

Characterisation and expression of the paired box protein 7 (*Pax7*) gene in polymorphic Arctic charr (*Salvelinus alpinus*)

Dean Sibthorpe^{a,1}, Rán Sturlaugsdóttir^b, Bjarni K. Kristjánsson^b,
Helgi Thorarensen^b, Skúli Skúlason^b, Ian A. Johnston^{a,*}

^a Gatty Marine Laboratory, School of Biology, University of St Andrews, St Andrews, Fife KY16 8LB, Scotland, UK

^b Hólar University College, Hólar, Hjalteadal, IS-551 Saudárkrókur, Iceland

Received 24 April 2006; received in revised form 24 August 2006; accepted 28 August 2006

Available online 6 September 2006

Abstract

Arctic charr (*Salvelinus alpinus* L.) from Lake Thingvallavatn, Iceland occur as four distinct morphs: large benthivorous (LB), dwarf benthivorous (DB), piscivorous (PI) and planktonivorous (PL). The morphs differ with respect to body size, head morphology, growth rate, and life history. The aim of this study was to investigate the paired box protein 7 (*Pax7*) gene as a candidate for such polymorphisms due to its importance in cranio-facial, skeletal muscle, and central nervous system development. No variation in coding and intronic sequences was found between morphs. We identified 10 alternate *Pax7* isoforms with insertions/deletions: a four-residue (GNRT) deletion, a GEASS insertion truncated by the first serine residue (GEAS), and a thirteen-residue insertion (GQYA/TGPEYVYCGT). The latter insertion with a threonine (T) contains a putative casein kinase II (CK-2) phosphorylation site. *Pax7* spatial expression patterns were identical in embryos of DB-, LB-, and PL-morphs, and were similar to those described for zebrafish *Pax7c*, but a difference in temporal expression for segmentation was observed between DB and LB morphs. At the end of segmentation, novel expression was observed in the mandibular region as two bilateral domains. The potential role of multiple alternative splicing of the *Pax7* gene for the generation of different Arctic charr morphs is briefly discussed.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Salmonid; Arctic charr; *Salvelinus alpinus*; Resource polymorphism; Morph; *Pax7*; Multiple splice variants; Evolution; Skeletal muscle; Development

1. Introduction

The salmonid species Arctic charr (*Salvelinus alpinus*) has a northern circumpolar distribution and occurs as anadromous and non-migratory populations (Johnson, 1980). Resource polymorphism is common in landlocked populations (Skúlason and Smith, 1995), and an extreme case of resource polymorphism is observed in Thingvallavatn, Iceland (64°11'N 21°08'W) where there are four distinct coexisting morphs: large benthivorous (LB), small or dwarf benthivorous (DB), piscivorous (PI) and planktonivorous (PL) (Skúlason et al., 1989). Further clustering of morphs, based on trophic morphology, segregates charr into

benthic (LB/DB) and limnetic (PI/PL) morphotypes (Snorrason et al., 1994). The morphs differ with respect to behaviour and life history characteristics including somatic growth, age at sexual maturity, head morphology, colour, and maximum body size (Skúlason et al., 1989; Sandlund, 1992; Skúlason and Smith, 1995). Intraspecific phenotypic variation is thought to have several mechanisms including phenotypic plasticity, maternal effects, genetic polymorphism and genetic differentiation (Skúlason et al., 1992). Hypotheses regarding intraspecific polymorphism based on phenotypic plasticity argue that the formation of stable, alternate phenotypes represents the early stages of the speciation process (West-Eberhard, 1998; Adams and Huntingford, 2002b; West-Eberhard, 2003).

Studies of molecular genetic heterogeneity using minisatellite data in the Thingvallavatn charr, show no intra-morphotypic differences, but do show significant differences between morphs (benthic versus limnetic) compared to other Icelandic lakes (Volpe and Ferguson, 1996). Recent microsatellite data also confirm that sympatric morphs have evolved locally following

* Corresponding author. Tel.: +44 1334 463440; fax: +44 1334 463443.

E-mail address: iaj@st-and.ac.uk (I.A. Johnston).

¹ Present address: Wellcome Trust Biocentre, Division of Cell and Developmental Biology, University of Dundee, Dundee, DDI 5EH, Scotland, UK.

post-glacial invasions from a postulated single charr lineage (Wilson et al., 2004).

In contrast, common-garden rearing experiments with the progeny of the four morphs from Thingvallavatn suggest genetic differences in morphology, body colour, growth, age at first sexual maturity, and foraging behaviour (Skúlason et al., 1989, 1993; Skúlason and Smith, 1995; Eiriksson et al., 1999). Inherited genetic differences have also been identified in polymorphic Arctic charr species from Loch Rannoch, Scotland (Adams and Huntingford, 2002b,a, 2004). Significant differences in head allometry between benthivorous and planktonivorous morphs has led Adams and Huntingford (2002a,b) to propose that Loch Rannoch morphs are separate species based on Kottelat's criteria (Kottelat, 1997). The differences in head morphology in the Arctic charr morphs in Thingvallavatn are shown in Fig. 1.

The lack of genetic differentiation between the Arctic charr morphs has led to the proposal that only a few regulatory genes may be responsible for the developmental differentiation that accounts for the observed phenotypic differences between morphs (Danzmann et al., 1991). More recent findings in cichlids support this assumption, which show that rapidly speciated lineages acquire a complex regulatory mechanism of alternative splicing over a very short evolutionary period (Terai et al., 2003). In this respect, we postulated that *Pax* proteins were interesting candidates as they are key regulators of early development and also exhibit extensive alternative splicing in mammals (Chi and Epstein, 2002) and latterly the Atlantic salmon (*Salmo salar* L.) (Gotensparre et al., 2006).

The *Pax* family of proteins are highly conserved transcription factors defined by a so-called 'paired box' or paired domain (PD) of 128-amino acids. The PD, first identified in the

Drosophila paired (*prd*) and *gooseberry* genes (Bopp et al., 1986; Frigerio et al., 1986), is a bipartite DNA-binding domain composed of two subdomains (PAI and RED) each consisting of helix-turn-helix (HTH) motifs. Further complexity in some *Pax* proteins is acquired by the addition of a conserved octapeptide (OP), and a *prd*-type homeodomain (HD) that also comprises a HTH motif. The important role of *Pax* proteins is highlighted by various gene mutations, which are linked to cancer (e.g. alveolar rhabdomyosarcoma), human disease (e.g. Waardenburg Syndrome), and physical defects (e.g. aniridia). Such profound effects confirm *Pax* proteins as 'master' controllers (Gehring, 1996; Underhill, 2000) and essential morphoregulators (Tremblay and Gruss, 1994).

In particular, we are interested in the *Pax7* gene as a possible candidate involved in the development/plasticity of polymorphic Arctic charr due to its involvement in cranio-facial, skeletal muscle, and central nervous system development (Mansouri et al., 1996), and myogenic stem cell specification (Seale et al., 2000). To date, no studies have compared gene sequences from different morphs of Arctic charr. Our *Pax7* study is the first to address this, with a detailed examination of this nuclear regulatory gene and its expression patterns in the Arctic charr morphs of Lake Thingvallavatn, Iceland.

2. Materials and methods

2.1. Collection, rearing, and embryo preparation

Arctic charr (*S. alpinus* Linnaeus, 1758) were sampled by gill-nets at various locations in Lake Thingvallavatn, Iceland. Adult fish were collected for analysis of *Pax7* and RNA content, and eggs were also stripped from ovulating females of DB-, LB-,



Fig. 1. The four morphs of Arctic charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland. (a) Planktonivorous morph (PL), (b) Piscivorous morph (PI), (c) Dwarf or small benthivorous morph (DB), and (d) Large benthivorous morph (LB). The two pelagic or limnetic morphotypes (a, b) have terminal mouths, pointed snouts and shorter pectoral fins than the two benthic morphotypes (c, d). Note the blunter snouts and sub-terminal mouths of the benthic morphotypes. Scale bars represent 1 cm.

and PL morphs at the Holar University College Aquarium, Holar, Iceland. For each morph, eggs from a minimum of twenty females were fertilised *in vitro*. The eggs of each female were fertilised with sperm from a unique male of the same morph. After 1 h of hydration in water, eggs from each family were split between two incubation systems supplied with running spring water. Each family was maintained in a separate compartment in the dark. One system was cooled to a near constant 3 °C and the other received water at ambient temperature which ranged from 6.8 to 3.8 °C (decreasing during development). After hatching, larvae were transferred to 1 m³ through-flow aquaculture tanks. The families were kept separated in 500 ml open containers with a 1 mm-mesh bottom at ambient temperature (4–7 °C) and natural photoperiod. Embryos were fixed in 4% (m/v) paraformaldehyde (PFA) in phosphate buffered saline (PBS) pH7.3 at 4 °C for 2 h. Embryos were dechorionated at room temperature (RT) under a dissecting microscope (Leica MZ12) and then washed 3 × 5 min in sterile PBS at RT and dehydrated in methanol.

2.2. Isolation of Pax7 cDNA sequence

Fast muscle was dissected from the dorsal epaxial myotome at approximately 0.7 fork length (FL) from eight adult wild-caught morphs, i.e. two each of LB, DB, PL, and PI. Total RNA and polyA⁺ RNA was extracted using TRI REAGENT (Sigma-Aldrich Company Ltd, Gillingham, UK) and Oligotex (Qiagen, Crawley, UK) respectively. Total RNA (TRI REAGENT method) was also extracted from pre-hatch LB morph embryos of mixed developmental stage (i.e. >50 pooled embryos from 369 h post-fertilisation (hpf) to 610 hpf for the two temperature groups of hatchery reared embryos, corresponding to daily samples from gastrulation to the end of segmentation). RT-PCR was carried out using the OneStep RT-PCR Kit (Qiagen, Crawley, UK) using the following degenerate primers designed from conserved regions (forward and reverse primers for paired box domain and homeodomain respectively) of human, mouse, and zebrafish *Pax7* nucleotide sequence alignments: 5'-GTTCAGCTGGGARATCCGGGAYAAAG-3' (forward) and 5'-TGACACGRGCCTCRGTCAGTTTGG-3' (reverse).

To obtain full-length coding sequences (from start to stop codons) for each morph (embryonic and adult tissue) the RETROscript 2-step RT-PCR was used (Ambion, Huntingdon, UK). Primer sequences were as follows: 5'-ATGGCTACTTACCAGGAACAG-3' (forward: including start codon) and 5'-GTAGGCCTGCCCGTCTCC-3' (reverse: not including stop codon). Final PCR products were purified by agarose gel electrophoresis and Montage DNA Gel Extraction (Millipore Corporation, MA, USA). cDNAs were cloned into the pCR4-TOPO vector using the TOPO TA Cloning system (Invitrogen Ltd, Paisley, UK) and transformed into One Shot TOP10 competent *Escherichia coli* (Invitrogen Ltd, Paisley, UK). Samples were analysed for 2 individuals per morph and 2 pooled embryo samples from the LB morph, representing daily samples from gastrulation to the end of segmentation. For individual adult fish 2–4 clones were sequenced per morph. All clones were sequenced at least twice on both strands.

2.3. Rapid amplification of cDNA ends (5' RACE only)

Clontech's SMART RACE (Clontech Laboratories UK Ltd) system was used to isolate the 5'-end of Arctic charr *Pax7* from LB morph pooled embryos of mixed developmental stage. First strand cDNA synthesis was carried out using 1 µg of polyA⁺ mRNA from muscle of adult stages, or 1 µg of total RNA from embryonic stages, according to the manufacturer's instructions. Primary PCR amplification was carried out using SMART UPM primer (Clontech supplied) and a reverse gene-specific primer (GSP) located in the homeodomain of *Pax7*: 5'-CCTGGTGTA GATGTCTGGGTAGTGTG-3'. Secondary PCR amplification was carried out using the same PCR conditions and nested primers as follows: SMART NUPM primer (kit supplied) and 5'-CGGCTGTGAAGGTGGTGC GACTGC-3' (reverse GSP). Secondary PCR products were purified, TA cloned and sequenced as described previously.

2.4. PCR amplification of Pax7 intronic sequences

Genomic DNA (gDNA) was extracted from LB morph muscle tissue using a protocol provided by Ambion (Huntingdon, UK) for extraction of gDNA from tissue stored in RNAlater (Ambion, Huntingdon, UK). gDNA for other Arctic charr morphs (DB, PL, and PI) was extracted from kidney using DNeasy (Qiagen, Crawley, UK). Intron 1 was PCR amplified from gDNA with exonic flanking primers 5'-ATGGCTACTTTACCAGGAACAG-3' (forward: exon 1) and 5'-GGCCTCCCGTTGATGAA-3' (reverse: exon 2) using standard PCR conditions. Similarly, intron 3 was amplified with primers: 5'-CCGGGATGTTTCAGTTGGGAGAT-3' (forward: exon 3) and 5'-CTATCGTCGTCGT-CATCTTTCTTG-3' (reverse: exon 4).

2.5. In silico analysis

Base calling and sequence assembly were performed using SeqMan (DNASTAR, Inc., Madison, WI, USA). Consensus sequences were submitted to nucleotide and protein databases at the European Bioinformatics Institute (EBI) and compared using WU BLAST Version 2.0 (Gish, 1996–2003). Multiple alignments were generated using the Clustal method (Higgins et al., 1992) in Megalign (DNASTAR, Inc., Madison, WI, USA) and refined manually. The Ensembl database of annotated zebrafish sequence (http://www.ensembl.org/Danio_rerio) was used to obtain annotated exon/intron borders for the zebrafish *Pax7* gene (exons 1–4). Zebrafish *Pax7* exonic sequences were aligned with Arctic charr *Pax7* genomic sequence to delineate exon/intron borders. Pattern searching of Arctic charr *Pax7* protein sequences was performed using the PROSITE database (Hofmann et al., 1999).

2.6. Whole mount in situ hybridisation

The 5' RACE *Pax7* clone (1.197 kb) was used for preparation of digoxigenin-labelled cRNA riboprobes using standard procedures. *In situ* hybridisation (ISH) was carried out using modified procedures (Wilkinson, 1992; Ennion et al., 1999; Hall

Table 1
Alternate *Pax7* isoforms of Arctic charr morphs indicating insertions/deletions in the *Pax7* protein

	Accession nos.	Morph	Source	Method	13 aa ins.	Q ins.	5 aa ins.	4 aa del.	CDS	5' UTR
<i>Pax7a</i>	AJ634764	LB	Embryonic	RT-PCR	–	+	–	+	497	–
<i>Pax7b</i>	AJ634765	LB	Embryonic	RT-PCR	–	+	+	+	502	–
<i>Pax7c</i>	AJ634766	LB	Adult	RT-PCR	+(T)	–	–	–	513	–
<i>Pax7d</i>	AJ634767	LB	Adult	RT-PCR	+(T)	+	–	+	510	–
<i>Pax7e</i>	AJ634768	LB	Embryonic	5' RACE	+(A)	+	+	–	519	470
<i>Pax7f</i>	AJ634769	LB	Embryonic	5' RACE	+(A)	+	+	+	515	342
<i>Pax7g</i>	AJ634770	LB	Adult	5' RACE	+(A)	+	+	+	515	470
<i>Pax7h</i>	AJ634771	DB	Adult	RT-PCR	–	+	+(GEAS)	–	505	–
<i>Pax7i</i>	AJ634772	DB	Adult	RT-PCR	+(A)	+	–	+	510	–
<i>Pax7j</i>	AJ634773	PI	Adult	RT-PCR	+(T)	+	–	–	514	–
<i>Pax7k</i>	AJ634775	PL	Adult	RT-PCR	–	+	+(GEAS)	+	501	–

Key: aa — amino acid residues, ins. — insertion, del. — deletion, Q — glutamine, CDS — total no. of amino acid residues in coding sequence, UTR — total no. of nucleotides in untranslated region, LB — large benthic, DB — dwarf benthic, PI — piscivorous, PL — planktonivorous. Protein sequence of insertions/deletions: 13 aa — GQY[A/T]GPEYVYCGT (version A or T is indicated in parentheses), 5 aa — GEAS (truncated insertion, GEAS, is indicated in parentheses), and 4 aa — GNRT. Note that the length of coding sequence for 5' RACE clones is projected from the combined analysis with RT-PCR full-length clones.

et al., 2004). Embryos from three Arctic charr morphs (LB, DB, and PL) at stages from 50% epiboly to post-somitogenesis were used for *in situ* hybridisation. For a negative control an Arctic charr *Pax7* sense probe was used. All procedures were performed in RNase-free conditions and hybridisations were carried out in a HM-4000 Multidizer hybridisation oven (Ultra-Violet Products Ltd, Cambridge, UK). Photographs were taken on binocular microscopes: (i) Leica Wild M3C (Milton Keynes, UK) and (ii) Leica Leitz DMRB (Milton Keynes, UK), using a Nikon CP4500 digital imaging system (Visual Inspection Technology LLP, Cardrona, Scotland, UK).

3. Results

3.1. Isolation of *Pax7* cDNA from Arctic charr morphs

OneStep RT-PCR using degenerate primers and LB embryonic total RNA produced several bands only one of which was homologous to zebrafish *Pax7* resulting in a sequence of 432 bp. This sequence enabled the design of charr-specific primers for 5' RACE producing a sequence of 1197 bp ranging from 5' untranslated region (UTR) to the N-terminal region of the homeodomain. PCR amplification of full-length *Pax7* (start to stop codons) from embryonic and adult RNA (all morphs) resulted in products of approximately 1.5 kb. BLAST analyses of Arctic charr LB full-length *Pax7* sequences identified homology to the zebrafish *Pax7c* gene. *Pax7* sequences for the remaining Arctic charr morphs (DB, PL, and PI) were isolated by RT-PCR from adult muscle tissue only. Products of approximately 1.5 kb

were obtained for these morphs which were also homologous to zebrafish *Pax7c*. Except for a small number of randomly occurring single nucleotide polymorphisms (SNPs), *Pax7* nucleotide sequences were identical for all Arctic charr morphs, but differences in transcripts were observed due to the presence or absence of known and unknown insertions/deletions indicating multiple *Pax7* splice variants.

The longest *Pax7* sequence for the LB morph comprises an ORF of 2027 nucleotides encoding a protein of 519 residues (56 kDa). The longest 5' UTR is 470 bp (Fig. 2), and one 5' RACE clone (*Pax7f*) contained a 128 bp deletion within the 5' UTR (underlined in Fig. 2).

3.2. Structural characterisation of *Pax7* transcripts from Arctic charr morphs

A multiple alignment of vertebrate *Pax7* proteins (charr, zebrafish, chicken, mouse, and human) is presented in Fig. 2, and includes the zebrafish *Pax3* protein. The coding region of Arctic charr *Pax7* exhibits all the signature features of *Pax* proteins, including a paired box domain (PD), a complete paired-type homeodomain (HD), and an octapeptide (HSIDGILG). There were no differences in the coding sequence of *Pax7* clones isolated from the different morphs. However, sequence analysis demonstrated multiple putative *Pax7* isoforms as reported for mouse (Jostes et al., 1990; Ziman et al., 1997; Ziman and Kay, 1998; Ziman et al., 2001), human (Schafer et al., 1994), zebrafish (Seo et al., 1998), and latterly Atlantic salmon (Gotensparre et al., 2006). Putative *Pax7* isoforms with known

Fig. 2. Multiple sequence alignment of vertebrate *Pax7* proteins including zebrafish *Pax3*. The consensus is the deduced amino acid sequence of *Pax7* cDNA from Arctic charr (*Salvelinus alpinus*) LB morph. The consensus is shown as a single-letter code with identical residues and gaps versus orthologues indicated by dots and dashes respectively. From top to bottom the signature features of *Pax7* are highlighted as follows: paired box domain (PD), octapeptide (OP), and homeodomain (HD). Boxed residues indicate insertion/deletion locations, which determine alternatively spliced transcripts. Included above the multiple alignment are nucleotides (lower case) for Arctic charr *Pax7* 5' untranslated region (UTR). The 128 bp deletion in the 5' UTR of one 5' RACE clone (*Pax7f*) is underlined. Exon boundaries are indicated for the first five exons as determined from the Ensembl database for the annotated zebrafish *Pax7* gene. A novel variation in the five-residue insertion (i.e. GEAS, a truncation of the first serine) is indicated by bold italic. Variations in the novel thirteen-residue insertion, i.e. alanine (A) and threonine (T) variants (GQYAGPEYVYCGT or GQYTGPEYVYCGT) are indicated in bold with the threonine residue indicated above the box. The threonine variant introduces a putative casein kinase II (CK-2) phosphorylation site (S/T-XX-D/E) within the insertion (TGPE) which is lost in alanine variants. Accession nos.: zebrafish *Pax7c* (O57418), chicken *Pax7* (O42349), mouse *Pax7* (Q9ES16), human *Pax7* (P23759), zebrafish *Pax3* (O57416).

insertions/deletions in the paired box were identified in Arctic charr morphs as follows: (i) a single glutamine (Q) insertion at position 75 (relative to the start of the PD), and (ii) a five-residue (GEASS) insertion at the likewise relative position of 118. The identification of a 75Q deletion in one adult clone (Arctic charr *Pax7c*) is a novel finding compared to zebrafish where no 75Q deletions were detected (Seo et al., 1998). Versions of the GEASS insertion were identified for two Arctic charr morphs (DB and PL) that were truncated by a single residue (GEAS) as latterly reported in Atlantic salmon (Gotensparre et al., 2006). Although the corresponding GEAS clones for the LB and PI morphs were not obtained from mRNA, the presence of the GEAS insertion was verified in these morphs from genomic DNA sequence (see Section 3.3).

Other insertions/deletions within the *Pax7* protein were also identified in Arctic charr morphs as follows: (i) a thirteen-residue insertion (GQYAGPEYVYCGT or GQYTGPEYVYCGT) located at position 29 (number of residues from the initiator methionine) and N-terminal to the PD, and (ii) a four-residue (GNRT) deletion located at position 212 and C-terminal to the octapeptide. Pattern searching with PROSITE revealed a protein motif present in the thirteen-residue insertion that contained the threonine (T) variant. This variant introduces a casein kinase II (CK-2) phosphorylation site (S/T-XX-D/E) within the insertion (TGPE) which is lost in alanine (A) variants.

Table 1 lists the putative alternate isoforms (including Accession nos.) identified to date (*Pax7a* to *Pax7k*) exhibiting

different combinations of the four types of insertion/deletion. Although the length of the coding sequence for 5' RACE clones is projected from the combined analysis with RT-PCR full-length clones, all 5' RACE sequences encompassed the four types of insertion/deletion and were 100% identical for overlapping nucleotides in the contig (726 bp). Arctic charr *Pax7f* and *Pax7g* are identical in coding sequence except for the 128 bp 5' UTR deletion in the former. Therefore, precluding either *Pax7f* or *Pax7g*, a total of ten putative alternate isoforms have been identified for the four types of insertion/deletion out of sixteen theoretical isoforms. Although not all possible combinations of insertion/deletion were identified in each morph from mRNA, nucleotides coding for two insertions were confirmed by genomic DNA sequence analysis in each morph, i.e. the GEA(S)S and thirteen-residue insertions (see Section 3.3). A schematic representation of Arctic charr *Pax7* isoforms is shown in Fig. 3.

3.3. Genomic organisation of *Pax7* insertions

Introns 1 and 3 were PCR amplified to identify alternative splice sites for the thirteen-residue and five-residue insertions in Arctic charr *Pax7* variants. Alignments of genomic sequences indicated nucleotides for the insertions as shown in Fig. 4a and 4b respectively. Splice donor (GT) and acceptor (AG) nucleotides are present and sequences generally adhere to the donor/acceptor consensus for intron/exon boundaries (Mount, 1982; Senapathy

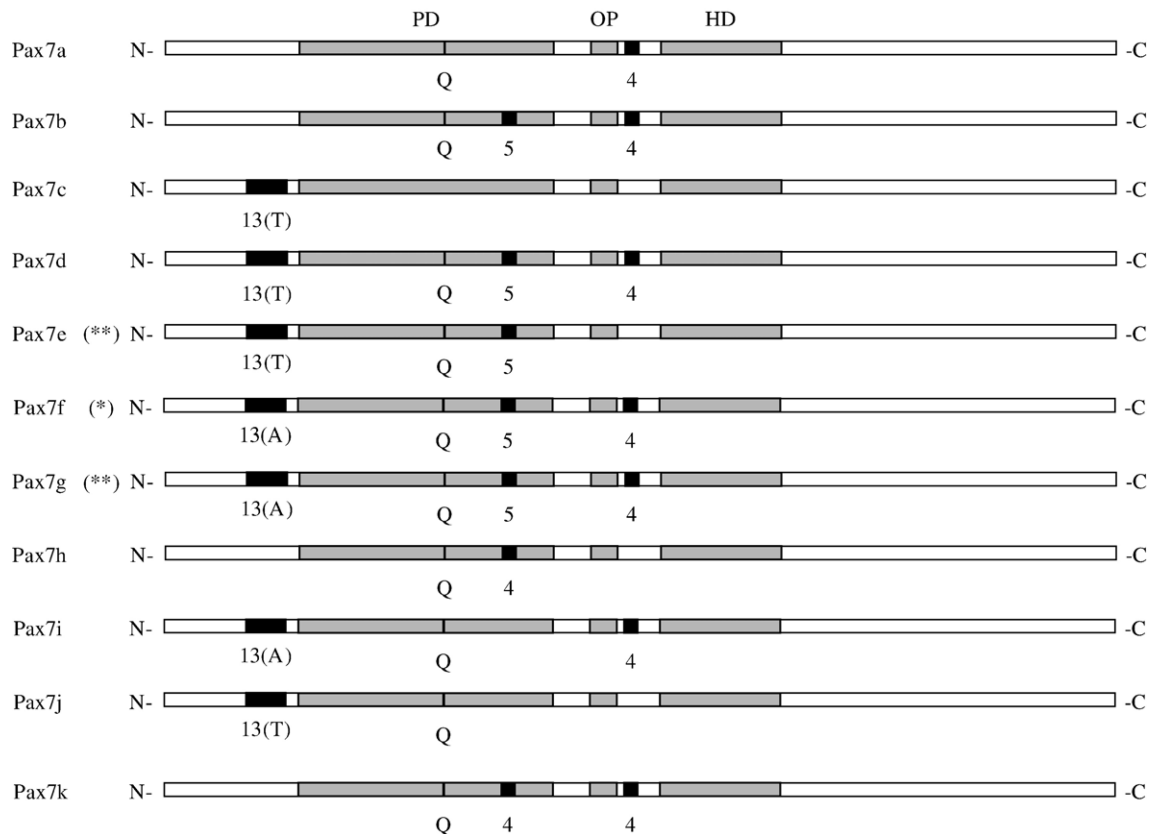


Fig. 3. Schematic overview of Arctic charr *Pax7* putative isoforms. Abbreviations: N-, Amino terminal, C-carboxy terminal, * 342 bp 5' UTR, ** 470 bp 5' UTR. Other abbreviations as for Table 1.

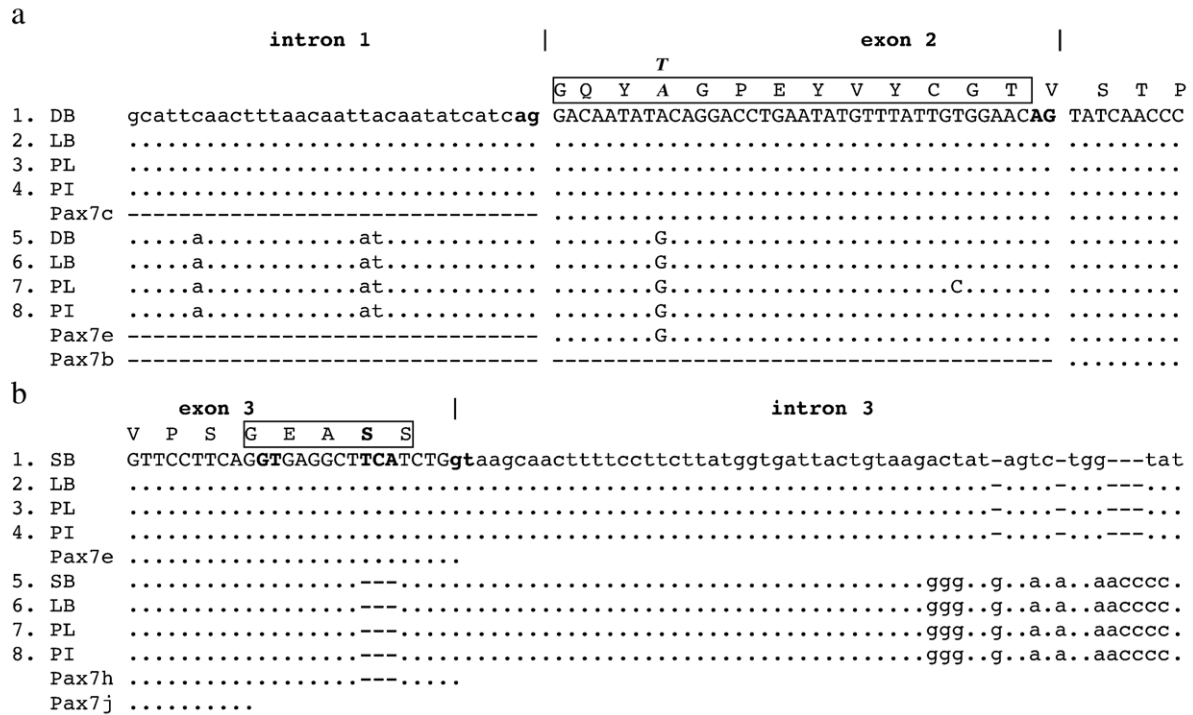


Fig. 4. Alignments of *Pax7* gene sequences from Arctic charr morphs (DB, LB, PL, and PI) depicting the genomic regions for the novel splice variants or insertions. Nucleotides for intronic and coding region sequences are in lower and upper case respectively. Dots and dashes indicate identical nucleotides and gaps respectively. Intron/exon splice sites for variants are indicated by '|', and splice site acceptors (ag) and donors (gt) are indicated in bold. The deduced amino acids for coding sequence are shown as single-letter code immediately above the nucleotide sequence, and novel insertions are boxed. (a) Thirteen-residue insertion alignment. Alanine (A) and threonine (T) variants in the insertion are in bold italic with the threonine residue indicated above the box. Genomic DNA sequences (numbered 1–8) identified two putative copies of *Pax7* in the genome of Arctic charr, which are shown for each morph (i.e. 1–4 and 5–8). Intron sizes to the first splice acceptor are 1329 bp and 1300 bp for copies 1–4 and 5–8 respectively. Each copy identified a single nucleotide change in the 1st codon position accounting for alanine and threonine variants, i.e. GCA and ACA respectively. Three mRNA transcripts are included in the alignment to highlight the three types of splice variants identified from this genomic region, i.e. insertion with threonine (*Pax7c*), insertion with alanine (*Pax7e*), and no insertion (*Pax7b*). (b) GEA(S)S insertion alignment (format as for Fig. 4A). Intron sizes from the second splice donor are 246 bp and 218 bp for copies 1–4 and 5–8 respectively. The truncated GEAS splice variant is due to TCA deletion on the interval for the 218 bp intronic copy. Three mRNA transcripts are included, i.e. insertion with TCA codon (*Pax7e*), insertion with TCA deletion (*Pax7h*), and no insertion (*Pax7j*).

et al., 1990). Nucleotides for the thirteen- and five-residue insertions are located at the 5'-end of exon 2 and 3'-end of exon 3 respectively.

Intron sizes were identical between morphs, and nucleotide sequences only differed with respect to a very few randomly occurring SNPs. Two sequences were obtained for intron 1 and two sequences for intron 3 for each Arctic charr morph. The sizes of these intronic sequences (not including nucleotides for the insertions) were 1300 bp and 1329 bp for intron 1, and 218 bp and 246 bp for intron 3. Pairwise alignments of the two distinct intronic sequences showed percentage nucleotide identities of 73.6% (1043/1418 nucleotides including gaps) for intron 1, and 71.9% (179/249 inc. gaps) for intron 3. The variation between these intronic sequences is partially shown in Fig. 4b, where the first 38 bp of intron 3 are identical but thereafter the two intronic sequences diverge. The differences between intronic sequences are mainly due to interspersed repeats and micro-insertions/deletions.

For the thirteen-residue insertion, nucleotides coding for alanine and threonine variants were identified by a single nucleotide change in the 1st codon position, i.e. GCA and ACA respectively. GCA codons were located on the 1300 bp intronic copy, and ACA codons on the 1329 bp copy. The truncated

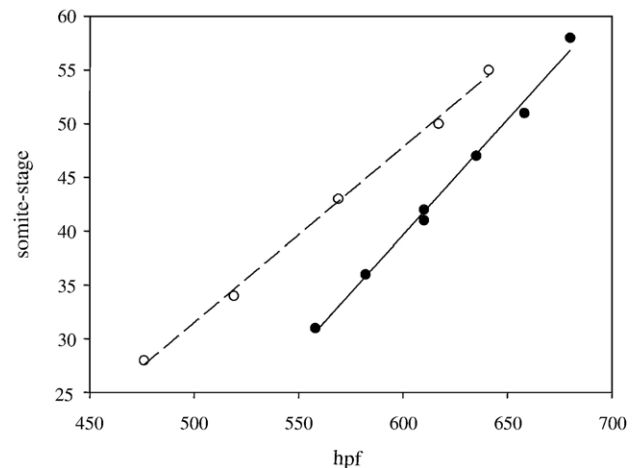


Fig. 5. The relationship between somite stage and age hours post-fertilization (hpf) for large benthic (solid circles) and dwarf benthic (open circles) morph embryos of the Arctic charr (*Salvelinus alpinus* L.). A first order linear regression was fitted to the data with the following equations: Large benthic morph: somite stage = -89.24 + 0.21 * hpf; (Rsq=0.99; $F_{1, 6} = 713.2$; $P < 0.0001$); Dwarf benthic morph: somite stage = -50.04 + 0.16 * hpf; (Rsq=0.99; $F_{1, 4} = 1249.4$; $P < 0.0001$).

GEAS splice variant was due to a trinucleotide (TCA) deletion resulting in the absence of the first serine residue. TCA codons were located on the 246 bp intronic copy, and TCA deletions on the 218 bp copy.

3.4. Somitogenesis

The relationship between somite stage and the age of the embryo (hpf) for the LB and DB morphs is shown in Fig. 5. The time required to form one somite pair was 285 min for the LB

morph and 375 min for the DB morph (one-way ANOVA with post-fertilisation time as a covariate; $F_{1,9}=59.95$; $P<0.0001$). Segmentation started at a significantly earlier age in the DB morph than the LB morph, although it was completed at approximately the same time.

3.5. Embryonic expression of Arctic charr *Pax7*

Arctic charr *Pax7* spatial expression patterns were similar for the morphs investigated (DB, LB, and PL). *Pax7* transcripts

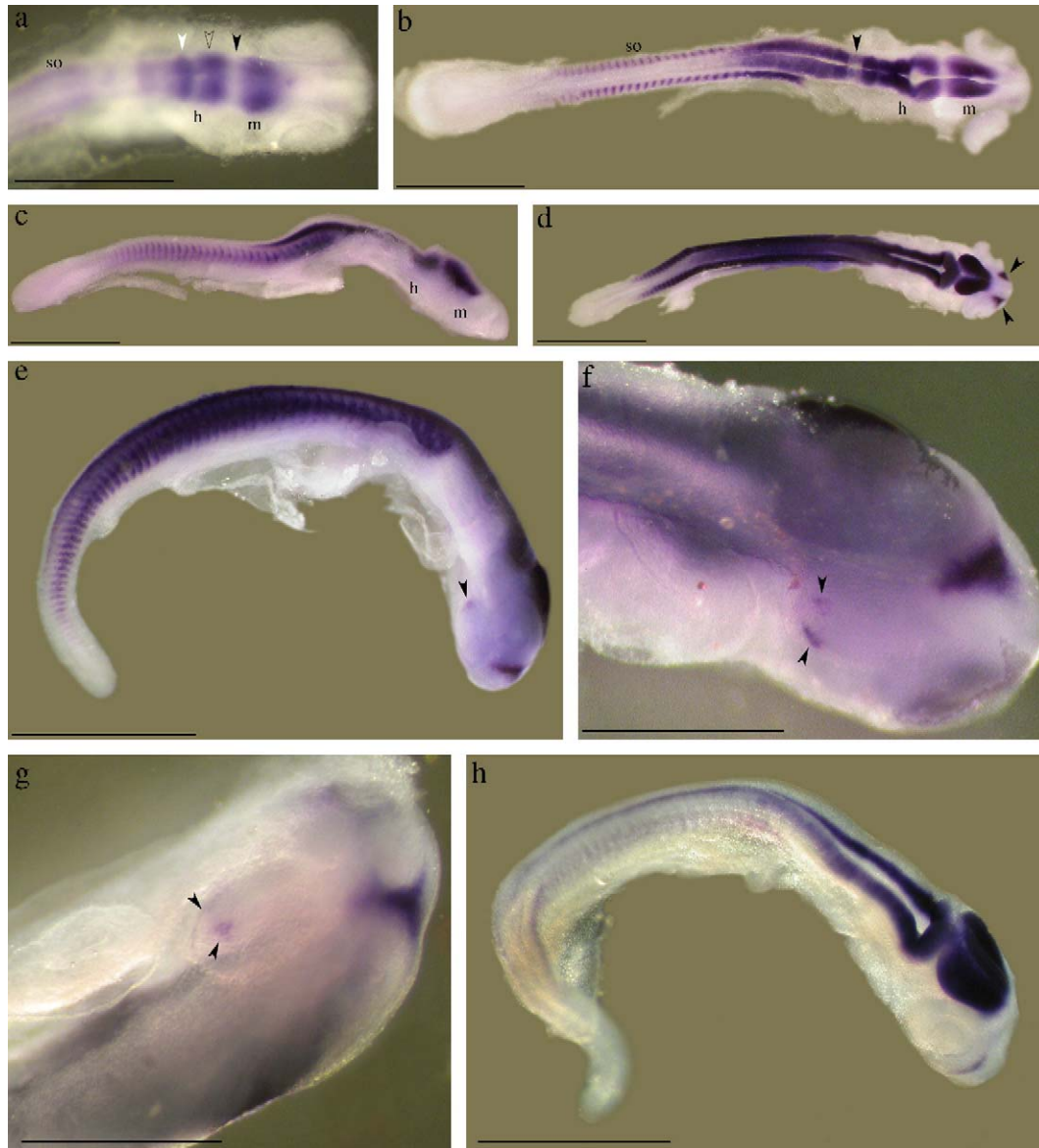


Fig. 6. Spatial expression patterns of Arctic charr *Pax7* transcripts in dwarf benthic (DB) and large benthic (LB) morphs. A composite of only two morphs is shown for brevity, but all morphs were assessed at all times and no differences between morphs were observed. Embryos are shown with anterior to the right. (a) Dorsal view of DB morph at 27 somites (476 hpf). Filled arrowhead (black) indicates isthmus of decreasing *Pax7* expression that separates midbrain and hindbrain boundary. Open arrowhead indicates the primordial cerebellum. Filled arrowhead (white) indicates the presumptive first rhombomere primordia. (b) Dorsal view of LB morph at 35 somites (582 hpf). Arrowhead indicates the rhombomere with reduced *Pax7* staining equivalent to observations in zebrafish *Pax7c* identified as rhombomere 6 (r6). (c) Lateral view of LB morph at 41 somites (610 hpf). (d) Dorsal view of LB morph at 47-somites (635 hpf). Arrowheads indicate *Pax7* transcripts in forebrain/frontonasal region. (e) Latero-ventral view of LB morph at 62-somites (704 hpf). Arrowhead indicates diffuse *Pax7* expression in mandibular region. (f, g) Ventral views of head of LB (728 hpf) and DB (787 hpf) embryos respectively. Arrowheads indicate two bilateral domains of *Pax7* expression in mandibular region. (h) Lateral view of DB morph after completion of segmentation showing gradual decline of *Pax7* expression in the somites. Other abbreviations: h, hindbrain; m, midbrain; so, somites. Scale bars: b, c, and d represent 50 μm ; a, e, f, g, and h represent 100 μm .

were first detected in Arctic charr morphs as a diffuse single transverse stripe in the developing brain at approximately 50% epiboly (not shown). Subsequent expression of Arctic charr *Pax7* was similar to that described for the zebrafish *Pax7c* gene (Seo et al., 1998). During early neurulation Arctic charr *Pax7* expression increased and widened in the dorsal half of the presumptive midbrain, but as with zebrafish *Pax7c*, expression boundaries and morphological subdivisions were poorly defined at early stages (not shown). Fig. 6 shows a composite of *Pax7* expression for DB and LB morphs from 476 to 787 hpf. The first appearance of segment-like domains was observed as transverse bands (Fig. 6a) highlighting the presumptive midbrain–hindbrain boundary. In the hindbrain, the primordial cerebellum and the first presumptive rhombomeres were observed. Low intensity staining was observed in the somites. At 35-somites (582 hpf) in the LB morph (Fig. 6b), *Pax7* expression was detected bilaterally along the rostrocaudal axis and rhombomeric boundaries were more defined. *Pax7* expression extended caudally in the spinal cord and somites,

but with decreasing signal intensity towards the tail bud. In the most caudal nascent somites (about 5–6 in number) *Pax7* transcription was not observed. Viewed at high magnification with DIC optics the expression of *Pax7* transcripts was observed across the somite width (so) in the lateral medial compartment of anterior somites (Fig. 7a), but was restricted to the rostral domain in somites towards the tail bud (Fig. 7b). This expression pattern continued in subsequent stages, but with a gradual increase in signal intensity rostrally and with further definition of neuromeres (Fig. 6c,d). The bilateral stripes in the midbrain were further extended laterally and appeared lobe-shaped by the 47-somite stage (Fig. 6d). The cerebellum was also lobe-shaped to a lesser extent but extended dorsoventrally. Strong *Pax7* expression was observed bilaterally in the forebrain/frontonasal regions. The posterior boundary of these regions was located slightly behind the eyes and extended anteriorly towards the nose and presumptive fronto-maxillary areas. At the end of segmentation (62-somites, Fig. 6e) a diffuse band of *Pax7* expression was observed in the developing lower

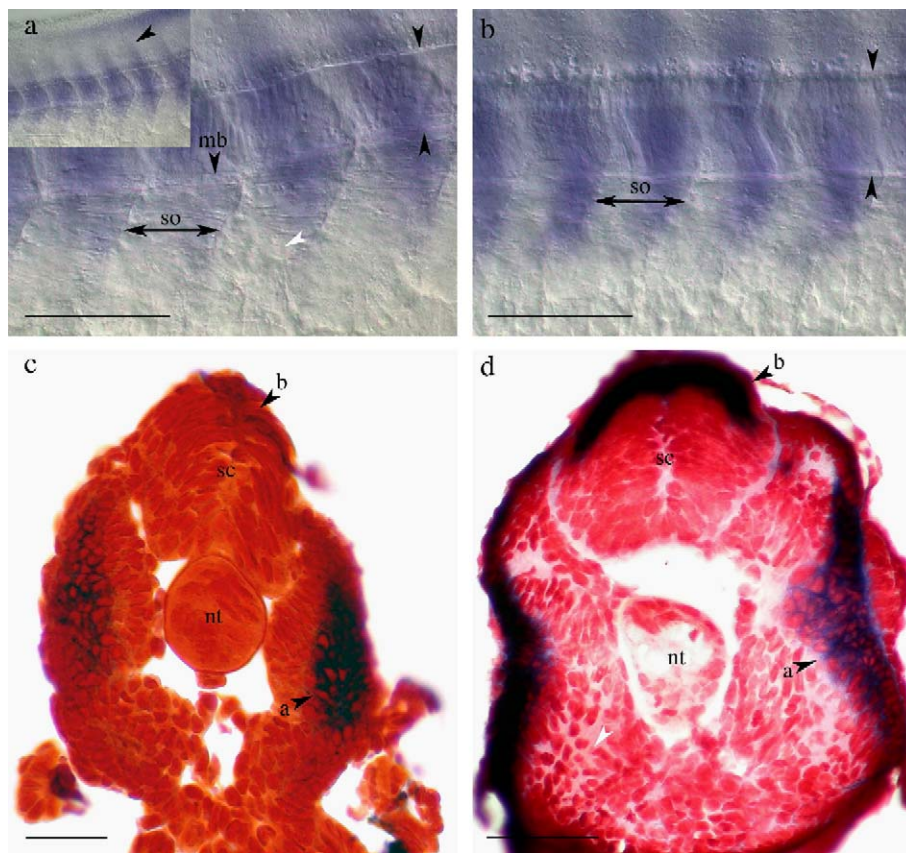


Fig. 7. (a,b) Whole mount *in situ* hybridisation of a 35-somite stage LB morph with *Pax7* cRNA probe viewed with Differential Interference Optics (DIC). The insert in (a) illustrates the position of the main image, with the arrowhead indicating the most rostral somite in the main image. The staining at the top in the insert corresponds to the spinal cord. Image (b) was taken 10 somites more posterior to image (a). Staining for *Pax7* transcripts was observed across the somite width (so) in the lateral medial compartment of somites (illustrated in Fig. 7a), but was restricted to the rostral domain in somites towards the tail bud (illustrated in Fig. 7b). The inward facing solid arrowheads show the position of the notochord. The white arrowhead illustrates a myotube. (c, d) Transverse wax embedded sections of whole mount *in situ* hybridisation of LB morphs stained with a *Pax7* cRNA probe and counterstained with Fast Red. In (c), which corresponds to somites 25–30 of a 42-somite stage (610 hpf) embryo, the *Pax7* transcripts were expressed in the lateral and medial regions of the myotome (arrow head 'a'). (d) Corresponds to the rostral somites at the end of segmentation in an embryo at 751 hpf. The arrowhead 'a' illustrates the intense staining for *Pax7* transcripts in the lateral margin of the myotome extending along the position of the major horizontal septum. The arrowhead 'b' shows intense staining in the dorsal region of the spinal cord, and the equivalent weakly stained region in the more posterior somites of the 610 hpf embryo is illustrated by 'b' in Fig. 6c. Other abbreviations: nt: notochord; sc: spinal cord. Scale bars represent 100 μ m.

jaw/mandibular region. In later stages, this expression appeared as two distinct bilateral domains (Fig. 6f, g). After segmentation, *Pax7* expression in the somites gradually declined (Fig. 6h). Transverse sections of wax embedded whole mounts revealed that the expression of *Pax7* transcripts in the muscle was restricted to the lateral and medial domains of the somite both during and at the end of the segmentation period (Fig. 7c, d). Staining for *Pax7* transcripts in the spinal cord was restricted to cells at the dorsal margin (labelled 'b' in Fig. 7d).

4. Discussion

The aim of this study was to investigate *Pax7* gene as a candidate involved in the intraspecific polymorphism of Arctic charr (*S. alpinus*) from Lake Thingvallavatn, Iceland. Orthologues of zebrafish *Pax7c* were identified in all Arctic charr morphs. A common feature of *Pax7* genes is the generation of different transcripts through alternative splicing (Kay et al., 1995; Ziman et al., 1997; Kay et al., 1998; Seo et al., 1998; Ziman and Kay, 1998; Barr et al., 1999; Kay and Ziman, 1999; Ziman et al., 2001; Gotensparre et al., 2006), a feature confirmed for the *Pax7* gene in Arctic charr. We identified at least 10 putative alternate *Pax7* isoforms (Fig. 3), which varied according to the presence and/or absence of known and novel insertions/deletions. Isoforms therefore ranged in size from 497 to 519 residues. Known variants that were identified included a single glutamine residue (Q+) at position 75 of the PD (Vogan et al., 1996; Ziman and Kay, 1998), and a GEASS insertion (Seo et al., 1998). The single glutamine residue is located in the linker region, which separates the DNA-binding subdomains PAI and RED. Studies have shown that the Q+/Q- isoforms influence DNA-binding, with the Q- isoform binding to subdomain consensus sequences with a two- to fivefold-higher affinity, whereas the introduction of a glutamine causes a fivefold reduction (Vogan et al., 1996). These studies indicate that the PD linker region is important in DNA target selection, and that such isoforms potentially mediate specific functions during embryogenesis. The Q- variant appears to be a basal character as there are no corresponding Q+ insertions in the amphioxus orthologue *AmphiPax3/7* (Holland et al., 1999). Thus, in the mouse and zebrafish, Q+ is the predominant isoform with Q+/Q- ratios of 2:1 (*Pax7*) and 3:2 (*Pax3*) for these taxa respectively. In our study, the predominant transcripts was also Q+, but one of the ten identified revealed a Q- variant (*Pax7c*, Table 1). Therefore, Q- variants are present in fish, but such transcripts appear to be relatively rare or confined to a brief developmental stage.

The five-residue GEASS insertion has previously been reported in all four of the zebrafish isoforms (Seo et al., 1998). The equivalent insertion in human *Pax7* protein (Schafer et al., 1994) has only two-residues (GL; see Fig. 2). Similar to zebrafish, the presence of the GEASS insertion is due to an alternative 3' splice site upstream of exon 3. We identified a truncated version of this insertion in Arctic charr, which is missing the first serine residue (GEAS) as latterly reported for Atlantic salmon (Gotensparre et al., 2006). Structurally, the GEA(S)S insertion is located between helices 5 and 6 of the

RED (C-terminal) domain. Studies of HTH domains suggest that the 'turn' between these helices can tolerate insertions, and are more flexible (Xu et al., 1995). *Pax* proteins have the ability to utilise all domains in the protein (PAI, RED, and HD), or multiple combinations thereof, to achieve different DNA-binding specificities for activation of specific target genes (Bertuccioli et al., 1996; Jun and Desplan, 1996; Jun et al., 1998). This combinatorial DNA-binding model allows flexibility at different stages of development for regulating various developmental processes (Jun and Desplan, 1996). As reported by Seo et al. (1998), it is unknown whether the GEA(S)S insertions affect DNA-binding specificity, however, their presence does increase the diversity of variants and different biological functions. It would be interesting to assess the effect of GEAS on the secondary structure of the isoform relative to the insertion of GL by structural studies. If these two protein modifications lead to similar effects on structure yet arise by separate modifications of the gene, then this would represent an interesting evolutionary phenomenon.

The following novel deletions/insertions within the *Pax7* protein were also identified in Arctic charr morphs: (i) a four-residue (GNRT) deletion, and (ii) a thirteen-residue insertion (GQYAGPEYVYCGT or GQYTGPEYVYCGT). Briefly, the four-residue (GNRT) deletion, located C-terminal to the octapeptide has not been reported in previous studies, although the *AmphiPax3/7* (Holland et al., 1999) protein exhibits a three-residue gap at the equivalent position (not shown). This deletion lies at the 3'-end of exon 4, and its close proximity to the conserved octapeptide may have some functional significance.

The thirteen-residue insertion is due to an alternative 5' splice site upstream of exon 2. Variation in this insertion occurs at the fourth residue in the sequence, where either an alanine (A) or a threonine (T) is present. Interestingly, threonine variants introduce a putative casein kinase II (CK-2) phosphorylation site (S/T-XX-D/E) within the insertion (TGPE) which is lost in alanine (A) variants. Further studies are required to ascertain if this 'in silico-identified' phosphorylation site is functional. Its N-terminal location to the PD is favourable, considering that this position would be less likely to interfere with DNA-binding, and would possibly be accessible to the CK-2 enzyme. CK-2 is a ubiquitously expressed serine/threonine protein kinase necessary for progression of the cell division cycle and cell viability (Litchfield, 2003). The importance of this enzyme in cellular functions is evident from observations that it phosphorylates as many as 300 known proteins (Pinna, 1990). Studies in carp, for example, have shown that CK-2 expression results in rearrangement of nucleolar components, leading to significant changes in the level of ribosomal RNA synthesis (Vera et al., 2000).

Pax7 gene coding sequences were identical between morphs, however, splice sites accounting for the Arctic charr *Pax7* A/T variants and the serine truncation were identified on separate intronic sequences (Fig. 4). This suggests two putative copies of the *Pax7* gene in the Arctic charr genome as proposed for *Salmo salar* (Gotensparre et al., 2006), although this would have to be confirmed by further genomic analyses (e.g. Southern blot, FISH). Two *Pax7* copies would be consistent with the

tetraploid status of salmonids (McKay et al., 2004) in which other gene pairs have been reported, for example, prolactin (Yasuda et al., 1986), MyoD (Rescan and Gauvry, 1996), somatostatin (Moore et al., 1999), and myostatin (Rescan et al., 2001). Such gene pairs have arisen from a whole-genome duplication in ancient ray-finned fishes (Actinopterygii), for which there is now strong evidence (Christoffels et al., 2004; Jaillon et al., 2004; Dehal and Boore, 2005).

Intronic sequencing for polymorphic Arctic charr *Pax7* showed two intronic copies for each morph (i.e. two intron 1 copies and two intron 3 copies) that were identical between morphs except for a few single nucleotide polymorphisms (SNPs). Pairwise alignments of the two intron 1 copies showed identity of 73.6%, and the two intron 3 copies showed 71.9%, and differences between copies were mainly composed of interspersed repeats and micro-insertions/deletions. The presence of two intronic copies may reflect an ancient duplication event. However, the lack of sequence variation between equivalent copies between morphs further emphasises that the genomes of each morph lack genetic differentiation as highlighted by Danzmann et al. (1991). Thus, it may be changes in control regions of key genes that are responsible for the observed phenotypes of Arctic charr morphs. This hypothesis is supported by recent observations in the three-spine stickleback, *Gasterosteus aculeatus*, where the *Pitx1* protein showed no changes in protein sequence between pelvic and pelvic-reduced morphs, but demonstrated site-specific regulatory changes in *Pitx1* expression (Shapiro et al., 2004). This study highlighted that mutations in the regulation of major developmental control genes may provide a mechanism for generating rapid changes in natural populations.

In our study, however, *Pax7* spatial expression patterns were identical in embryos of the DB-, LB-, and PL-morphs, and were similar to those described for zebrafish *Pax7c*. But, a difference in temporal expression patterns was observed for the segmentation period that started at a significantly earlier age in the DB morph than the LB morph (Fig. 5). Interestingly, the DB morph has also been shown to have a reduction in muscle fibre number compared to LB and PL morphs, with potential savings in the energy costs of ionic homeostasis (Johnston et al., 2004).

Throughout the segmentation period, *Pax7* expression generally proceeded in a rostral–caudal direction with expression observed in the midbrain, hindbrain, rhombomeres, spinal cord, and somites. Interestingly, the expression of *Pax7* transcripts was restricted to the lateral and medial domains of the somite and did not correspond to the fields of developing slow or fast muscle fibres (Fig. 6a–d). By the 47-somite stage, strong *Pax7* expression was observed bilaterally in the forebrain/frontonasal regions (Fig. 6d). *Pax7* expression in these areas has previously been reported in the mouse, where its expression has been detected in the nasal pit and nasal neuroepithelium (Jostes et al., 1990). In a recent study of murine limb-muscle development, where *Pax3* is replaced by *Pax7*, it is notable that *Pax7* is also significantly expressed in the frontonasal region (Relaix et al., 2004). Moreover, *Pax7*^{-/-} mutant mice exhibit malformations in facial structures involving the maxilla and nose (Mansouri et al., 1996). Mansouri and

co-workers suggest that these malformations could be related to defects in neural crest cell derivatives, and conclude that *Pax7* regulates the development of a subpopulation of these cells. Thus, it appears that *Pax7* spatial expression patterns are conserved in the above areas from higher to lower vertebrates such as fish. At the end of segmentation, we identified novel expression of *Pax7* in the lower jaw/mandibular region as two distinct bilateral domains. *Pax7* expression in this region in the zebrafish was not reported by Seo et al. (1998).

The pointed snout of the pelagic morphotypes is thought to represent the ancestral condition of Arctic charr whereas the blunt snout, sub-terminal mouth and stocky body of the benthic morphotypes are considered to be pedomorphs (Skúlason et al., 1989). It has been suggested that the retention of embryonic and juvenile characters in the adult phenotype is related to differential timing and/or speed of jaw and skull ossification in late embryonic stages (Eiriksson et al., 1999).

In summary, this study found no differences in *Pax7* coding sequence or spatial expression patterns between morphs, although there appears to be a difference in temporal expression at least between DB and LB morphs. We have identified novel expression of *Pax7* in the lower jaw/mandibular region that may be relevant to the resource polymorphisms exhibited by this species. Further work is required to investigate potential differences in the levels, and spatial and temporal expression patterns, of individual *Pax7* transcripts in the different morphs. Such work may highlight a correlation between delayed rates of development/growth and *Pax7* expression. With regard to the gross morphology of Arctic charr, further studies are also required to investigate what other genes, and/or regulatory regions, are implicated in the rapid evolution of polymorphic Arctic charr.

Acknowledgements

We are grateful to the Natural Environmental Research Council of the UK for funding the project (Grant NER/A/S/2000/00558) and to Susan Gotensparre for the generous donation of degenerate primers, and Sean Earnshaw for help with preparing the illustrations.

References

- Adams, C.E., Huntingford, F.A., 2002a. The functional significance of inherited differences in feeding morphology in a sympatric polymorphic population of Arctic charr. *Evol. Ecol.* 16, 15–25.
- Adams, C.E., Huntingford, F.A., 2002b. Inherited differences in head allometry in polymorphic Arctic charr from Loch Rannoch Scotland. *J. Fish Biol.* 60, 515–520.
- Adams, C.E., Huntingford, F.A., 2004. Incipient speciation driven by phenotypic plasticity? Evidence from sympatric populations of Arctic charr. *Biol. J. Linn. Soc.* 81, 611–618.
- Barr, F.G., Fitzgerald, J.C., Ginsberg, J.P., Vanella, M.L., Davis, R.J., Bennicelli, J.L., 1999. Predominant expression of alternative PAX3 and PAX7 forms in myogenic and neural tumor cell lines. *Cancer Res.* 59, 5443–5448.
- Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G., Desplan, C., 1996. In vivo requirement for the paired domain and homeodomain of the paired segmentation gene product. *Development* 122, 2673–2685.
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., Noll, M., 1986. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of *Drosophila*. *Cell* 47, 1033–1040.

- Chi, N., Epstein, J.A., 2002. Getting your Pax straight: Pax proteins in development and disease. *Trends Genet.* 18, 41–47.
- 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. *Mol. Biol. Evol.* 21, 1146–1151.
- Danzmann, R.G., Ferguson, M.M., Skúlason, S., Snorrason, S.S., Noakes, D.L.G., 1991. Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn Iceland. *J. Fish Biol.* 39, 649–659.
- Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3, 1700–1708.
- Eiriksson, G.M., Skulason, S., Snorrason, S.S., 1999. Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn Iceland. *J. Fish Biol.* 55, 175–185.
- Ennion, S., Wilkes, D., Gauvry, L., Alami-Durante, H., Goldspink, G., 1999. Identification and expression analysis of two developmentally regulated myosin heavy chain gene transcripts in carp (*Cyprinus carpio*). *J. Exp. Biol.* 202, 1081–1090.
- Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., Noll, M., 1986. Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. *Cell* 47, 735–746.
- Gehring, W.J., 1996. The master control gene for morphogenesis and evolution of the eye. *Genes Cells* 1, 11–15.
- Gish, W., 1996–2003. Washington University Basic Local Alignment Search Tool Version 2.0.
- Gotensparre, S.M., Andersson, E., Wargelius, A., Hansen, T., Johnston, I.A., 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.
- Hall, T.E., Smith, P., Johnston, I.A., 2004. Stages of embryonic development in the Atlantic cod *Gadus morhua*. *J. Morphol.* 259, 255–270.
- Higgins, D.G., Bleasby, A.J., Fuchs, R., 1992. Clustal V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* 8, 189–191.
- Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215–219.
- Holland, L.Z., Schubert, M., Kozmik, Z., Holland, N.D., 1999. Amphipax3/7, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. *Evol. Dev.* 1, 153–165.
- Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffé, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J., Roest Crolius, H., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946–957.
- Johnson, L., 1980. The Arctic Charr, *Salvelinus alpinus*. Dr W. Junk Publishers, The Hague.
- Johnston, I.A., Abercromby, M., Vieira, V.L.A., Sigursteindóttir, R.J., Kristjánsson, B.K., Sibthorpe, D., Skúlason, S., 2004. Rapid evolution of muscle fibre number in post-glacial populations of Arctic charr *Salvelinus alpinus*. *J. Exp. Biol.* 207, 4343–4360.
- Jostes, B., Walther, C., Gruss, P., 1990. The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* 33, 27–37.
- Jun, S., Desplan, C., 1996. Cooperative interactions between paired domain and homeodomain. *Development* 122, 2639–2650.
- Jun, S., Wallen, R.V., Goriely, A., Kalionis, B., Desplan, C., 1998. Lune/eye gene, a Pax-like protein, uses a partial paired domain and a homeodomain for DNA recognition. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13720–13725.
- Kay, P.H., Mitchell, C.A., Akkari, A., Papadimitriou, J.M., 1995. Association of an unusual form of a Pax7-like gene with increased efficiency of skeletal muscle regeneration. *Gene* 163, 171–177.
- Kay, P.H., Harmon, D., Fletcher, S., Robertson, T., Ziman, M., Papadimitriou, J.M., 1998. Pax7 includes two polymorphic homeoboxes which contain rearrangements associated with differences in the ability to regenerate damaged skeletal muscle in adult mice. *Int. J. Biochem. Cell Biol.* 30, 261–269.
- Kay, P.H., Ziman, M.R., 1999. Alternate Pax7 paired box transcripts which include a trinucleotide or a hexanucleotide are generated by use of alternate 3' intronic splice sites which are not utilized in the ancestral homologue. *Gene* 230, 55–60.
- Kottelat, M., 1997. European freshwater fishes. *Biologia* 52, 1–271.
- Litchfield, D.W., 2003. Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem. J.* 369, 1–15.
- Mansouri, A., Stoykova, A., Torres, M., Gruss, P., 1996. Dysgenesis of cephalic neural crest derivatives in Pax7^{-/-} mutant mice. *Development* 122, 831–838.
- McKay, S.J., Trautner, J., Smith, M.J., Koop, B.F., Devlin, R.H., 2004. Evolution of duplicated growth hormone genes in autotetraploid salmonid fishes. *Genome* 47, 714–723.
- Moore, C.A., Kittilson, J.D., Ehrman, M.M., Sheridan, M.A., 1999. Rainbow trout (*Oncorhynchus mykiss*) possess two somatostatin mRNAs that are differentially expressed. *Am. J. Physiol.* 277, R1553–R1561.
- Mount, S.M., 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10, 459–472.
- Pinna, L.A., 1990. Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta* 1054, 267–284.
- Relaix, F., Rocancourt, D., Mansouri, A., Buckingham, M., 2004. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes Dev.* 18, 1088–1105.
- Rescan, P.Y., Gauvry, L., 1996. Genome of the rainbow trout (*Oncorhynchus mykiss*) encodes two distinct muscle regulatory factors with homology to myoD. *Comp. Biochem. Physiol.* B 113, 711–715.
- Rescan, P.Y., Jutel, I., Ralliere, C., 2001. Two myostatin genes are differentially expressed in myotomal muscles of the trout (*Oncorhynchus mykiss*). *J. Exp. Biol.* 204, 3523–3529.
- Sandlund, O.T., 1992. The arctic charr *Salvelinus alpinus* in Thingvallavatn. In: Jonasson, P.M. (Ed.), *Ecology of Oligotrophic, Subarctic Thingvallavatn*, Oikos, Odense, pp. 305–351.
- Schafer, B., Czerny, T., Bernasconi, M., Genini, M., Busslinger, M., 1994. Molecular cloning and characterisation of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. *Nucleic Acids Res.* 22, 4574–4582.
- Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., Rudnicki, M.A., 2000. Pax7 is required for the specification of myogenic satellite cells. *Cell* 102, 777–786.
- Senapathy, P., Shapiro, M.B., Harris, N.L., 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* 183, 252–278.
- Seo, H.-C., Saetre, B.O., Havik, B., Ellingsen, S., Fjose, A., 1998. The zebrafish Pax3 and Pax7 homologues are highly conserved, encode multiple isoforms and show dynamic segment-like expression in the developing brain. *Mech. Dev.* 70, 49–63.
- Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in three spine sticklebacks. *Nature* 428, 717–723.
- Skúlason, S., Smith, T.B., 1995. Resource polymorphisms in vertebrates. *Trends Ecol. Evol.* 10, 366–370.
- Skúlason, S., Noakes, D.L.G., Snorrason, S.S., 1989. Ontogeny of trophic morphology in four sympatric morphs of arctic charr *Salvelinus alpinus* in Thingvallavatn Iceland. *Biol. J. Linn. Soc.* 38, 281–301.
- Skúlason, S., Antonsson, T., Gudbergsson, G., Malmquist, H.J., Snorrason, S.S., 1992. Variability in Icelandic Arctic charr. *ICEL. AGR. SCI.* 6, 143–153.
- Skúlason, S., Snorrason, S.S., Ota, D., Noakes, D.L.G., 1993. Genetically based differences in foraging behaviour among sympatric morphs of arctic charr (Pisces: Salmonidae). *Anim. Behav.* 45, 1179–1192.
- Snorrason, S.S., Skúlason, S., Jonsson, B., Malmquist, H.J., Jonasson, P.M., Sandlund, O.T., Lindem, T., 1994. Trophic specialisation in Arctic charr *Salvelinus alpinus* (Pisces; Salmonidae): morphological divergence and ontogenetic niche shifts. *Biol. J. Linn. Soc.* 52, 1–18.

- Terai, Y., Morikawa, N., Kawakami, K., Okada, N., 2003. The complexity of alternative splicing of hageromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12798–12803.
- Tremblay, P., Gruss, P., 1994. Pax: genes for mice and men. *Pharmacol. Ther.* 61, 205–226.
- Underhill, D.A., 2000. Genetic and biochemical diversity in the Pax gene family. *Biochem. Cell Biol.* 78, 629–638.
- Vera, M.I., Kausel, G., Barrera, R., Leal, S., Figueroa, J., Quezada, C., 2000. Seasonal adaptation modulates the expression of the protein kinase CK2 beta subunit gene in the carp. *Biochem. Biophys. Res. Commun.* 271, 735–740.
- Vogan, K.J., Underhill, D.A., Gros, P., 1996. An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol. Cell Biol.* 16, 6677–6686.
- Volpe, J.P., Ferguson, M.M., 1996. Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. *Mol. Ecol.* 5, 763–772.
- West-Eberhard, M.J., 1998. Evolution in the light of developmental and cell biology, and vice versa. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8417–8419.
- West-Eberhard, M.J., 2003. *Developmental Plasticity and Evolution*. Oxford University Press, Oxford.
- Wilkinson, D.G., 1992. Whole mount in situ hybridisation of vertebrate embryos. In: Wilkinson, D.G. (Ed.), *In Situ Hybridisation: A Practical Approach*. IRL Press, Oxford, pp. 939–947.
- Wilson, A.J., Gislason, D., Skúlason, S., Snorrason, S.S., Adams, C.E., Alexander, G., Danzmann, R.G., Ferguson, M.M., 2004. Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Mol. Ecol.* 13, 1129–1142.
- Xu, W., Rould, M.A., Jun, S., Desplan, C., Pabo, C.O., 1995. Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80, 639–650.
- Yasuda, A., Itoh, H., Kawauchi, H., 1986. Primary structure of chum salmon prolactins: occurrence of highly conserved regions. *Arch. Biochem. Biophys.* 244, 528–541.
- Ziman, M.R., Kay, P.H., 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.
- Ziman, M.R., Fletcher, S., Kay, P.H., 1997. Alternate Pax7 transcripts are expressed specifically in skeletal muscle, brain and other organs of adult mice. *Int. J. Biochem. Cell Biol.* 29, 1029–1036.
- Ziman, M.R., Thomas, M., Jacobsen, P., Beazley, L., 2001. A key role for Pax7 transcripts in determination of muscle and nerve cells. *Exp. Cell Res.* 268, 220–229.