-PCR on cDNA synthesized from skeletal muscle total RNA produced an expected band of ~1,500 bp that was subcloned and sequenced. Subsequent analysis of all cloned cDNA yielded two distinct cDNA forms. Blast analysis showed high homology to pax7 paralogs of other teleost fish; Atlantic salmon, Arctic char and zebrafish. We refer to these two transcript forms as rtPax7α and rtPax7β. A 5′-RACE protocol revealed two 5′-untranslated region (UTR) sequences for rtPax7α (identical sequence of variable length, 135 and 387 bp) and one for rtPax7β (388 bp). Assimilation of the respective 5′-UTR sequences with that of the longest transcript variant of rtPax7α1 (Figure 1a) and rtPax7β1 (Figure 1b) (Contig assembly application of vector NTI advance 11.5) resulted in 1935 and 1945 nucleotide long sequence, respectively. Conceptual translation of rtPax7α1 (ATG at 388 nt.) and rtPax7β1 (ATG at 389 nt.) resulted in putative proteins of 515 and 518 amino acids, respectively. Analyses of pattern identification using Prosite revealed the presence of conserved features in both putative protein forms that are characteristic of Pax7 protein; paired domain (PD), homeodomain (HD) and an octapeptide (OP) (Figure 1). A pairwise comparison of the nucleotide sequences showed that both forms share 83% identity in the 5′-UTR sequences, while the protein coding sequences share 97% identity. However, putative amino acid sequences are 98% identical. Therefore not all nucleotide polymorphisms in the coding regions resulted in amino acid substitutions. Indeed 47 polymorphic nucleotides were observed in the coding sequence, while only seven putative amino acid variations including a serine 167 deletion in rtPax7β1 (as compared to rtPax7α1) were observed between the two.

**Multiple splice variants**

Multiple transcripts of each pax7 gene differed in length due to an alternate indel of three regions, which suggests these transcripts represent alternate splice variants of each gene. Specifically, rtPax7α variants differed due to an indel of 39 bp (GQY(T)GPEYVYCGT), 15 bp (GEASS) or 12 bp (GNRT). Similarly, variants of rtPax7β differed by an indel of 39 bp (GQY(A)GPEYVYCGT), 12 bp (GEAS) or 12 bp (GNRT). Additionally, the majority of the rtPax7β clones sequenced (6/10) had a putative arginine substituted for lysine 197 due to an AAG to AGG transition, which suggests a potential allelic variation in the gene. The sequences of all the variants were deposited in the GenBank (Table 2).
Partial genomic characterization

Employing a genome walker universal kit (Clontech) and a set of gene specific primers, genomic sequences corresponding to the 5' UTR sequences and up to around the third exon were amplified (Figure 2). Sequenced partial gDNA of both rpax7α and rpax7β were individually aligned with corresponding transcripts using Spidey (NCBI) that delineated the exon/intron boundaries. While the substitution of threonine in rpax7β for alanine 32 (rpax7α) that occurs due to an ACA to GCA transition resides at the 5'-end of second exon, the substitution of serine in rpax7β for alanine 123 (rpax7α) as a result of a GCA to TCA transversion, resides at the 3'-end of third exon (Figure 2). Although the presence of the trinucleotide sequence coding to serine 167 in rpax7α was confirmed by our study, we did not sequence the gDNA corresponding to deleted serine 167 in rpax7β. However, evidence from studies in other salmonids indicates the genomic fidelity of this deletion, given that it has been mapped to the gDNA of one of the two putative pax7 paralogs reported for Atlantic salmon and Arctic char (Gottensparrer et al. 2006; Sibthorpe et al. 2006). Further, the 39 bp indel maps to the 5'-end of second exon in both forms, while the 15 bp indel of rpax7α maps to the 3' end of third exon. In both cases, the splice junctions are consistent with the GT-AG rule. Further, the first introns of rpax7α and rpax7β were 1,309- and 1,333-bp, respectively, and had 80% sequence identity overall (data not shown). Similarly, the second introns were 550 and 617 bp long for rpax7α and rpax7β, respectively.
Table 2  Splice variants of pax7 paralogs

| cDNA variants | 13 aa (±) | 5/4 aa (±) | 4 aa (±) | SNP | CDs (bp)* | Protein (aa) | Accession number |
|---------------|----------|-----------|-----------|-----|-----------|--------------|-----------------|
| rtpax7α       | +        | +         | –         | K   | 1,548     | 515          | JQ303311        |
| rtpax7β       | +        | –         | +         | K   | 1,545     | 514          | JQ303312        |
| rtpax7β       | –        | +         | –         | K   | 1,521     | 506          | JQ303313        |
| rtpax7β       | –        | –         | +         | K   | 1,509     | 502          | JQ303314        |

* Including stop, TGA; † indels corresponding to 13 aa, 5/4 aa and 4 aa residues; SNP single nucleotide polymorphism in rtpax7β.

Figure 2  Partial genomic characterization of rainbow trout (rt) pax7 DNA forms; rtpax7α and rtpax7β. Multiple cDNA of both rtpax7α and rtpax7β are aligned with the respective genomic DNA (gDNA) sequences. Deduced amino acid sequences are shown on the top of each alignment. Exon boundaries are indicated by an arrow (▼). Polymorphic codons are underlined; amino acid variations are highlighted in black. Amino acid residues corresponding to the insertion or deletion (indel) of the 39 bp in exon 2 or 15 bp in exon 3 are boxed. Identical nucleotide residues in all sequences are indicated by dotted line.

and had 77% overall identity (data not shown). The 3rd intron sequenced only for rtpax7α was 216 bp. Although the genomic fidelity of 12 bp (GNRT) region is not verified in this study, previous studies indicate that this 12 bp indel maps to the 5’-end of fifth exon in salmonid pax7. These observations collectively indicate that rtpax7α and rtpax7β CDAs are transcribed from two pax7 genes that likely arose as a result of whole genome duplication.

Phylogenetic associations

All teleost pax7 clustered differently from their mammalian orthologs and into two separate groups with high confidence (100%) (Figure 3). As suggested previously (Minchin and Hughes 2008), the observed topology indicates the existence of two pax7 gene clades in teleost fish. The trout and Atlantic salmon pax7 clustered into the same clade, and isoforms of both genes showed greater identity between the species than within the species isoform comparisons. This suggests a salmonid pax7 gene paralogy that likely arose during a second round of presumed fish specific whole genome duplication. Clustering of zpax7a, but not zpax7b, sequences into same clade further suggests that rtpax7α and rtpax7β are co-orthologs of zebrafish pax7a.

In silico analysis of pax7 promoter regions

Analysis of the rtpax7α (Figure 4) and the rtpax7β (Figure 5) promoter regions using MatInspector software
Figure 3  Phylogenetic analysis of Pax7 sequences of rainbow trout and other vertebrate species. Phylogenetic analysis of various vertebrate pax7 sequences including rainbow trout (pax7α and pax7β) was performed using phylogeny.fr pipeline with default parameters. Amino acid sequences of rainbow trout pax7, pax7α and pax7β, Atlantic salmon pax7α (CAF02090), Atlantic salmon pax7β (CAH04385), zebrafish pax7 (NP_571400.1), stickleback pax7 (1/2) (ENSGACP00000017071), stickleback pax7 (2/2) (ENSGLP00000005345), human PAX7 (DQ322591.1) and mouse PAX7 (NP_035169.1) were used. Branch support values (%) were reported on each branch.

Figure 4  Nucleotide sequence of rainbow trout pax7α (rtα) promoter region. Putative transcription factor binding sites are highlighted in colors and labeled above the consensus sequence. Putative E-box consensus sequences (CAN(T/A)TG) are boxed and TATA-box binding motif is underlined. Transcription start site is identified in this study. 

Figure 5  Nucleotide sequence of rainbow trout pax7β (rtβ) promoter region. Putative transcription factor binding sites elements are highlighted in colors and labeled above the consensus sequence. Putative E-box consensus sequences (CAN(T/A)TG) are boxed. TATA box sequence is underlined. Nucleotide positions are relative to initiation ATG (+1). Transcription start site is double underlined.

showed consensus binding sites for several TFs in both the promoters. In the rtpax7α promoter region, a consensus TATA binding sequence is located between −452 and −468, 63 bases upstream of the TSS identified in this study. Important binding sites includes consensus
sites for the following TFs: Octamer binding transcription factor-4 (Oct4), Nanog, androgen response element (ARE), muscle specific mitogen binding factor (MtBF), Sine-oculis homeobox 1 homolog (Six1) and CCAAT/ enhancer binding protein (C/ebp) (Figure 4). In the rt$\text{pax7}\beta$ promoter region, we found a consensus TATA binding sequence 219 bp upstream of the TSS identified in this study. However, a TATA-like sequence AATTAA is also present 66 bp upstream of the TSS. Important binding sites includes consensus sites for the following TFs: a cAMP Responsive Element-Binding protein (Creb), Oct4, MtBF, E-box protein homodimer (E47), Myocyte specific enhancer factor-2 (Mef2), Six1, SRY related HMG box factors -5 (Sox5) and -15 (Sox15), glucocorticoid response element (GRE), progesterone response element (PRE), Nuclear factor of activated T-cells 5 (Nfat5), Krüppel-like zinc-finger transcription factor (Zbp89) and C/ebp (Figure 5). Also, both promoters have multiple putative E-boxes with consensus CAN(T/A)TG sequence.

Comparative in silico analysis of trout $\text{pax7}$ promoters with that of the putative zebrafish promoter regions; zf$pax7a$ and zf$pax7b$, show that several if not all, TFBS are similarly present in these two zebrafish $\text{pax7}$ promoter regions (Figure 6). Binding sites for Oct4 and Six1 (except zf$pax7b$) are observed in all promoters analyzed, although multiple Oct4 sites are present in zebrafish $\text{pax7}$ promoters. All promoters with the exception of rt$pax7\alpha$ also have more than one Sox binding site. Specifically, Sox15 and Sox5 binding site/s are predicted in both rt$pax7\beta$ and zf$pax7a$ promoter regions. Additionally, binding sites for Sox2 and Sox9 transcription factors are predicted in the zf$pax7b$ promoter region. Although the various promoters differ in the genetic nature of these cis-elements, the presence of Sox binding elements appears to be a common feature in these fish promoters as the rainbow trout $\text{pax7}$ promoter region (sequence that is highly similar to rt$pax7\alpha$) also have binding sites for Sox9 and Sox15. Comparative analyses further show that these fish promoters have MRF binding sites. Although we did not find putative sites for MRFs within the sequences we cloned, a myogenic factor 6 (Myf6) binding site in zf$pax7a$ promoter and a MyoD/E47 heterodimer binding site in zf$pax7b$ promoter were identified. Also, an ARE that is observed in rt$pax7\alpha$ is also present in zf$pax7b$ promoter. However, a Zbp89 binding site is identified only in rt$pax7\alpha$, while rt$pax7\beta$ and zf$pax7b$ promoters each have one Nfat5 binding site. Although the binding and functional relevance have to be experimentally derived, presence of these cis-regulatory elements in trout promoters suggests some degree of evolutionary conservation in $\text{pax7}$ gene regulation.

**Discussion**

Paired box-7 transcription factor has been implicated in vertebrate skeletal muscle growth and development. It is expressed in skeletal muscle SCs and plays a principal regulatory role in adult skeletal myogenesis. In this study, we identified two $\text{pax7}$ genes (rt$pax7\alpha$ and rt$pax7\beta$) and their putative splice variants in rainbow trout. Further, we sequenced the promoter regions of both genes and performed an in silico analysis to identify the putative TFBS in the promoter regions of both genes as compared to duplicate zebrafish $\text{pax7}$ gene promoters.

An important finding of this study was the identification and characterization of two highly homologous $\text{pax7}$ genes in rainbow trout, suggesting both genes likely arose from a salmonid whole genome duplication.
event (Macqueen and Johnston 2014). With the recent availability of the trout genome sequences (Berthelot et al. 2014), one recent study reported the presence of three mammalian pax7 co-orthologs in rainbow trout genome: pax7a1, pax7a2 and pax7b1 (Seiliez et al. 2014). Comparison of our sequences with the partial cDNA sequences annotated in this study showed nucleotide identities of 97.7% between rt Pax7a and pax7a2 and 99.7% between rt Pax7β and pax7a1, indicating that we have successfully identified the two genes of the pax7a clade of rainbow trout. As also reported (Seiliez et al. 2014), our phylogenetic analysis further support this observation because both rt Pax7a and rt Pax7β cladistically belong with zlPax7a. Although our analysis also supports the existence of additional pax7 gene/s in second clade, we did not identify sequences that are homologous to pax7b1 in this study. This is however not surprising given the sequence dissimilarities between pax7a members and pax7b1 at the extreme 5’ end of the coding sequence where we targeted our forward primer to amplify rtPax7a and rtPax7β sequences.

Both pax7 paralogs also express multiple splice variants in the adult skeletal muscle, which adds another layer of complexity to pax gene function in trout. Expression of splice variants appears to be a common mode of pax gene regulation (Barber et al. 1999; Pritchard et al. 2003; Ziman and Kay 1998). For instance in mice, the single gene expresses 4 alternate spliced transcripts in adult skeletal muscle tissue (Ziman and Kay 1998). These variants show different expression levels and altered DNA binding and transactivation properties (Du et al. 2005; Ziman and Kay 1998). Similarly, expression of multiple pax7 splice variants has been reported in zebrafish, Atlantic salmon and Arctic char (Gotensparre et al. 2006; Seo et al. 1998; Sibthorpe et al. 2006). Moreover, the alternately skipped 14 amino acid residues at the N-terminus appear to be a common feature of salmonid fish Pax7. Additionally, the substitution of threonine for alanine 123 in rtPax7a resulted in the inclusion of two additional casein kinase—II phosphorylation sites with a consensus sequence of S/T-X-X-D/E. CK2 is a common serine threonine kinase and phosphorylates multiple factors involved in vertebrate myogenesis (Johnson et al. 1996; Molkentin et al. 1996; Winter et al. 1997). Also, the deletion of GNRT residues from either form also results in the elimination of an Asn-glycosylation site that has a consensus, Asn-X-S/T-Y, sequence (NRTD). These features are particularly interesting as the previous studies on Pax3/7-FKHR fusion proteins revealed that a cis-acting functional transcriptional repression domain exists at the N-terminus of both Pax3 and Pax7 (Bennicelli et al. 1999). Although the relevance of these variations in Pax7 post-translational modifications, DNA binding affinity and transactivation properties has to be functionally determined; the production of multiple splice variants may provide enormous diversity in Pax7 target gene regulation.

Extant paralog genes may develop unique expression patterns and acquire diverse fates including gain of novel function (neo-functionalization) or retain a portion of the original gene function (sub-functionalization) (Conant and Wolfe 2008). These diverse fates of the paralogs can at least in part arise by divergence in their regulatory/promoter regions (Van Hellemont et al. 2007). Recently, Seiliez et al. (2014) showed that pax7 paralogs of rainbow trout differ in their expression pattern during the course of satellite cell conversion to myocytes that has been attributed to the differential epigenetic histone modifications in the pax7 gene loci. Our in silico examination reveals putative binding sites for several important TFs in the promoter regions of both trout pax7 genes. Comparative analyses suggest that several of these sites are also present in zebrafish pax7 gene promoters. The presence of one or more Oct4 binding sites in all promoters analyzed in the present study suggests an evolutionarily conserved role of this transcription factor in pax7 gene regulation. Oct4 is one of the master inducers of pluripotency in embryonic stem cells, and studies show that Oct4 binds mouse pax7 promoter and functions as a transcriptional repressor (Lang et al. 2009). Further, the presence of a putative binding site for Nanog that is also an inducer of pluripotency in close proximity to TSS of rtPax7a is especially interesting, as co-expression of Nanog with Oct4 significantly inhibits myogenic cell differentiation, although binding of Nanog itself to pax7 promoter was not observed (Lang et al. 2009). Further, putative binding sites for various members of Sox family were observed in these fish pax7 promoters. Past studies, although primarily in mammalian models, showed that satellite cells express various sox genes. Sox8 has been implicated in maintaining the satellite cell progenitor population (Schmidt et al. 2003). In vitro cell culture studies using P19 cell lines showed that both Sox15 and Sox7 influence Pax3/Pax7 expression and myogenesis (Savage et al. 2009). Also, genetic knock-out studies in mice show that Sox15 is essential for adult skeletal muscle regeneration (Lee et al. 2004). The identification of a putative binding site for Zbp89 that binds human pax7 promoter was also noted in rtPax7β promoter (Salmon et al. 2009). Zbp89 is expressed in skeletal muscle of various mammalian species (Merchant et al. 1996) and acts as a pax7 transcriptional repressor while enhancing myogenic differentiation (Salmon et al. 2009). While the presence of commonality in the genetic nature of cis-elements in these fish promoters suggests significant degree of conservation in pax7 gene regulation at the promoter
level, presence of unique cis-elements in each of trout pax7 promoter regions may point to the potential differences in their regulation. Nevertheless, a detailed functional characterization is required to unambiguously ascertain the role of these and other TFBS in trout pax7 promoters.

Conclusions
In conclusion, the sequence information of rainbow trout putative pax7 paralog genes and their corresponding splice variants will facilitate future studies designed to characterize the tissue and stage-specific expression profiles of these transcript variants and their consequent function. Although the physical presence of the reported TFBS in these teleost pax7 promoters may not necessarily confer binding by the corresponding TFs, existing literature report the expression of many of these TFs in the context of vertebrate skeletal myogenesis that is mediated by SCs, which indicates a functional relevance of these TFs in teleost pax7 gene regulation. Therefore, future studies that delineate the minimal promoter regions and experimental characterization of these putative cis-elements would shed more light on the functional aspects of pax7 gene regulation.

Authors’ contributions
KCC: Obtained UI Student research grant, performed all benchwork, sequence analyses, data interpretation and manuscript preparation; BMM: Performed sequence analyses, data interpretation and manuscript editing; BDR: Performed data interpretation and manuscript editing; RAH: Manuscript editing; GKM: Intellectual idea, obtained funding and covered all experimental costs, provided laboratory facilities, performed data interpretation and manuscript preparation. All authors read and approved the final manuscript.

Author details
1 Department of Animal and Veterinary Science, University of Idaho, PO Box 442330, Moscow, ID 83844-3051, USA. 2 Department of Biological Sciences, University of Idaho, 875 Perimeter MS 3051, PO Box 442330, Moscow, ID 83844-3051, USA. 3 Ag Biotech 309, Department of Animal and Veterinary Science, University of Idaho, PO Box 442330, Moscow, ID 83844-3051, USA. 4 Present Address: University of South Florida, 12901 Bruce B. Downs, Tampa, FL 33612, USA.

Acknowledgements
We thank the University of Idaho Student Grant Program (SGP) for supporting this study under the award number BC2038 as well as National Science Foundation and the Idaho Experimental Program to Stimulate Competitive Research (NSF-EPSCoR) under award number EPS0447689 as well as the USDA Hatch Grant NC-1184.

Compliance with ethical guidelines
Competing interests
The authors declare that they have no competing interests.

Received: 1 November 2014 Accepted: 12 May 2015 Published online: 17 June 2015

References
Barber TD, Barber MC, Cloutier TE, Friedman TB (1999) PAX3 gene structure, alternative splicing and evolution. Gene 237:311–319

Bennecelli JL, Advani S, Schafer BW, Barr FG (1999) PAX3 and PAX7 exhibit conserved cis-acting transcription repression domains and utilize a common gain of function mechanism in alveolar rhabdomyosarcoma. Oncogene 18:4348–4356. doi:10.1038/sj.onc.1202812

Berthelot C, Brunet F, Chalepin D, Juanchich A, Bernard M, Noël B et al (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun 5:3657. doi:10.1038/ncomms4657

Collins CA, Olien I, Zammit PS, Heslop L, Petrie A, Partridge TA et al (2005) Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. Cell 122:289–301. doi:10.1016/j.cell.2005.09.010

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. doi:10.1093/nar/gkn180

Devoto SH, Stoiber W, Hammond CL, Steinbacher P, Hasley, JR, Barresi MJ et al (2006) Generality of vertebrate developmental patterns: evidence for a dermomyotome in fish. Evol Dev 8:101–110. doi:10.1111/j.1525-142X.2006.05079.x

Du S, Lawrence EJ, Sztrelecki D, Rajput P, Kita SJ, Gottesman DM et al (2005) Co-expression of alternatively spliced forms of PAX3, PAX7, PAX3-FKHR and PAX7-FKHR with distinct DNA binding and transactivation properties in rhabdomyosarcoma. Int J Cancer 115:85–92. doi:10.1002/ijc.20844

Froehlich JM, Galt NJ, Charging MJ, Meyer BM, Bipa GP (2013) In vitro indeterminate teleost myogenesis appears to be dependent on Pax3. In Vitro Cell Dev Biol Anim 49:371–385. doi:10.1007/s11626-013-9616-2

Gottensparre SM, Anderson E, Wangelius A, Hansen T, Johnston IA (2006) Insight into the complex genetic network of tetraploid Atlantic salmon (Salmo salar L). description of multiple novel Pax-7 splice variants. Gene 373:8–15. doi:10.1016/j.gene.2005.12.026

Günther S, Kim J, Kostin S, Lepper C, Fan C-M, Braun T (2013) Myf5-positive satellite cells contribute to Pax7-dependent long-term maintenance of adult muscle stem cells. Cell Stem Cell 13:590–601. doi:10.1016/j.stem.2013.07.016

Hulo N, Baroosh A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS et al (2006) The PROSITE database. Nucleic Acids Res 34:D227–D230. doi:10.1093/nar/gkj063

Jaillon O, Aury JM, Brunet F, Petit JL, Ange-Thomsen N, Mauceli E et al (2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431:946–957. http://www.nature.com/nature/journal/v431/n7011/suppinfo/nature03025_S1.html

Johnson SE, Wang X, Hardy S, Taparowsky EJ, Konieczny SF (1996) Casein kinase II expression, alternative splicing and evolution. Gene 237:311–319

Johnson SE, Wang X, Hardy S, Taparowsky EJ, Konieczny SF (1996) Casein kinase II. FEBS Lett 36:141–145. doi:10.1016/0014-5793(96)00222-3

Koumans JTM, Akster HA (1995) Myogenic cells in development and growth of fish. Comp Biochem Physiol A Physiol 110:3–20

Kuang S, Charge SB, Seale P, Huh M, Rudnicki MA (2006) Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. J Cell Biol 172:103–113. doi:10.1083/jcb.200508001

Lang KC, Lin IH, Teng HF, Huang YC, Li CL, Tang KT et al (2009) Simultaneous overexpression of Oct4 and Nanog abrogates terminal myogenesis. Am J Physiol Cell Physiol 297:C43–C54. doi:10.1152/ajpcell.00468.2008

Lee HJ, Göring W, Ochi M, Mühlfeld C, Steeding G, Paprotta I et al (2004) Sox15 is required for skeletal muscle regeneration. Mol Cell Biol 24:8428–8436. doi:10.1128/mcb.24.19.8428-8436.2004

Lepper C, Partridge TA, Fan C-M (2011) An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. Development 138:3639–3646. doi:10.1242/dev.067595

Macqueen DJ, Johnston IA (2006) A novel salmonid myod gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization. FEBS Lett 580:4996–5002. doi:10.1016/j.febslet.2006.08.016

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole-genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 281:7. doi:10.1098/rspb.2013.2981

Marschallinger J, Obermayer A, Sanger AM, Stoiber W, Steinbacher P (2009) Postembryonic fast muscle growth of teleost fish depends upon a...
nonuniformly distributed population of mitotically active Pax7+/+ precursors. Dev Dyn 238:2442–2448. doi:10.1002/dvdy.22049
Megeney LA, Kablar B, Garrett K, Anderson JE, Rudnicki MA (1996) MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev 10:1712–1718
Merchant JL, Jyer GR, Taylor BR, Kitchen JR, Mortensen ER, Wang Z et al (1996) ZBP-89, a Kruppel-like zinc finger protein, inhibits epidermal growth factor induction of the gastrin promoter. Mol Cell Biol 16:6644–6653
Minchin JE, Hughes SM (2008) Sequential actions of Pax3 and Pax7 drive xenotransplantation in zebrafish neural crest. Dev Biol 317:506–522. doi:10.1016/j.ydbio.2008.02.058
Molkentin JD, LL Olson EN (1996) Phosphorylation of the MADS-Box transcription factor MEF2C enhances its DNA binding activity. J Biol Chem 271:17199–17204. doi:10.1074/jbc.271.29.17199
Mommens TP (2001) Paradigms of growth in fish. Comp Biochem Physiol B: Biochem Mol Biol 129:207–219. (pii: S109649591003128)
Montarrant D, Lindon C, Piniset C, Domneyre F (2000) Cultured myf5 null and myoD null muscle precursor cells display distinct growth defects. Biocell 23:565–567
Moss FP, Leblond CP (1971) Satellite cells as the source of nuclei in muscles of growing rats. Anat Rec 170:421–435. doi:10.1002/ar.1091700405
Motohashi N, Asakura A (2014) Muscle satellite cell heterogeneity and self-renewal. Front Cell Dev Biol. 2:00001. doi:10.3389/fcell.2014.00001
Mozdzik PE, Schultz E, Cassens RG (1997) Myonuclear accretion is a major determinant of avian skeletal muscle growth. Am J Physiol 272:C565–C571
Olguin HC, Yang Z, Tapscott SJ, Olwin BB (2007) Reciprocal inhibition between myoD and myf5 regulates the myogenic program in P19 cells. Stem Cells 27:1231–1243. doi:10.1602/jcb.200508044
Passam S, de Lima JE, Leslie JD, Hughes SM, Saude L (2013) Notch signalling is required for activation, proliferation, and differentiation of rat skeletal muscle satellite cells. J Cell Physiol 219:114–128. doi:10.1002/jcp.24592
Pascoal S, de Lima JE, Leslie JD, Hughes SM, Saude L (2013) Notch signalling is required for the formation of structurally stable muscle fibres in zebrafish. PLoS One 8:e68021. doi:10.1371/journal.pone.0068021
Pichardt C, Grosz P, Hollebenbach AD (2003) Alternative splicing of MyoD produces a transcriptionally inactive protein. Gene 305:61–69
Relaix F, Montarras D, Zaffran S, Gayraud-Morel B, Rocancourt D, Tajbakhsh S et al (2008) Sequential actions of Pax3 and Pax7 drive xenotransplantation in zebrafish neural crest. Dev Biol 317:506–522. doi:10.1016/j.ydbio.2008.02.058
Salmon M, Owens GK, Zehner ZE (2009) Over-expression of the transcription factor MEF2C enhances its DNA binding activity. J Biol Chem 274:17199–17204. doi:10.1074/jbc.274.29.17199
Sambasivan R, Yao R, Kissenpfennig A, Van Wittenberghe L, Paldi A, Gayraud-Morel B et al (2006) Pax7+/+ nonuniformly distributed population of mitotically active Pax7 expressing myogenic cells in zebrafish muscle development, injury, and models of disease. Dev Dyn 240:2440–2451. doi:10.1002/dvdy.22745
Seiliez I, Froehlich JM, Marandel L, Gabbard JC, Biga PR (2014) Evolutionary history and epigenetic regulation of the three paralogous pax7 genes in rainbow trout. Cell Tissue Res 10:10. doi:10.1007/s00441-014-2060-0
Seo HC, Saetre BO, Hawl B, Ellingsen S, Fjosne A (1998) The zebrafish Pax7 and Pax9 homologues are highly conserved, encode multiple isoforms and show dynamic segment-like expression in the developing brain. Mech Dev 70:49–63
Sibthorpe D, Sturlaugsdottir R, Kristjansson BK, Thoraensen H, Skulason S, Johnston IA (2006) Characterisation and expression of the paired box protein 7 (Pax7) gene in polymorphic Arctic char (Salvelinus alpinus). Comp Biochem Physiol B Biochem Mol Biol 145:371–383. doi:10.1016/j.cbpb.2006.08.013
Smith CR, 2nd, Janney MJ, Allen RE (1994) Temporal expression of myogenic regulatory genes during activation, proliferation, and differentiation of rat skeletal muscle satellite cells. J Cell Physiol 159:379–385. doi:10.1002/jcp.104590222
Tatsunami R, Anderson JE, Neeret CJ, Halevy O, Allen RE (1998) HGF/SF is present in normal adult skeletal muscle and is capable of activating satellite cells. Dev Biol 194:114–128. doi:10.1006/dbio.1997.8803
Valente LMP, Gomes EFS, Faconnneau B (1998) Biochemical growth characterization of fast and slow-growing rainbow trout strains: effect of cell proliferation and size. Fish Physiol Biochem 18:213–224. doi:10.1006/fpbi.1998.1664
Van Hellenmont G, Blomme T, Van de Peer Y, Marchal K (2007) Divergence of regulatory sequences in duplicated fish genes. Genome Dyn 3:81–100. doi:10.1015/j.jcb.2006.08.013
von Maltzahn J, Jones AE, Parks RJ, Rudnicki MA (2013) Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. Proc Natl Acad Sci USA 2011:16474–16479. doi:10.1073/pnas.1307680101
White RB, Berins AS, Gnocchi VS, Zammit FS (2010) Dynamics of muscle fibre growth during postnatal mouse development. BMC Dev Biol 10:21. doi:10.1186/1471-213X-10-21
Winter B, Kautzner L, Issinger O-G, Arnold H-H (1997) Two putative protein kinase CK2 phosphorylation sites are important for Myf-5 activity. Biol Chem 378:1445–1456. doi:10.1515/bchm.1997.378.12.1445
Zimmer MR, Kay-PH (1998) Differential expression of four alternate Pax7 paired box transcripts is influenced by organ- and strain-specific factors in adult mice. Gene 217:77–81
