



High variability in parvalbumin beta 1 genes offers new molecular options for controlling the mislabeling in commercial Salmonids

Marta Muñoz-Colmenero¹ · Sharmeen Rahman¹ · Jose Luis Martínez² · Eva Garcia-Vazquez¹

Received: 24 September 2018 / Revised: 24 March 2019 / Accepted: 31 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Fish consumption has many health benefits, but may encompass some risks such as allergic reactions. The main responsible of fish allergy is parvalbumins, being one of the most common parvalbumin beta 1 proteins (*Pvalb1*) in many fish species. IgE-mediated cross-reactivity may occur between *Pvalb1* of different species, and there are also cases of species-specific allergic reactions for only a few fish species. The frequent exchanges *Salmo* and *Oncorhynchus* genera found in commercial products demonstrate the great need to continue investigating new tools for the control of mislabeling. In this study, we have amplified by PCR, the parvalbumin beta 1 gene, and analyzed the intron and exon diversity in three genera (*Oncorhynchus*, *Salmo*, and *Salvelinus*), to evaluate the variability of this gene in Salmonids' group. High variability and occurrence of species-specific SNPs found make this gene suitable as powerful species identification tool. In addition, three putative proteins were inferred from translated DNA sequences, which are worthy to investigate, because of their potentially different effect in the fish allergic reaction of salmonid-sensitive consumers.

Keywords Parvalbumin beta 1 gene · Mislabeling · Salmonids · PCR cloning–sequencing · Fish allergy

Introduction

Seafood is an important element of human diet for numerous populations in the world, especially in seaside regions [1]. Fish consumption has been growing worldwide during the last 5 decades [2]. This growth is partly due to health benefits associated with fish consumption. Fish are a relevant source of vitamins, high-assimilation proteins, and polyunsaturated fatty acids such as omega-3, which can reduce the risk of cardiovascular disease [3]. However, there are also drawbacks related to fish consumption. Some of them are the

accumulation of toxic contaminants in fish tissue [4], or the food protein-induced enterocolitis syndrome (FPIES), which are most common for cow's milk and soybean, but may also occur after eating fish [5]. Also the IgE-mediated hypersensitive allergic reactions are a regular problem, which are increasing along with the increase of fish consumption and now affect up to 0.2% of the population in the world [6, 7].

The IgE-mediated allergy reactions to fish are varied: urticarial and gastrointestinal symptoms, asthma, oral anaphylaxis, etc. [8]. These symptoms usually start immediately after the intake of fish, in contrast to FPIES' symptoms. In this last case, the symptoms, as vomiting and diarrhea among others, are presented delayed respect to the food intake [9]. The major allergen responsible to fish allergy in humans is parvalbumin [10]. This calcium-binding protein of about 12 kDa, which has a helix-loop-helix structural domain (EF-hand) where the calcium is bound, is involved in muscle contraction and relaxation and in the rapid cycling of calcium by the sarcoplasmic reticulum [11]. Parvalbumins are abundant in the white muscle of lower vertebrates, although their content is variable. Gazzaz and Rasco [12] published that the content of this protein in Salmonids constitutes 4–50 mg/kg; however, years later, Kuehn et al. [13] published 0.5–2 mg/g for salmon and trout, classifying these

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00217-019-03278-0>) contains supplementary material, which is available to authorized users.

✉ Marta Muñoz-Colmenero
a.martam.colmenero@gmail.com

- ¹ Laboratory of Genetics of Natural Resources, Area of Genetics, Department of Functional Biology, University of Oviedo, C/Julian Claveria s/n, 33006 Oviedo, Asturias, Spain
- ² Sequencing Unit of the University of Oviedo, Severo Ochoa Building, C/Julian Claveria s/n, 33006 Oviedo, Asturias, Spain

fishes within the group of intermediate content of parvalbumin respect to other fish groups. Based on their amino acid sequence, the parvalbumins are divided in two lineages, α and β , with an isoelectric point (pI) of 5.0 and 4.5, respectively [14]. The β -parvalbumins are considered the main allergens [15]. They are exceptionally resistant to heat and are able to trigger allergy even after boiling [16].

There is IgE-mediated allergenic cross-reactivity among parvalbumins from various fish species, which means that a patient allergic to a species may suffer an allergic reaction when eating a different fish [17]. However, this is not generalizable to all the cases. Some fish-hypersensitive patients may be allergic to some species, but tolerate other species without experiencing adverse effects [5, 18, 19]. Cross-reactivity can occur between parvalbumins having similar amino acid sequences [20], but some patients are sensitized to only one type of fish [7]. The occurrence of species-specific allergic reactions makes the exchange of species in commercial seafood a serious health risk for allergic consumers that can tolerate some fish species. This risk has been repeatedly mentioned in studies of species substitution in commercial seafood (e.g., [2, 21]), bringing to light that the species-specific allergic consumer may eat inadvertently the problematic species if they are mislabeled and sold as other species.

Salmonids (salmon, trout, and char) are the first fish group in world production and consumption [22]. Their invaluable contribution to human nutrition may also encompass some risks of allergy amongst some of the millions of consumers. In the last decades, allergic reactions to Salmonids are increasing, especially in countries where consumption is high [23]. According to the known specificity in several fish allergies, some consumers are sensitive only to a few species [24], for example trout or Atlantic salmon, but not to other fish [25]. The regions containing the amino acids 3–10 and 27–38 of the parvalbumin beta 1 have been proposed as potential salmonid-specific binding sites of IgE, an immunoglobulin involved in fish allergy, because they are identical in Salmonids and different in other fish [25]. Perez-Gordo et al. [23] described three possible allergenic epitopes in parvalbumin beta 1 protein. The first epitope (amino acids 1–18) would be responsible for specific allergy of Salmonids, coincident with the 3–10 region identified by Kuehn et al. [25]. The second one, which corresponds to the positions 61–81, would be associated with strong allergic reactions [23]. A third region located between 13 and 39 AA for Saptarshi et al. [6] and 21 and 40 AA for Kuehn et al. [26] was determined as the main reactive epitope against monoclonal antibodies.

On the other hand, commercial mislabeling in Salmonids and substitution between species and even between genera has been described. For example, rainbow trout (*O. mykiss*) is sold as Atlantic salmon (*Salmo salar*) (8.3% mislabeling in commercial products) [27], and vice versa, for Atlantic

salmon being sold as different species of *Oncorhynchus* (4.6% mislabeling reported by Warner et al. [28]; 11.11% by Cline [29]). In all those examples, the Cytochrome oxidase subunit I (COI) gene was used to identify the marketed species, which is one of the most used molecular markers in species identification and mislabeling studies, as part of the Fish Barcode of Life project (FISH-BOLD). However, new molecular markers that can be used with this purpose are always welcome to widen the available tools.

Taking into account the high consumption of salmonid products, the problem of mislabeling found in different studies worldwide, and the potential variability of the Pvalb1 protein, one of the main responsible of the fish allergy in humans, the present study was focused on the diversity of parvalbumin beta 1 gene, pursuing two objectives. In one hand, knowing the variability of Pvalb1 gene in three genera of Salmonids (*Oncorhynchus*, *Salmo*, and *Salvelinus*) as well as their potential power as molecular tool of identification in mislabeling cases. For such, its introns and exons were sequenced and evaluated “in silico” as new molecular tool to discriminate among the Salmonids species. On the other hand, the amino acid sequences were obtained to know the differences in the resultant proteins and their potential risk of allergy derived from exchanges between different genera in this group of fish in mislabeling cases.

Materials and methods

Sample collection

The samples analyzed in this work belong to the collection stored in the Laboratory of Genetics of Natural Resources at the Department of Functional Biology of the University of Oviedo (Spain). Four species of the Salmonidae family were considered: Arctic char, *Salvelinus alpinus* (from Lake Leman; Alps); brown trout, *Salmo trutta*, and Atlantic salmon, *Salmo salar* (wild North Spain populations); rainbow trout, *Oncorhynchus mykiss* (some individuals from a wild population in Vancouver Island, Canada, and others from a farm stock in Spain). Five individuals of each population were analyzed.

Design of primers

To design the primers, complete parvalbumin beta 1 gene sequences (*Pvalb1*) of Atlantic salmon and Coho salmon (*Oncorhynchus kisutch*) (accession numbers FN544080, FN555150.1) were retrieved from GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>). The *Pvalb1* gene in fish contains four exons and three introns [30]. Two pairs of primers were designed manually and then checked with the program PRIMER-3 [31]. One pair amplifies a DNA

fragment that includes exon 1 to exon 3 and the two introns in the middle (Online Resource S1). The other pair amplifies the DNA fragment comprised between the 3' end of the third exon and the end of the fourth exon, including the third intron (Online Resource S1).

DNA extraction, PCR amplification, and sequencing of partial sequence of *Pvalb1*

DNA was extracted from muscle using the protocol developed by Estoup et al. [32]. The extractions were kept at 4 °C until analysis and aliquots were stored at – 20 °C for long time preservation.

The polymerase chain reaction (PCR) was performed in a total volume of 40 µl. We added 4 µl from the extractions of DNA, 2.5 mM of MgCl₂, 4 µl of dNTPs 2.5 mM, 2 µl of DNA Taq polymerase (Promega, Madison WI), buffer 1 × of the polymerase, and 20 pmol of each primer. The PCR conditions for the first primer pair were: an initial denaturation at 95 °C for 5 min, then 35 cycles with denaturation at 95 °C for 20 s, annealing at 61 °C for 20 s, and extension at 72 °C for 30 s, followed by a final extension at 72 °C for 10 min. For the second primer pair, the annealing temperature was 60 °C.

PCR products were visualized and purified from 2% agarose gel with the kit Wizard SV Gel and PCR Clean-Up System (Promega). Sequencing was performed at the sequencing facilities of the University of Oviedo employing BigDye terminator chemistry and an ABI Prism 3130xl genetic analyzer. When we found evidence of haplotype mixture (i.e., heterozygotes that yield overlapped peaks in the sequence chromatogram), cloning was carried out after sequencing in the two directions to determine reliably the heterozygous positions.

Cloning and PCR colony

To separate the different *Pvalb1* haplotypes in heterozygotes, cloning was carried out using the Novagen AcceptorTM Vector Kit with the pSTBlueTM-1 (Merck Millipore) for

the ligation. The DNA quantity was determined using the formula specified by the manufacturer. The transformation process was done by thermal shock with the NovaBlue SinglesTM Competent Cells of *Escherichia coli* following the instructions specified in the kit (Merck Millipore). The transformed bacteria were grown in liquid SOC medium during 1 h and spread onto solid LB medium with ampicillin. The next day, the white colonies were picked in 50 µl bidistilled water and heated at 99 °C for 5 min to break and denature the DNA for the colony PCR. PCR of the colonies was performed in a total volume of 20 µl including 3 µl of DNA, 0.6 µl MgCl₂ (50 mM), 2 µl dNTPs (2.5 mM), 0.3 µl of DNA Taq polymerase (Biotools, reference 10.047), buffer 1 × of the Taq polymerase, and 2 µl of each T7 and SP6 primers (10 µM). These primers are designed to amplify the insert and part of the vector. PCR conditions were: one cycle at 95 °C for 5 min; 35 cycles at 95 °C for 30 s, 55 °C 30 s, and 72 °C for 30 s, and a final cycle at 72 °C for 10 min. DNA purification was carried out with Illustra Exostar 1-Step (GE Healthcare Life Sciences) and sequencing was performed as explained above. We sequenced a number of clones proportional to the variation (number of haplotypes) found per individual (Table 1). The number of clones sequenced from each individual was determined based on having at least two sequences of each haplotype. If one haplotype had no replicate, new clones were sequenced until finding at least one more sequence of the same haplotype. Additional clones were sequenced from one individual of each species to check if the number of clones was sufficient to detect all the haplotypes.

Sequencing of complete parvalbumin beta 1 gene in *Salmo salar*

To sequence the complete gene of the *Pvalb1* (primers 1–4; Online resource S1), we worked on one *Salmo salar* individual (SS46). Quality DNA extraction was performed with QIAamp DNA Mini Kit (Quiagen). DNA was eluted in the last step with double quantity of elution buffer than indicated by the manufacturer. The PCR mix was 4 µl of DNA, 4 µl of

Table 1 Haplotypes and intrapopulation variation found for *Pvalb1* gene in this study

Population	S		Hap		PS		Hd
	Total	Mean (SD)	Total	Mean (SD)	Total	Mean (SD)	
Farm <i>Oncorhynchus mykiss</i>	24	4.8 (1.5)	3	2.0 (0.0)	4	1.4 (0.9)	0.5978
Wild <i>Oncorhynchus mykiss</i>	86	17.2 (6.9)	10	4.6 (0.9)	6	5.0 (0.7)	0.7715
Wild <i>Salmo salar</i>	66	13.2 (4.7)	8	4.0 (1.4)	12	12.0 (0.0)	0.7445
Wild <i>Salmo trutta</i>	85	17.0 (6.0)	17	6.0 (0.7)	9	7.6 (1.5)	0.8207
Wild <i>Salvelinus alpinus</i>	103	20.6 (6.7)	19	7.2 (2.8)	8	7.8 (0.4)	0.7921

Data are presented as total per population ($N=5$), mean per individual, and SD in parenthesis for the number of clones sequenced (S), haplotypes (Hap), and polymorphic sites (PS). Mean heterozygosity (Hd) per population is also given

dNTPs (2.5 mM), 2 µl of the Buffer of the Taq polymerase with MgCl₂, 0.75 µl of the Taq polymerase Expand long template PCR system (ROCHE, reference 11681834001), 1 µl of each primer (20 µM), and bidistilled water to complete 20 µl of total volume. The PCR conditions were: an initial denaturation at 94 °C for 2 min, then ten cycles with denaturation at 94 °C for 10 s, annealing at 55 °C for 30 s, and extension at 68 °C for 90 s; then other 25 cycles with denaturation at 94 °C 15 s, annealing 55 °C 30 s, and extension at 68 °C 90 s; finally, the last extension at 68 °C for 7 min. The process of cloning, PCR colony, and sequencing was the same that the previous section. The clones with inserts were sequenced twice in the two directions.

Sequence analysis and phylogenetic trees

The different haplotypes in each individual were established analyzing the sequences obtained after the PCR colony. Consensus haplotype sequences were obtained for each individual by alignment of forward and reverse sequences. From these haplotype sequences, the heterozygote positions were identified in comparison with the consensus complete parvalbumin beta 1 gene. To align sequences, we used BioEdit Sequence Alignment Editor Software 7.0.5.3 [33] and SeqScape software v2.6 (Applied Biosystems, 2004).

The sequences were checked first to confirm that they were parvalbumin beta 1 sequences, using the BLAST tool of GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>). The minimum identity found was 98%.

The best model of nucleotide substitution was obtained with the program jModelTest 0.1.1 [34] with the Akaike information criterion (AIC) [35]. For this gene and group of species, it was Tamura-Nei model [36]. The phylogenetic analyses were performed with MEGA 6 [37]. Phylogenetic trees were reconstructed from the sequences to inferring their evolutionary relationships, using Neighbor-Joining and Maximum-Likelihood methods with 5000 bootstrap replicates. The positions with less than 95% site coverage were removed. A *Pvalb1* sequence of *Gadus morhua* from GenBank ([24]; accession number FN544076.1) was used as outgroup for reconstructing the phylogenetic trees.

Inference of putative proteins

To translate the haplotypes to amino acid sequences, introns were removed, taking as a reference the sequence of parvalbumin beta 1 of *Salmo salar* from GenBank of accession number FN544080. The resulting amino acid sequences correspond to the three first exons, missing the fourth (26 base pairs, bp). The DNA sequences were translated to amino acids with the program MEGA 6. Calcium-binding sites were identified within the amino acid sequences following Swoboda et al. [24].

Tridimensional models of the putative proteins were generated from the different haplotypes using the MODEL-ER program [38], through the ModWeb [39]. The resulting models were visualized with the graphics program Chimera 1.6.1 [40].

Results

Amplification of *Pvalb1* amplicons

To amplify the 5' part of the parvalbumin beta 1 gene, the pair of primers designed was: Forward, Psa1 (5'-CTGTGCCATCTGTGCAAGG- 3') and Reverse, Psa2 (5'-CCAATCATGCCATCACCATCG-3'). These primers amplify the three first exons and two introns (Online Resource S1), from sites 6–654. The amplicon size was 650 bp long for *O. mykiss* and 649 bp for the other species, except for some *Salmo trutta* individuals that exhibited haplotypes with a single-nucleotide deletion in the second intron at position 382.

For amplifying the third intron and the fourth exon, the primer pair was: Forward Psa3 (5'-TACCGATGCAGAGACAAAGG-3') and Reverse Psa4 (5'-GTCTTGGGCAATATTGTTCC-3'). These primers amplify from position 592 to ~ 20 nucleotides past the 3' end of the parvalbumin beta 1 gene. The amplicons were around 931 bp long and vary depending on the length of the third intron.

The complete *Pvalb1* gene obtained in this work was of 1495 bp. It is available in GenBank database, <http://www.ncbi.nlm.nih.gov/genbank/> with the accession number KT756204. The haplotypes obtained from the five individuals of each population analyzed are available with the following accession numbers: JX123033–JX123035 for farmed *Oncorhynchus mykiss*; KT756171–KT756180 for wild *Oncorhynchus mykiss*; JX123036–JX123040 and KT756181–KT756183 for wild *Salmo salar*; JX123043, JX123045–JX123051 and KT756184–KT756192 for wild *Salmo trutta*; JX123052–JX123059 and KT756193–KT756203 for wild *Salvelinus alpinus*.

The results of the BLAST tool confirmed that our clone sequences were parvalbumin beta 1, matching with sequences as FN544260.1, FN544080.1 or FM994927.1 of the GenBank database with 99% of identity.

Variability of the partial parvalbumin beta 1 gene sequence

The genetic polymorphisms for this gene were analyzed for the DNA region including exons 1–3. This region includes the three first exons and the complete second intron, which is the most variable part of this gene [41].

All the haplotypes obtained in this study are disclosed in Online Resource S3 in detail. The number of haplotypes found was different for the different populations and species (Table 1). The species with more polymorphic positions was *S. salar* and the species with more haplotypes was *S. alpinus*. In contrast, higher haplotype diversity was found for *S. trutta* (Table 1). *Salmo trutta* shared one polymorphic site with *S. salar* and another with *Salvelinus alpinus* (Table 2). On the other hand, some polymorphic sites in wild *O. mykiss* were not variable in the individuals analyzed of the farmed stock. The number of haplotypes per individual was > 2 in most cases. The highest number of haplotypes corresponded to one *S. alpinus* (individual SA3) with 12 haplotypes, which means that at least six copies of the gene occur in *S. alpinus* genome.

A total of 25 nucleotide positions were different between species but identical for all the haplotypes of the same species, representing species-specific single-nucleotide polymorphisms (SNP). Four of them were located within exons and only one of those was synonymous (Table 3), and thus, three non-synonymous changes were found. For coding sequence, Codon-based test of neutrality for analysis averaging over all sequence pairs was carried out using MEGA v6 program, resulting in significant non-neutrality (-0.789 ; $p=0.431$), showing that this gene is under selection, as it is usual in important functional genes.

Phylogenetic analysis

The Neighbor-Joining phylogenetic tree reconstructed from the haplotypes obtained in this study exhibited the same topology (Fig. 1), and was perfectly concordant with the taxonomic status of the individuals analyzed, clustering by species and then by genera. *Salmo trutta* and *Salmo salar* clustered in the same branch. In another branch, close to them were located all the *O. mykiss* haplotypes (wild and farmed haplotypes mixed). The more distant

cluster contained the haplotypes of *Salvelinus alpinus* (Fig. 1).

Inferred variants of the parvalbumin beta 1 protein

After translating the haplotypes from nucleotides to amino acids, three different putative proteins of 96 amino acids were obtained. The 2 first amino acids methionine and alanine coded by the 5' six first nucleotides of exon 1, and the 11 last amino acids coded in exon 4 are missing in our partial sequences. The inferred partial protein was the same for all the haplotypes within species (Table 2), as expected for functionally important proteins. It was also the same for *O. mykiss* and *S. alpinus*. Differences occurred between the genus *Salmo* and the other two (Fig. 2). *Salmo trutta* and *S. salar* exhibited glutamic acid (E) instead the valine (V) of *O. mykiss/S. alpinus* in position 77. E and V are not in the same group of amino acids, E being acid and V a-polar. In addition, *S. salar* had threonine (T) instead serine (S) in position 24 (Fig. 2), which is a change unique for this species and shows, therefore, variation between species within the genus *Salmo*. T and S are both polar amino acids with no electric charge.

A further comparison with protein information from public databases was made for the final part of the protein inferred in this study (amino acids 99–109). We found two changes compared with other sequences of Salmonids available in public databases in December 2015 (Online Resource S2), for the amino acids present in positions 101 and 106.

Calcium-binding sites were identified using the information in Swoboda et al. [24]. In the putative proteins of this study, they are located in the amino acid positions 52–63 and 91–102, and are both 12 amino acid long (Fig. 2). Non-synonymous changes were not found in these positions in our study. The tridimensional models inferred from the protein sequences were similar to each other (Fig. 2) despite different types of amino acid substitutions, as expected for this type of functionally important proteins.

Table 2 Summary of polymorphic sites found in the parvalbumin beta 1 sequence within each species

	Polymorphic sites											
Farm <i>O. mykiss</i>	91 R	133 M	218 M	389 A	445 C	488 R						
Wild <i>O. mykiss</i>	91 R	133 M	218 M	389 R	445 M	488 R						
Wild <i>S. salar</i>	60 K	67 K	152 M	397 Y	410 K	429 W	454 S	477 R	502 R	505 Y	530 K	587 Y
Wild <i>S. trutta</i>	40 W	131 K	134 R	261 Y	382 Gap/R	465 R	477 R	486 Y	545 M			
Wild <i>S. alpinus</i>	95 M	209 Y	221 M	431 Y	462 K	465 R	492 Y	524 K				

Red: sites monomorphic in farmed rainbow trout (*O. mykiss*); Green: sites shared between *Salmo salar* and *Salmo trutta*; Orange: sites shared between *Salmo trutta* and *Salvelinus alpinus*. The number corresponds to the nucleotide position in the sequence and the letter is the nucleotide occurring in that position: K (G or T), M (A or C), R (A or G), S (G or C), W (A or T), and Y (C or T)

Table 3 Difference of nucleotide positions between species and their location in the gene (intron or exon), OM, *O. mykiss*; SA, *S. alpinus*; ST, *S. trutta*; SS, *S. salar*

Position	Nucleotides	Species	Location	Position	Nucleotides	Species	Location
70	C	OM	Intron 1	443	T	Rest	Intron 2
	T	Rest			C	SA	
99	C	OM, SA	Intron 1	444	C	SA, OM	Intron 2
	A	SS, ST			T	SS, ST	
107	A	OM	Intron 1	469	G	OM	Intron 2
	C	Rest			C	Rest	
111	G	OM	Intron 1	476	T	Rest	Intron 2
	T	Rest			G	SA	
167	G	Rest	Intron 1	480	T	Rest	Intron 2
	A	SA			G	OM	
188	T	OM	Intron 1	482	T	Rest	Intron 2
	gap	Rest			C	SA	
204	A	OM	Intron 1	483	G	Rest	Intron 2
	T	SA			A	OM	
	G	SS, ST		498	T	SA, OM	Intron 2
208	A	SS, ST	Intron 1		C	SS, ST	
	G	OM, SA		515	T	OM	Intron 2
244	T	SA	Exon 2, non-synonymous		G	Rest	
	A	Rest		535	G	Rest	Intron 2
245	C	SA, SS	Exon 2, non-synonymous		T	SA	
	G	ST, OM		546	T	SA, OM	Intron 2
271	C	SS, ST	Exon 2, synonymous		C	SS, ST	
	T	SA, OM		584	T	SA, OM	Exon 3, non-synonymous
414	T	SA, OM	Intron 2		A	SS, ST	
	G	SS, ST					
418	C	OM	Intron 2				
	T	Rest					

Risk of allergy derived from salmonid mislabeling

For this assessment, we have considered published mislabeling cases and use the information obtained in this study for differences between species at parvalbumin beta 1 gene. One of the non-synonymous changes detected in this study affects the amino acid 77 (Fig. 2), within the 61–81 epitope identified by Perez-Gordo et al. [23] as responsible of strong allergic reactions. *Salmo* species exhibit an *E* (acid), while *Oncorhynchus* and *Salvelinus* have a *V* (a-polar) in that position. The other non-synonymous change (*T* for *S* at position 24 in *Salmo salar*) also affects one region involved in allergy [6, 26]. Therefore, changes involving substitutions of *Salmo* for *Oncorhynchus* species or vice versa would imply consumers being exposed to different variants of the parvalbumin antigen.

From mislabeling reported in three studies [27–29], we found that 8.3, 4.7, and 11.1% commercial samples sold in South Africa and in different regions of USA were replaced by another species with a change in amino acid 77 within an

epitope involved in strong allergy reactions (Table 4). This could imply potential modifications in the type of allergy reaction suffered for a consumer with specific or stronger allergy to species of *Salmo* genus when inadvertently eating *Oncorhynchus* species and vice versa. However, the 3D structure obtained was very similar among all proteins, so the reaction against them is unknown yet, until more analyses over the different protein variants are conducted.

Discussion

The two pairs of primers Psa11–Psa12 and Psa13–Psa14 designed in this work amplify a partial sequence of parvalbumin beta 1 gene in different species of Salmonids (subfamily Salmonidae), with high identity to reference sequences of the coding gene available in databases. Stop codons and/or insertion-deletions were not found in the exons for any haplotype. The combined amplicon codes for 109 amino acids, typical of beta-parvalbumins, and has the

Fig. 1 Neighbor-joining tree reconstructed from for the 59 haplotypes of parvalbumin beta 1 nucleotide sequences found in this study and the reference sequence of *S. salar* (GenBank accession number FN544080). The evolution model used was Tamura-Nei with gamma parameter of 0.8656. A sequence of *Gadus morhua* (GenBank accession number FN544076) was included as outgroup

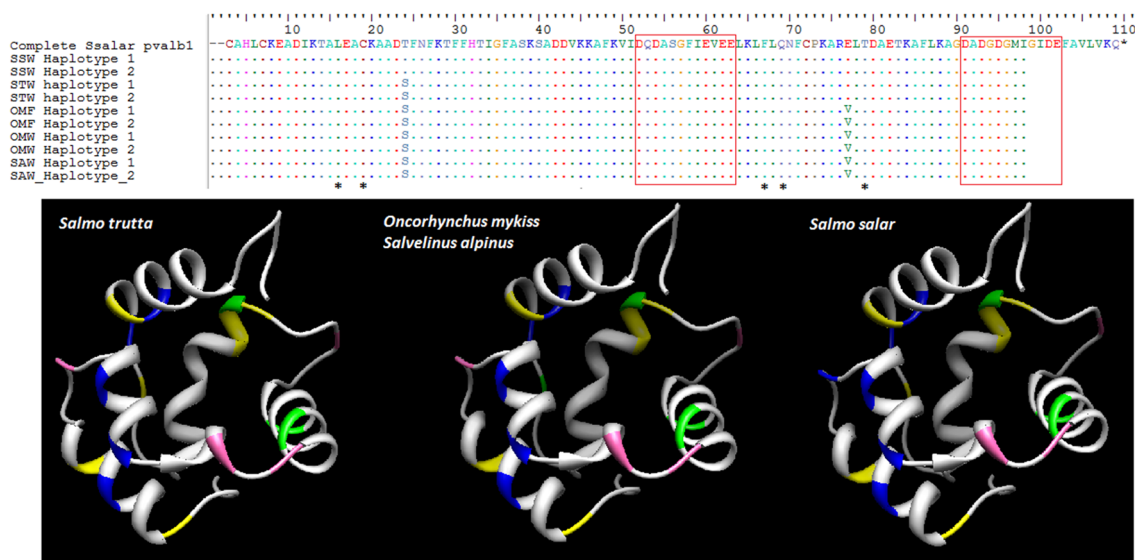
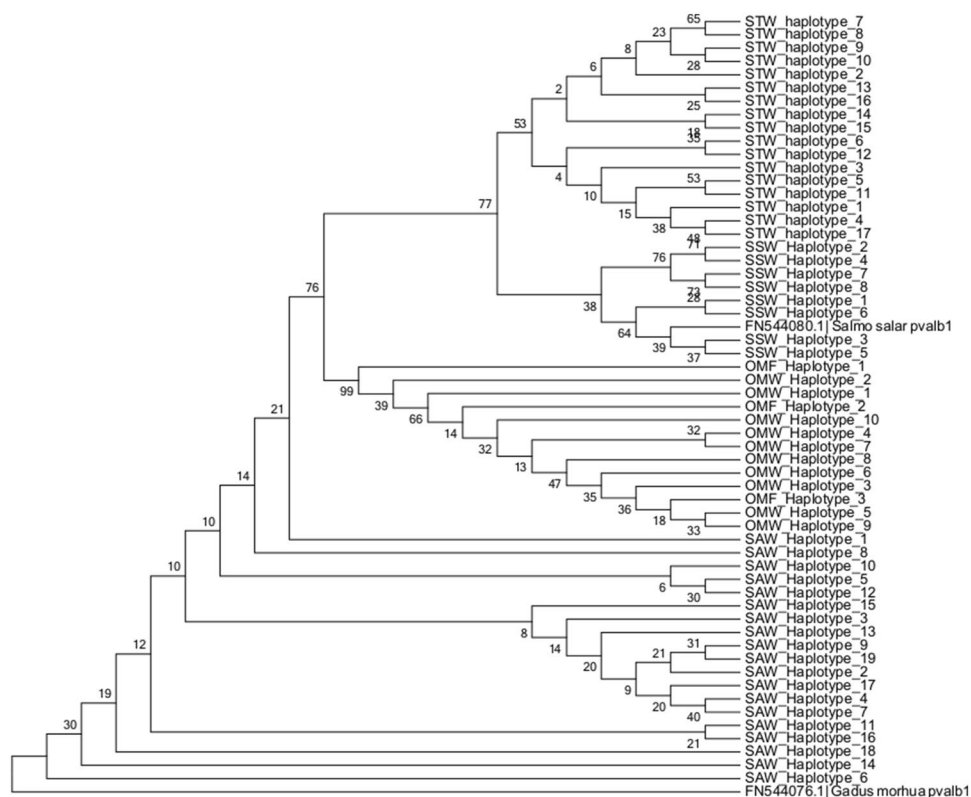


Fig. 2 Alignment of the putative protein sequences. Translations of two haplotypes per species were used. They correspond to farmed *O. mykiss* (OMF), wild *O. mykiss* (OMW), wild *Salvelinus alpinus* (SAW), wild *Salmo salar* (SSW), and wild *Salmo trutta* (STW). Cal-

cium-binding sites are marked with red boxes and β -specific amino acid residues are indicated with asterisks. In the tridimensional models reconstructed with MODELLER [38], serine, threonine, valine, and glutamic acid are in pink, blue, green, and yellow, respectively

characteristic beta isoform residues (Fig. 2). All together confirms that we have sequenced a functional gene from the new primer pairs.

The study of both introns and exons of this gene directly from genomic DNA is a considerable novelty, since most studies of parvalbumin beta 1 have been focused on cDNA

Table 4 Expected change in allergy epitopes for mislabeling reported in Salmonids, based on non-synonymous changes in epitopes involved in allergy

Study (country)	Label species	Substitute species	% cases	Change in epitope
[27] (South Africa)	<i>Salmo salar</i>	<i>Oncorhynchus mykiss</i>	8.3	Yes
[28] (USA)	<i>Oncorhynchus</i> spp.	<i>Salmo salar</i>	4.7	Yes
	<i>Oncorhynchus</i> spp.	Other <i>Oncorhynchus</i> spp.	4.9	?
[29] (USA)	<i>Oncorhynchus</i> spp.	<i>Salmo salar</i>	11.1	Yes
	<i>Oncorhynchus</i> spp.	Other <i>Oncorhynchus</i> spp.	9.1	?

[42] or directly on the protein, but similar studies based on DNA tools have been performed in other fish species [43]. This new approach revealed high intra-individual variability in the DNA sequence of this gene, which allows the possible application of this study about the parvalbumin beta-1 gene in the field of seafood traceability [41, 44, 45]. Many SNPs found in the short amplicon of this study seem to be species-specific (Tables 2, 3) and could be employed to authenticate marketed Salmonids, which is a clear need for this products [27, 28]. The SNPs tool is a rapid and efficient assay for genotyping and traceability studies [46], so the information collected here can be a good basis for the development of SNPs discovered here and their application in the identification of salmonid species. SNPs' tool, compared with other molecular tools like DNA barcoding, is useful not only for species identification if not for population studies. According to that, in this work, some position has been fixed in the case of rainbow trout from aquaculture, resulting in an interesting topic to explore in future studies with the objective to discriminate trout/salmon from aquaculture respect to wild trout/salmon, which is other common problem in Salmonids traceability [47]. Other traceability tool for Salmonids, using parvalbumin beta gene, which comes from our results, could be performed through the determination of gene copy number. This strategy using the Pvalb1 gene would allow simultaneously the authentication of commercial products derived from Salmonids and the assessment of their allergenic potential after clinical assays.

Multiple copies in the same individual have been detected in the cases that more than two haplotypes per individual have been found. This is consistent with the previous studies in fish parvalbumins; for example, [48] found nine putative parvalbumin genes in zebrafish. In our study, the exception to multiple copies was farmed *O. mykiss* where we found only two haplotypes per individual, but this could be explained by diversity losses that are well known in farmed fish (e.g., [49]). The relationship between the number of copies of parvalbumin genes and the allergenic potential of a fish is an emergent question [44]. The amount of parvalbumin beta 1 protein is related with allergenicity [50], although it is possible that the different copies

are tissue-specific and the number of copies is not related with the level of gene expression, so this subject should be explored in future studies. The evolution by duplication is known in many genes and different organisms [51], and in the case of Salmonids, a large part of the genome has evolved by deletions and duplications from the existing genes (e.g., [52]). Several copies of mRNA of parvalbumin beta gene have been reported in raw fillets of rainbow trout, which has been interpreted as this gene is being transcribed continuously in white muscle [53]. It is possible that a high number of gene copies of the gene enable high transcription levels in fish muscle. From this study, we cannot know if all Salmonids have the same number of copies, but, in view of the differences found between species (Tables 1, 3; Fig. 2) and the high degree of mislabeling occurring in this fish group, it would be interesting to find out whether the number of copies varies between species and the possible effect of such variation on species' allergenic potential. Given the high number of potential parvalbumin genes in fish [48], more studies would be necessary, especially in the important field of allergies.

Finally, from the point of view of fish allergies, this study raises interesting perspectives, although, with only in silico results, so more assays in vitro will be needed to verify our hypothesis. The putative amino acid changes in positions 77 and 24 found between *Salmo salar* and *Oncorhynchus/Salvelinus* genera fall within regions involved in allergic reactions [6, 23, 26]. The exchanges between *Salmo salar* and *Oncorhynchus* species are frequent in salmonid commercial products [27, 29], bestowing to these non-synonymous changes special importance. Although the final protein cannot be directly inferred from the genomic sequence, because the gene transcripts could suffer post-transcriptional and even post-translational editions, our results are worthy to be further explored. They may suggest causes of differences in allergenic capacity of salmonid species, or potential changes in the strength of the reaction suffered for a sensible patient, especially important in the species most marketed and, therefore, more prone to be mislabeled.

Conclusions

The parvalbumin beta-1 gene analyzed directly from genomic DNA exhibited high variability and multiple gene copies in some of the most consumed salmonid species. Many species-specific SNPs in introns and exons were found, being some of those found in exons non-synonymous and located in areas of the protein identified as important epitopes binding IgE antibodies in allergic reactions. The high variability of the *Pvalb1* gene discovered here offers a valuable catalog of SNPs and copy-number variation that might be the basis for the development of authentication tools for commercial salmonid seafood. These tools could be used to control mislabeling between different species or genera as well as between wild and farmed species due to the variability found in the nuclear gene *Pvalb1* seems to be potentially capable of discriminating both ways. On the other hand, the differences discovered in the amino acid sequence among species could encompass different degree of allergic reactions for the sensible consumers, which is interesting to check with in vitro assays.

Acknowledgements The authors are grateful to the assistance of Daniel Serna Fuente (Sequencing Unit of the University of Oviedo).

Funding This study was supported by the Regional Government of Asturias (Grant GRUPIN-2014-093). M. Muñoz-Colmenero holds a national Spanish Grant (reference AP-2010-5211).

Compliance with ethical standards

Conflict of interest The authors have no competing interests regarding this study.

Compliance with ethics requirements The samples analyzed in this work belong to the tissue collection stored in the Laboratory of Genetics of Natural Resources at the Department of Functional Biology of the University of Oviedo (Spain). Therefore this article does not contain any studies with human or animal subjects.

References

- Perez-Gordo M, Cuesta-Herranz J, Maroto AS, Cases B, Ibáñez MD, Vivanco F et al (2011) Identification of sole parvalbumin as a major allergen: study of cross-reactivity between parvalbumins in a Spanish fish-allergic population. *Clin Exp Allergy* 41:750–758
- Costa Leal M, Pimentel T, Ricardo F, Rosa R, Calado R (2015) Seafood traceability: current needs, available tools, and biotechnological challenges for origin certification. *Trends Biotechnol* 33:331–336
- Wang Y, Crawford MA, Chen J, Li J, Ghebremeskel K, Campbell TC et al (2003) Fish consumption, blood docosahexaenoic acid and chronic diseases in Chinese rural populations. *Comp Biochem Physiol A: Mol Integr Physiol* 136:127–140
- Cawthorn DM, Steinman HA, Witthuhn RC (2011) Evaluating the availability of fish species on the South African market and the factors undermining sustainability and consumer choice. *Food Control* 22:1748–1759
- Zapatero Remón L, Alonso Lebrero A, Martín Fernández E, Martínez Molero MI (2005) Food protein-induced enterocolitis syndrome caused by fish. *Allergol Immunopathol* 33:312–316
- Saptarshi S, Sharp MF, Kamath SD, Lopata AL (2014) Antibody reactivity to the major fish allergen parvalbumin is determined by isoforms and impact of thermal processing. *Food Chem* 148:321–328
- Sharp MF, Lopata AL (2014) Fish allergy: in review. *Clinic Rev Allerg Immunol* 46:258–271
- Romano A, Di Fonso M, Giuffreda F, Quarantino D, Papa G, Palmieri V et al (1995) Diagnostic work-up for food-dependent, exercise-induced anaphylaxis. *Allergy* 50:817–824
- Leonard SA, Pecora V, Giovanni Fiocchi A, Nowak-Węgrzyn A (2018) Food protein-induced enterocolitis syndrome: a review of the new guidelines. *World Allergy Organ J* 11:4
- Aas K, Jebsen JW (1967) Studies of hypersensitivity to fish: partial purification and crystallization of a major allergenic component of cod. *Int Arch Allergy Immunol* 32:1–20
- Heffron JK, Moerland TS (2008) Parvalbumin characterization from the euryhaline stingray *Dasyatis Sabina*. *Comp Biochem Physiol A Mol Integr Physiol* 150:339–346
- Gazzaz SS, Rasco BA (1993) Parvalbumins in fish and their role as food allergens: a review. *Rev Fish Sci* 1:1–26
- Kuehn A, Scheuermann T, Hilger C, Hentges F (2010) Important variations in parvalbumin content in common fish species: a factor possibly contributing to variable allergenicity. *Int Arch Allergy Immunol* 153:359–366
- Cai QF, Liu GM, Li T, Hara K, Wang XC, Su WS et al (2010) Purification and characterization of parvalbumins, the major allergens in red stingray (*Dasyatis akajei*). *J Agr Food Chem* 58:12964–12969
- Radauer C, Biblin M, Wagner S, Mari A, Breiteneder H (2008) Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol* 121:847–852
- Arif SH, Hasnain A (2010) A major cross-reactive fish allergen with exceptional stability: parvalbumin. *Afr J Food Sci* 4:109–114
- Van Do T, Elsayed S, Florvaag E, Hordvik I, Endresen C (2005) Allergy to fish parvalbumins: studies on the cross-reactivity of allergens from 9 commonly consumed fish. *J Allergy Clin Immunol* 116:1314–1320
- Sicherer SH (2001) Clinical implications of cross-reactive food allergens. *J Allergy Clin Immunol* 108:881–890
- Raith M, Klug C, Sesztak-Greinecker G, Balic N, Focke M, Linhart B, Hemmer W, Swoboda I (2014) Unusual sensitization to parvalbumins from certain fish species. *Ann Allergy Asthma Immunol* 113:571–578
- Helbling A, Haydel R Jr, McCants ML, Musmand JJ, El-Dahr J, Lehrer SB (1999) Fish allergy: is cross-reactivity among fish species relevant? Double-blind placebo-controlled food challenge studies of fish allergic adults. *Ann Allergy Asthma Immunol* 83:517–523
- Muñoz-Colmenero M, Klett-Mingo M, Diaz E, Blanco O, Martínez JL, García-Vázquez E (2015) Evolution of hake mislabeling niches in commercial markets. *Food Control* 54:267–274
- Naylor RL, Goldberg RJ, Primavera JH, Kautsky N, Beveridge MCM, Clay J et al (2000) Effect of aquaculture on world fish supplies. *Nature* 405:1017–1024
- Perez-Gordo M, Lin J, Bardina L, Pastor-Vargas C, Cases B, Vivanco F et al (2012) Epitope mapping of Atlantic salmon major allergen by peptide microarray immunoassay. *Int Arch Allergy Immunol* 157:31–40

24. Swoboda I, Bugajska-Schretter A, Verdino P, Keller W, Sperr WR, Valent P et al (2002) Recombinant Carp parvalbumin, the major cross-reactive fish allergen: a tool for diagnosis and therapy of fish allergy. *J Immunol* 168:4576–4584
25. Kuehn A, Hutt-Kempf E, Hilger C, Hentges F (2011) Clinical monosensitivity to salmonid fish linked to specific Ig-E epitopes on salmon and trout beta-parvalbumins. *Allergy* 66:298–305
26. Kuehn A, Swoboda I, Arumugan K, Hilger C, Hentges F (2014) Fish allergens at a glance: variable allergenicity of parvalbumins, the major fish allergens. *Front Immunol* 5:1–8
27. Cawthorn DM, Steinman HA, Witthuhn RC (2012) DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Res Int* 46:30–40
28. Warner K, Timme W, Lowell B, Hirshfield M (2013) Oceana study reveals seafood fraud nationwide. OCEANA Publications. <http://oceana.org/reports/oceana-study-reveals-seafood-fraud-nationwide>. Accessed Feb 2016
29. Cline E (2012) Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. *Food Res Int* 45:388–393
30. Xu YX, Zhu ZY, Lo LC, Wang CM, Lin G, Feng F et al (2006) Characterization of two parvalbumin genes and their association with growth traits in Asian seabass (*Lates calcarifer*). *Anim Genet* 37:266–268
31. Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users for biologist programmers. In: Karwetz S, Misener S (eds) *Bioinformatics methods and protocols: methods in molecular biology*. Human Press, Totowa, pp 365–386
32. Estoup A, Largiader CR, Perrot E, Chourrout D (1996) Rapid one tub DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Mol Mar Biol Biotechnol* 5:295–298
33. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
34. Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253–1256
35. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–722
36. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
38. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
39. ModWeb: a server for protein structure modeling. <https://modbase.compbio.ucsf.edu/modweb/>. Accessed 12 Mar 2015
40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and Analysis. *J Comput Chem* 25:1605–1612
41. Hanák P, Laknerová I, Svátora M (2012) Second intron in the protein-coding region of the fish parvalbumin gene—a promising platform for polymerase chain reaction-based discrimination of fish meat of various species. *J Food Nutr Res* 51:81–88
42. Lindstrom C, Hordvik I, Endresen C, Elsayed S (1996) Cloning of two cDNAs encoding parvalbumin, the major allergen of Atlantic salmon (*Salmo salar*). *Scand J Immunol* 44:335–344
43. Rencova E, Kostelnikova D, Tremlova B (2013) Detection of allergenic parvalbumin of Atlantic and Pacific herrings in fish productions by PCR. *Food Addit Contam Part A* 30:1679–1683
44. Sun M, Liang C, Gao H, Lin CDM (2009) Detection of parvalbumin, a common fish allergen gene in food, by real-time polymerase chain reaction. *J AOAC Int* 92:234–240
45. Rehbein H (2013) Differentiation of fish species by PCR-based DNA analysis of nuclear genes. *Eur Food Res Technol* 236:979–990
46. Machado-Schiaffino G, Martinez JL, Garcia-Vazquez E (2008) Detection of mislabeling in hake seafood employing mtSNPs-based methodology with identification of eleven hake species of the genus *Merluccius*. *J Agric Food Chem* 56:5091–5095
47. Muñoz-Colmenero M, Juanes F, Dopico E, Martínez JL, Garcia-Vazquez E (2017) Economy matters: a study of mislabeling in salmon products from two regions, Alaska and Canada (Northwest of America) and Asturias (Northwest of Spain). *Fish Res* 195:180–185
48. Friedberg F (2005) Parvalbumin isoforms in zebrafish. *Mol Biol Rep* 32:167–175
49. Wang S, Hard JJ, Utter F (2001) Salmonid inbreeding: a review. *Rev Fish Biol Fisher* 11:301–319
50. Griesmeier U, Vázquez-Cortés S, Bublin M, Radauer C, Ma Y, Briza P et al (2010) Expression levels of parvalbumins determine allergenicity of fish species. *Allergy* 65:191–198
51. Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292–298
52. Koop BF, Davidson WS (2008) Genomics and the genome duplication in Salmonids. In: *Fisheries for global welfare and environment*, 5th World Fisheries Congress, pp 77–86
53. Rehbein H (2007) Parvalbumin mRNA in raw and processed fillet of rainbow trout *Oncorhynchus mykiss*: determination of copy number and stability against heat and acid. *Fish Sci* 73:160–1165

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

FN544080.1: 1..1.5K (1.5Kbp) ▾

100 200 300 400 500 600 700 800 900 1 K 1,100 1,200 1,300 1,479

Genes



exon 2

exon 3

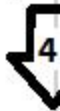
exon 4

Intron Features

number=1

/number=2

/number=3



Online Resource S2: Putative aminoacids translated from partial parvalbumin beta 1 gene. The sequences were obtained from the GenBank database and one from our study (marked with an asterisk). Polymorphic amino acid sites (101, 106) are in bold.

	Aminoacids from 99 to 109
<i>S. salar</i> KT756204*	G I DEF A VLVKQ
<i>S. salar</i> FN544080	G I NEFAVLVKQ
<i>O. kisutch</i> FN555150	G I DEF A VLVKQ
<i>O. nerka</i> FN555149	G I DEF A VLVKQ
<i>O. mykiss</i> FN544260	G I DEF A V W VKQ

Supplementary Table S1: Disclosed different haplotypes found within each species of Salmonids. Red: fixed sites in farm rainbow trout; Green: sites share between *Salmo salar* and *Salmo trutta*; Orange: sites shared between *Salmo trutta* and *Salvelinus alpinus*.

POLYMORPHIC SITES

Farm *O. mykiss* 91 R 133 M 218 M 389 A 445 C 488 R

Haplotype 1	A	A	C			G
Haplotype 2	A	A	C			A
Haplotype 3	G	C	A			A

Wild *O. mykiss* 91 R 133 M 218 M 389 R 445 M 488 R

Haplotype 1	A	A	C	A	C	A
Haplotype 2	A	A	C	A	C	G
Haplotype 3	G	C	A	A	C	A
Haplotype 4	A	A	C	G	C	A
Haplotype 5	G	C	A	A	A	A
Haplotype 6	G	C	A	A	C	G
Haplotype 7	A	A	A	G	C	A
Haplotype 8	G	C	C	A	C	A
Haplotype 9	G	C	A	G	C	A
Haplotype 10	A	A	A	A	C	A

Wild *S. salar* **60 K 67 K 152 M 397 Y 410 K 429 W 454 S 477 R 502 R 505 Y 530 K 587 Y**

Haplotype 1	T	G	C	T	T	A	G	A	A	C	T	T
Haplotype 2	G	T	A	C	G	T	C	G	G	T	T	C
Haplotype 3	T	G	C	T	T	A	G	A	A	C	G	T
Haplotype 4	G	T	A	C	G	T	C	G	G	T	T	T
Haplotype 5	T	G	C	C	T	A	G	A	A	C	G	T
Haplotype 6	T	G	C	T	T	A	G	G	A	C	T	T
Haplotype 7	T	G	C	C	G	T	C	G	G	T	T	C
Haplotype 8	T	G	C	T	G	T	C	G	G	T	T	C

Wild *S. trutta* **40 W 131 K 134 R 261 Y 382 465 R 477 R 486 Y 545 M**
Gap/R

Haplotype 1	A	G	A	T	A	G	A	C	C
Haplotype 2	T	G	A	C	-	A	G	T	C
Haplotype 3	T	G	A	T	A	A	A	C	C
Haplotype 4	A	G	A	C	A	G	A	C	C
Haplotype 5	T	G	A	C	A	G	A	C	C
Haplotype 6	T	T	G	C	-	G	A	C	C
Haplotype 7	A	G	A	T	A	A	G	T	C
Haplotype 8	A	G	A	T	-	A	G	T	C
Haplotype 9	A	G	A	C	A	A	G	T	C
Haplotype 10	A	G	A	C	-	A	G	T	C
Haplotype 11	T	G	A	C	-	G	A	C	C
Haplotype 12	T	T	G	C	-	A	G	T	A
Haplotype 13	A	G	A	C	A	A	G	T	A
Haplotype 14	A	G	A	C	A	A	G	C	C
Haplotype 15	T	G	A	C	A	A	G	C	C
Haplotype 16	T	G	A	C	-	A	G	T	A
Haplotype 17	A	G	A	C	G	G	A	C	C

Wild *S. alpinus* 95 M 209 Y 221 M 431 Y 462 K 465 R 492 Y 524 K

Haplotype 1	A	C	A	T	T	A	A	T
Haplotype 2	C	C	A	T	G	G	A	G
Haplotype 3	A	C	A	T	G	G	A	G
Haplotype 4	A	T	C	C	G	G	G	G
Haplotype 5	C	C	A	T	T	A	A	T
Haplotype 6	A	T	C	T	T	A	A	T
Haplotype 7	A	C	C	C	G	G	G	G
Haplotype 8	A	C	A	C	T	A	A	T
Haplotype 9	C	C	A	T	G	G	G	G
Haplotype 10	A	C	A	T	T	A	A	G
Haplotype 11	A	T	C	T	G	G	A	G
Haplotype 12	C	C	A	T	T	A	A	G
Haplotype 13	A	C	A	C	G	G	A	G
Haplotype 14	C	T	C	T	T	A	A	T
Haplotype 15	A	C	A	T	G	G	A	T
Haplotype 16	A	T	C	T	G	G	A	T
Haplotype 17	A	C	A	C	G	G	G	G
Haplotype 18	A	C	C	T	T	A	A	T
Haplotype 19	C	C	A	C	G	G	G	G