

Additional file 1: Figure S1. Teleost MHC I haplotypes

Table of Contents		Page
S1a	Summary of previously published zebrafish and medaka MHC I haplotypes	1
S1b	MHC Ia and Ib haplotypes from previously published Atlantic salmon and rainbow trout BACs.	2
S1c	MHC class I haplotypes in Atlantic salmon, rainbow trout, coho salmon and Northern pike genomes.	2

Figure S1a. Summary of previously published zebrafish and medaka MHC I haplotypes.

Zebrafish (*Danio rerio*) haplotype amino acid sequences used in this study originate from McConnell et al.(1) and medaka (*Oryzias latipes*) Hd-rR, HN1 and cab haplotype sequences are from Matsuo et al.(2), Tsukamoto et al.(3) and Nonaka & Nonaka (4) respectively. Genes are color coded dark green for proteasome PSMBs, yellow for TAP2, red for MHC I, blue for TAPBP and grey for other genes. Ψ denotes pseudogenes while red font shows location of PSMB8F sequences. Haplotype numbers and chromosomal location are indicated above and on the side of each region respectively. Zebrafish haplotypes are from (1) and medaka haplotypes from (2) (Hd-rR), (3) (HN1) and (4) (cab).

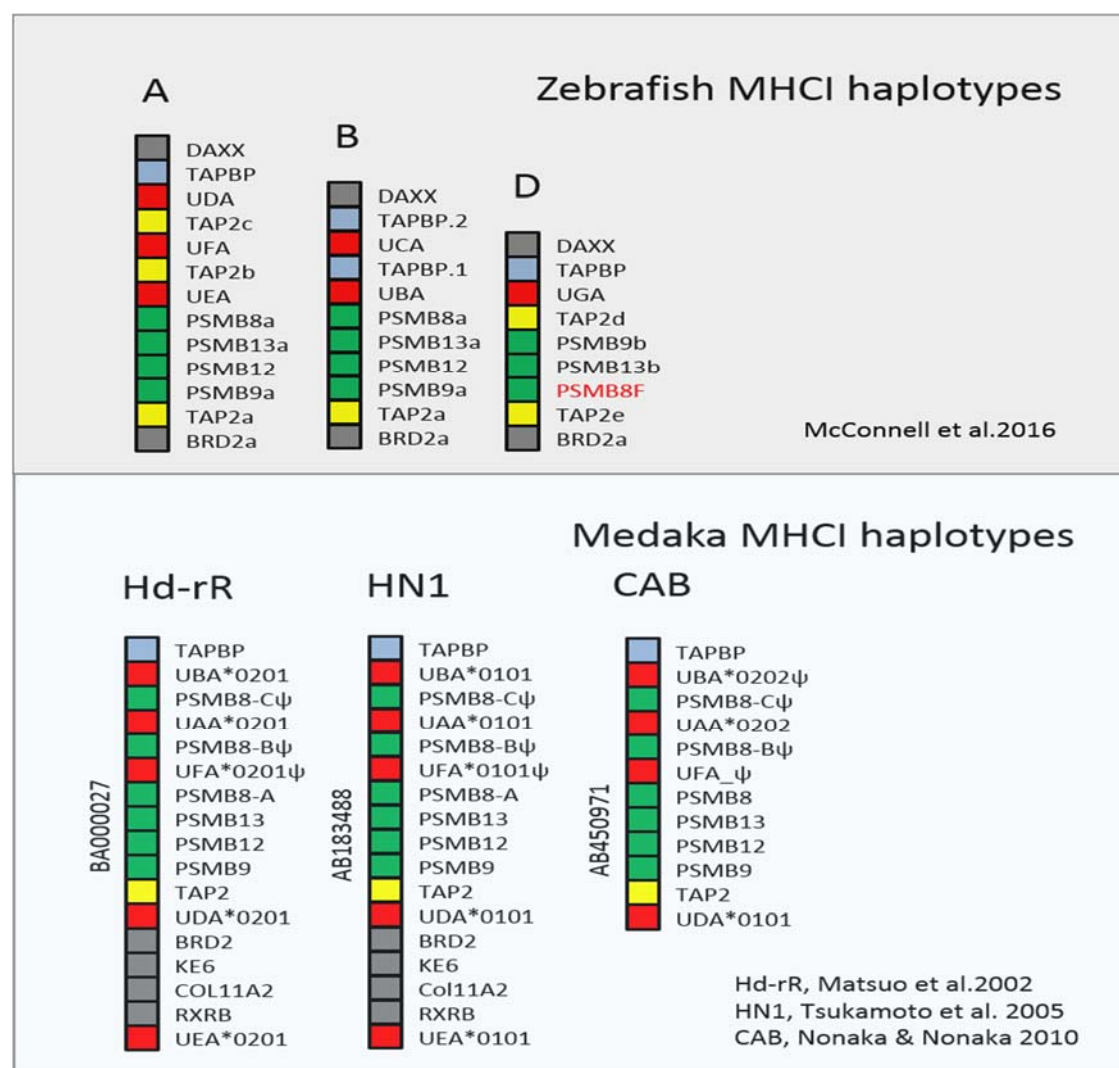


Figure S1b. MHC Ia and Ib haplotypes from previously published Atlantic salmon and rainbow trout BACs.

Schematic presentation of previously sequenced BAC sequences from Atlantic salmon and rainbow trout. Salmonid and pike MHC class I haplotype amino acid gene sequences used in this study originate from either previously sequenced BAC clones or genomes available in Genbank as follows: Atlantic salmon (*Salmo salar*) haplotypes from BACs (5, 6): Ia_#A BACs 868O01 (EF441211) and 92I04 (EF427384.1); Ia_#B BAC 714P22 (EF210363); Ib_#A BAC 8I14 (EF427379); Ib_#B BAC 438J08 (FJ969490); Atlantic salmon genome GCA_000233375.4 (7) haplotypes: Ia_#C Chr.27 NC_027326: 10.000.000-10.656.000; Ib_#C Chr.14 region NC_027313: 50.800.000- 59.500.000. Rainbow trout (*Oncorhynchus mykiss*) MHC I haplotypes from BACs (8): Ia_#A BAC (AB162342); Ib_#A BAC sequence (AB162343); Rainbow trout haplotypes from genome GCA_002163495.1 (unpublished GenBank assembly): Ia_#B genome Chr.18 (CM007952.1); Ib_#B genome Chr.14 (CM007948.1). Coho salmon (*Oncorhynchus kisutch*) genome GCA_002021735.1 (unpublished) haplotypes: Ia region LG17 NC_034190: 25300000-26000000, Ib MHC Ib region LG14 NC_034187: 22.700.000-23.500.000. Northern pike (*Esox Lucius*) genome GCA_000721915.3 (9) MHC I haplotypes: MHC Ia region LG10 (NC_025977.3), presumed allelic haplotype represented by unplaced genomic scaffold NW_017859580.1, and unplaced genomic scaffold NW_017859271.1. More details about each region can be found in Dataset S1. Sequence references are shown vertically alongside each region. Genes are color coded dark green for proteasome PSMBs, yellow for TAP2, red for MHC I, blue for TAPBP and grey for other genes. Ψ denotes pseudogenes. Haplotype numbers are indicated above the regions.

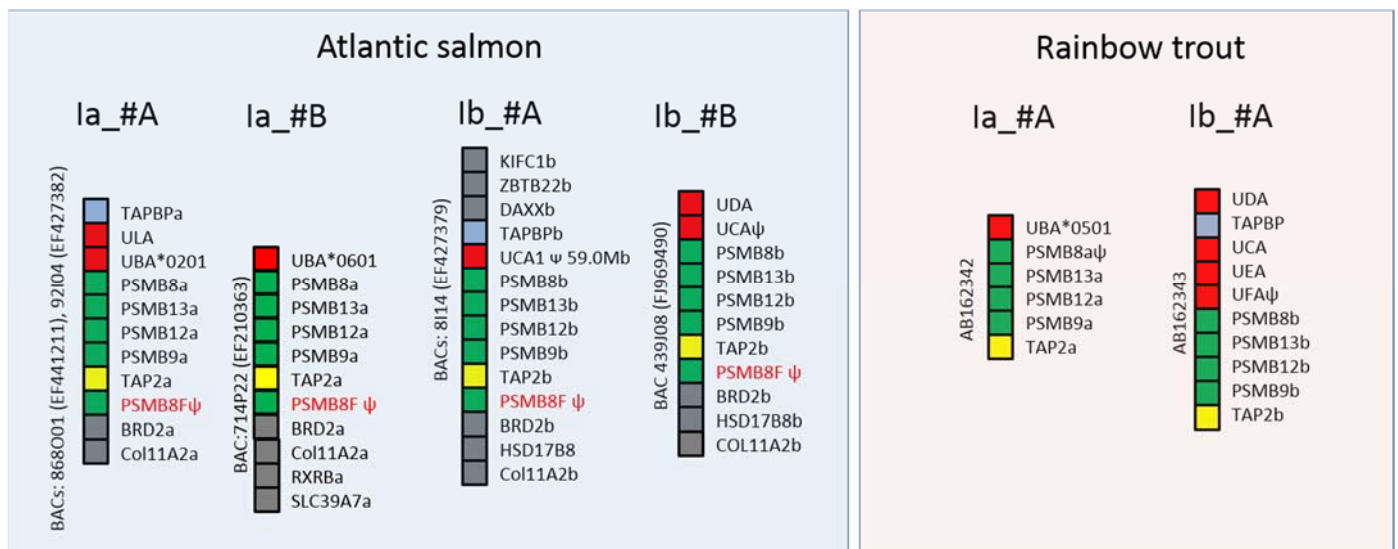
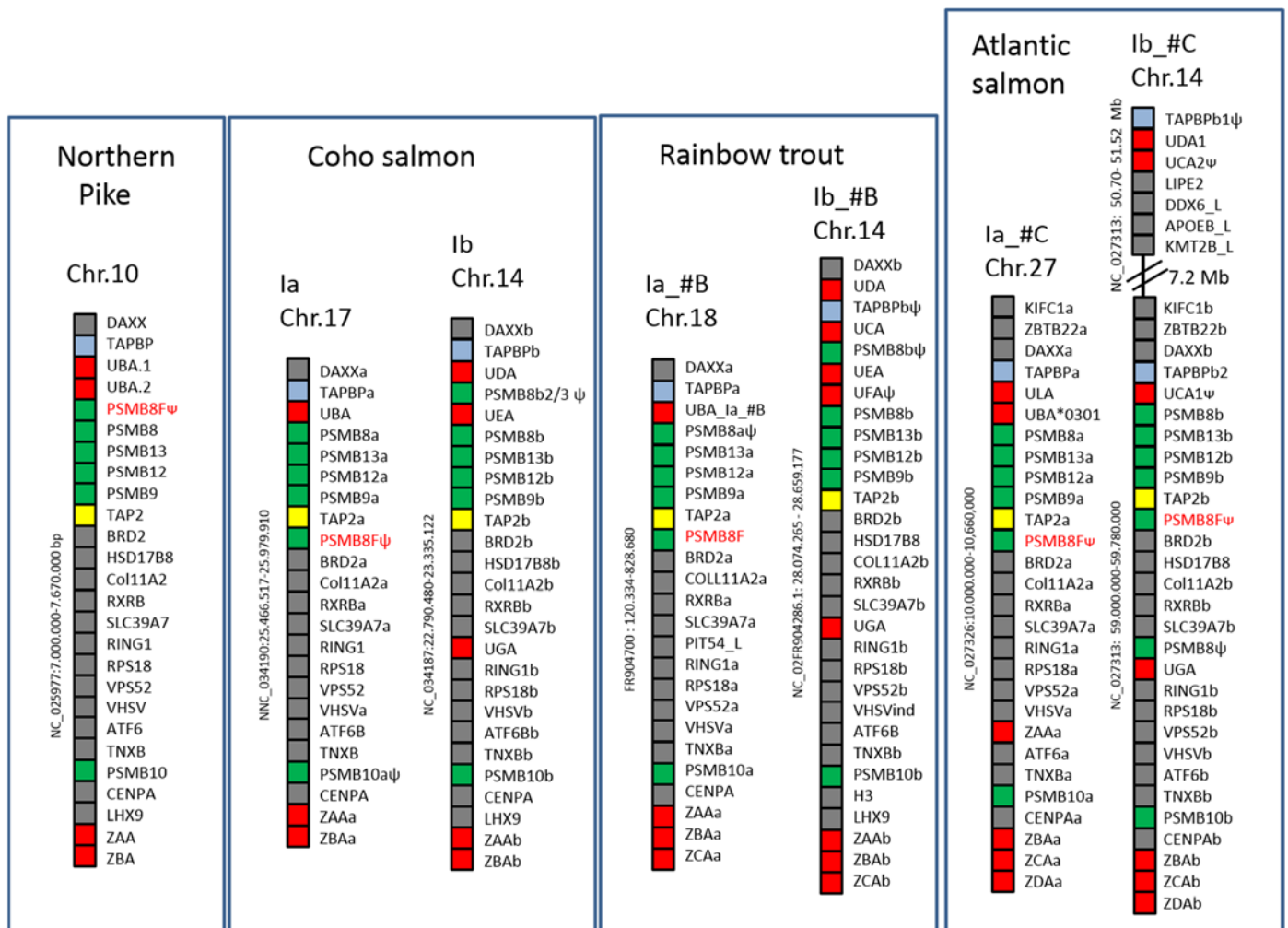


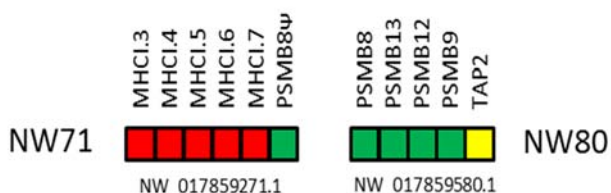
Figure S1c. MHC class I haplotypes in Atlantic salmon, rainbow trout, coho salmon and Northern pike genomes.

Schematic view of duplicate MHC Ia and Ib regions in Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), coho salmon (*Oncorhynchus kisutch*) and MHC Ia regions in Northern pike (*Esox lucius*) genomes. Regional references are as follows: Atlantic salmon (*Salmo salar*) haplotypes from genome GCA_000233375.4 (7) haplotypes: Ia_#C Chr.27 NC_027326: 10.000.000-10.656.000; Ib_#C Chr.14 region NC_027313: 50.800.000- 59.500.000. Rainbow trout (*Oncorhynchus mykiss*) MHC I haplotypes from genome GCA_002163495.1 (unpublished GenBank assembly): Ia_#B genome Chr.18 (CM007952.1); Ib_#B genome Chr.14 (CM007948.1). Coho salmon (*Oncorhynchus kisutch*) genome GCA_002021735.1

(unpublished) haplotypes: Ia region LG17 NC_034190: 25300000-26000000, Ib MHC1b region LG14 NC_034187: 22.700.000-23.500.000. Northern pike (*Esox Lucius*) genome GCA_000721915.3 (9) MHC1 haplotypes: MHC1a region LG10 (NC_025977.3), presumed allelic haplotype represented by unplaced genomic scaffolds NW_017859580.1, and NW_017859271.1. Sequence references are shown vertically alongside each region. Genes are color coded dark green for proteasome PSMBs, yellow for TAP2, red for MHC1, blue for TAPBP and grey for other genes. Ψ denotes pseudogenes. Haplotype numbers and chromosomal location are indicated above the regions. The trout Ia_#B haplotype contains a new yet undefined UBA allele. The two additional Northern pike scaffolds are most likely haplotypic variation, although organization is undefined in the genome assembly.



Additional Northern pike scaffolds



References:

1. McConnell SC, *et al.* (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
2. Matsuo M, Asakawa S, Shimizu N, Kimura H, & Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). *Immunogenetics* 53:930-940.
3. Tsukamoto K, *et al.* (2005) Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. *Immunogenetics* 57:420-431.
4. Nonaka MI & Nonaka M (2010) Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics* 62(5):319-332.
5. Lukacs MF, *et al.* (2010) Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. *BMC genomics* 11:154.
6. Lukacs MF, *et al.* (2007) Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). *BMC genomics* 8:251.
7. Lien S, *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* 533(7602):200-205.
8. Shiina T, *et al.* (2005) Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. *Immunogenetics* 56(12):878-893.
9. Rondeau EB, *et al.* (2014) The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS one* 9(7):e102089.

Additional file 2: Text S1. Deduced amino acid sequences used in this study

Region	Deduced amino acid sequences	Page
Atlantic salmon (<i>Salmo salar</i>) haplotypes		
Ia region*	Haplotype Ia_A; BACs 868O01 (EF441211) and 92I04 (EF427384.1) deduced amino acid sequences	2
Ia region*	Haplotype Ia_#B; BAC 714P22 (EF210363) deduced amino acid sequences	3
Ia region	Haplotype Ia_#C; Chr.27 genome NC_027326: 10.000.000-10.656.000; deduced amino acid sequences	4
Ib region*	Haplotype Ib_#A; BAC 8I14 (EF427379) deduced amino acid sequences	5
Ib region*	Haplotype Ib_#B; BAC 438J08 (FJ969490) deduced amino acid sequences	6
Ib region	Haplotype Ib_#C Chr.14 Region NC_027313:50.800.000-51.100.000 deduced amino acid sequences	7
Ib region	Haplotype Ib_#C continued Chr.14 region NC_027313: 59.000.000-59.498.000; deduced amino acid sequences	7
	Other deduced Atlantic salmon amino acid sequences	8
Rainbow trout (<i>Oncorhynchus mykiss</i>) haplotypes		
Ia region*	Haplotype Ia_#A; BAC (AB162342) deduced amino acid sequences	13
Ia region	Haplotype Ia_#B; Genome (GCA_002163495.1) Chr.18 (CM007952.1); deduced amino acid sequences	14
Ib region*	Haplotype Ib_#A; BAC sequence (AB162343) deduced amino acid sequences	15
Ib region	Haplotype Ib_#B; Genome Chr.14 (CM007948.1) deduced amino acid sequences	16
Coho salmon (<i>Oncorhynchus kisutch</i>) genome GCA_002021735.1		
Ia region	LG17 region NC_034190: 25300000-26000000 deduced amino acid sequences	18
Ib region	LG14 region NC_034187: 22.700.000-23.500.000 deduced amino acid sequences	19
Northern pike (<i>Esox Lucius</i>) genome GCA_000721915.3		
	MHC I region LG10 (NC_025977.3) deduced amino acid sequences	21
	Unplaced genomic scaffold Un_scaffold0763 NW_017859271.1 deduced amino acid sequences	22
	Unplaced genomic scaffold Un_scaffold1132 NW_017859580.1 deduced amino acid sequences	23
Deduced amino acid sequences from other species		
	Spotted gar (<i>Lepisosteus oculatus</i>) sequences	23
	TAPBP-like sequences from selected species	24
Zebrafish haplotypes 19B, 19D and other published sequences*		25
Medaka Hd-rR, HN1, cab haplotype published sequences*		27
References		29

*Grey coloured sequences are previously published.

A.salmon la region Haplotype la_A; BACs 868O01 (EF441211) and 92104 (EF427384.1) deduced amino acid sequences

BAC sequence reference: see Lukacs et al. [1, 2].

```
>SalmonTAPBP_a_#A ABO13869.1 EF441211:64.923-69.673
MANISTIIYKLSFLAFTYFIHVYGTSCPVLCEWVQEKPGRGGGFPAAMIQEKSLLYINTD
PESEETKSQQGPSADINHDRVYYVTDPAAILCSSLHPPEGSVHKPQCEINPFMPQ PSTV
QWVVP LTDSAHSPIYLQADWYSAALQGLDGLGLSSVMRAPTATKEPTVVL SVSSRTPLV
RSRLGEPVVLDCGFWM EATSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTD RFAETKEKRA
GLDFTALHETGNASLILQEAQVRHTGTIYICTVYLPYLLAQVAVELEIVEPPSL SIFPSPL
PLSVPGQVVMVKVQCEASGFFPLSLDFHWELTGPDGKVRPLGQGSVTGHRQGP DNTYSQT
SRLELDSAKLDLGRGGEVTCVAVHPGGTTRASVTLN VIGINGPSIEDSMAMVAVALGLYG
LIKIVSWTFSSGSDDTNSQEKKVK
>SalmonULA*0102_#A ABO13870.1 EF441211: 79.632-94.364
MKCFILLLLSISLHAASAMHSLRYVYTATSGMPDFPEFMTVGLVNGEPI SYDYDSIIRRE
TPRQDWMKEAVDPDYWNRNTQTSIGDEQTFKANIDVAKQRFNQTGGVHVYQNM YGCEWDD
EAGVTEGFDQYGYDGEDFLAFDLKTLKWIAPTPQSLITKLKWDNNMAQIQ QDKHYLTQTC
IEWLK KYLDY GKSTLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQ KDGQDHED
VEYGETLQNDGTFQKSSHLTVTPEEWKNNKYQCVVQVTG VKEDFIKVLTESEIKTNWGN
TNIGFVPANTSDVGSNSSHNTVPKE
>SalmonUBA*0201_#A EF441211: 133.909-140.046 a1 domain and EF427384.1:1-
8.682 a2-TM domains
MKSCILLFLGLIVLLHTVSAATNTLQYFYTATSGIDNFPEFVTMGIVNGH QIDHYDSITK
RAIQKAEWISGAVDPDYWKNTNTQIYAGTETVFVNNINVAKS RFNQTGVHVNQKMYGCEWD
DETGVTGFDQDGYDGEDFLAFDLKTLTWIAPTPQAVITKLKWD SNTAQNEYRKNYLTQT
CIEWLK KYLDY GKSTLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQ KDGQDHHE
DVEHGETLQNDGTFQKSSHLTVTPEEWKNNKYQCVVQVTG LQEDFIKVLTESEIKTNWN
DPNIVLIIGVVVALLLVVVAVVVGVVIWKKKSKKGFVPASTSDT DSDNSGRAAQMT
>SalmonPSMB8a_#A EF427384.1:15.837-18.682 ABQ59680.1
MALFDVSGYKSYSELRGQIIIGTGVGHFIDRPNKQFSVPVGVDP SGFLKSCSREGGVSIDL
NHGTTTLAFTFRHGVI VAVDSRASAGSYIASKEANKVIEINPYLLGTMSG SAADCQYWER
LLAKECRLYKLRNKQRI SVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKG PGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHR DAYS SGGVVNLY
HMQEDGWIKVCKEDVSELIHRYRKGMF
>SalmonPSMB13a_#A EF427384.1:19.476-24.495 ABQ59681.1
MALSNVLEIPTSGFNFKNVARNVALEGLLEGGHTKTLKPMKTGTTIAGL VCKEGVVLGAD
TRATSGEVVADKMC AKIHYISPNIYCCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVQY LAMGSGDLAALGI
LEDRFKPNMEMEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGV DYIRPYQESEYKDK
RRKRYKYGTGTT SILTEKIVPLELEVQKTQVRMDTA
>SalmonPSMB12a_#A EF427384.1:24.652-29.374 ABQ59686.1
MERHLMDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGSYVSSKTINKLIQVH DRIFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAAASVLKELCYNNKEELQAGFITAGW
DRKKGPQVYTVALGGM LLSQPFTIGGSGSTIYIGYADAKYKPDMSKEECLQFAKNALALA
MGRDNVSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE
>SalmonPSMB9a_#A EF427384.1:32.328-36.287 ABQ59682.1
MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETV VNRVMNKL SLLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDRKGGGQVYVTLNGLLSRQPFVAVGSGSAYVYGFDVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLV TIDEKGAEKCILGNELPTFYDQ
>SalmonTAP2a_#A EF427384.1:37.057-41.156 ABQ59683.1
MTKMMRLRTCALAMAVGLCIDITTLAFGASISKTPRTFDAFGNVVRLWV VAGIRLVLLLG
LTLTLTGSIKPVFKRWLAVHCF LAPVYETGRRMLYGSSPERVYGS LGLGSPSLWLLCTAA
AAAAALFWETTFPDSNGESNGKQKTQKARVLFMRVLYFYRPDTLLL VGAFIFLSLAVLCE
MFIFPYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMK
VELFGALVKQEIGFFETMKTGDITSRLSTDITLMGRAVALN VNVLLRTLIKTVGMLS LMM
SLSWKLTLLMLMETPITGLLQSVHDNYYQRLSKEVQDSMARANEAAGETVGGIRT VRSFK
TEQHEAGRYNDR LMDTHNLKTRRDTVRAVYLLLRRLTAVVMQVAMLYYGR LFIQRGQMST
GNLVSFILYQSDLADNIRTLIYIFGDM LNSVGAAGKVFEYLDREPQVSTKGT LQPETLTG
HVHFNNLSFSYPTQRKVLQGFSL ELRPGQLTALVGPSGGGKSTCVSLLERFYQPQQGE
```

```
ILLDGLPLQSYQHHLHRKIAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAH
SFISQLEKGYDTDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQ
EALASCPSQTLVIAHRLKTIERADQIILIDQGTVEQGTQELMDRKGSSYYKLKERLFT
EDDAPH
>SalmonPSMB8Fa_#A pseudogene EF427384.1: 50.877-50.960
HKRNTGWIKVSQEDVGDLYHRFYNDKK*
```

A.salmon Ia region Haplotype Ia_#B; BAC 714P22 (EF210363) deduced amino acid sequences

Bac sequence reference: see Lukacs et al. [1, 2].

```
>SalmonUBA*0601_#B ABQ01995.1 EF210363:10.646-12.879
VHVNQWMYGCEWDDAEAGVTEGFEQWGYDGEDFIAFDLKTCSWIAPTPQAVITKLKWDSDT
AQNEHRKNYYTQICIEWLKKYVDYGKSTLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGV
MVSQKQDGDHEDVEYGETLQNDGTFQKSSHLTVTPEEWKNNKYQCQVQVTVGKEDFI
KVLTESEIKTNWNDPNIVLIIVVVVALLLVAVVVGVVIWKKKSCKGFVPASTSDTSD
NSGRAAQMT
>SalmonPSMB8a_#B ABQ01989.1 EF210363:19.903-22.781
MALFDVSGYKSYSELRGQIIIGTGVGHFIDRPNKQFSVPVGVDPGFLKSCSREGGVSIDL
NHGTTTLAFTFRHGVIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYS GG VV NLY
HMQEDGWIKVCKEDVSELIHRYRKGMF
>SalmonPSMB13a_#B ABQ01990.1 EF210363:23.293-28.664
MALSNVLEIPTTGFFNFENVARNVALEGLLEGHTKTLKPMKTGTIAGLVCKEGVVLGAD
TRATSGEVVADKMCACIHYISPNIYCCGAGTAADTEKTTDLSSNLTIFSMNSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVQYLAMGSGDLAALGI
LEDRFKPNMEMEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQSEYKDK
RRKRYKYGTGTTISILTEKIVPLELEVQKTQVRMDTA
>SalmonPSMB12a_#B ABQ01993.1 EF210363:28.644-33.489
MERHLMDSQIKGVSTGTITLAVTFNGGVIIGSDSRASIGGSYVSSKTINKLIQVHDIRFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAASVLKELCYNNEELQAGFITAGW
DRKKGPQVYTVALGGMLLSQPFTIGGSGSTYIYGADAKYKPDMSKEECLQFAKNALALA
MGRDNVSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE
>SalmonPSMB9a_#B ABQ01991.1 EF210363:36.392-40.350
MLEESSEPGWLSEEVKTGTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ
>SalmonTAP2a_#B ABQ01992.1 EF210363:41.120-45.219
MTKMMLRTCALAMAVGLCIDITTLAFGASISKTPRTFADFNGNVRLWVWVAGIRLVLLLG
LTLTLTGSIKPVFKRWLAVHCFAPVYETGRMRLYGSSPERVYGSGLGLGSPSLWLLCTAA
AAAAALFWETTFPDSNGESNGKQKTQKARVLFMRVLYFYRPDTLLLVGAFIFLSLAVLCE
MFIFPYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMK
VELFGALVKQEI GFFETIKTGDITSRLSTDITLMGRAVALNVNVLRLTIKTVGMLSLMM
SLSWKLTLLMLMETPITGLLQSVHDNYYQRLSKEVQDSMARANEAAGETVGGIRTVRSFK
TEQHEAGRYNDRMLDTHNLKTRRDTVRVYLLLRRLTAVVMQVAMLYYGRLLFIQRGQMST
GNLVSFILIYQSDLADNIRTLYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTG
HVHFNNLSFSYPTQRERKVLQGFSLRLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGE
ILLDGLPLQSYQHHLHRKIAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAH
SFISQLEKGYDTDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQ
EALASCPSQTLVIAHRLKTIERADQIILIDQGTVEQGTQELMDRKGSSYYKLKERLFT
EDDAPH
>SalmonPSMB8Fa_#B EF210363:54.985-55.056 no predicted ORF
HKRNTGWIKVSQEDVGDLYHRFYNDKK*
```

A.salmon MHC Ia Haplotype Ia_#C; Chr.27 genome NC_027326: 10.000.000-10.656.000 deduced amino acid sequences.

Atlantic salmon genome assembly: See Lien et al. [3].

>SalmonTAPBP_a_#C NP_001117077.1 NC_027326:10.023.359-10.039.558
MANISTIIYKLSFLAFTYFIHVGTSQCVLECFVQEKPGRGGGFPAAMIQEKSLLYINTD
PESEETKSQQGPSADINHDRVYVTDPAAILCSSLHPPEGSVHKPQCEINPFMPQ PSTV
QWVPLTDSAHSPYIYLQADWYSAALQGLDGLSSVMRAPATKEPTVVLSSVSRTPLV
RSRLGEPVVLDCGFWMPEATSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTDRAETKEKRA
GLDFTALHETGNASLILQEAQVRHTGTNICTVYLPYLLAQVAVELEIVEPPSLSIFFPSPL
PLSVPGQVVKVQCEASGFFPLSLDFHWELTGPDKVRPLGQGSVTGHRQGPNTYSQTSR
LELDSAKLDLGRGGEVTCVAVHPGGTTRASVTNLVIGINGPSIEDSMAMVAVALGLYGLI
KIVSWTFSSGSDDTNSQEKVK"

>SalmonULA_#C XP_014032820.1 NC_027326:10.037.566-10.054.430
MKCFILLLLSISLHAASAAMHSLRYVYTATSGMPDFPEFVTVGLVNGEPISYYDSIIRSE
TAPQDWMKEAVDPDYWNRTQTSIGDEQTFKANIDVAKQRFNQTTGGVHVYQNMYGCEWDD
EAGVTEGFDQYGDGEDFLAFDLKTLKWIAPTPQSLITKLKWDNNMAQIQQDKHYLTQTS
IEWLKLYLDYGKSTLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQKDGQDHED
VEYGETLQNDGTFQKSSHLTVTPPEWKNNKYQCVVQVTGKEDFIKVLTESEIKTNWGN
TNIGFVPANTSDVGSNSSHNTAPKE"

>SalmonUBA*0301_#C XP_014032819.1 NC_027326:10.122.009-10.149.394
MKCFILLLLGIALHSSAATHSLRYVYTATSGIPDFPEFVTVGLVNGEPISYYDSIIRR
ETPRQDWMKTEGSDYWESQTQVSIGSEQTFKANIDVAKQRFNQTTGGVHVNQKMYGCEWD
DETGVTGFDQDGYDGEDFLAFDLKTLTWIAPTPQAVITKLKWDNNTAQNEYRKNYLTQT
CIEWLKLYLDYGKSTLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQKDGQDHHE
DVEHGETLQNDGTFQKSSHLTVTPPEWKNNKYQCVVQVTGLQEDFIKVLTESEIKTNWN
DPNIVLIIGVVVALLLVVAVVVGVIWKKKSKKGFVPASTSDTSDNSGRAAQM

>SalmonPSMB8a_#C XP_014032817.1 NC_027326:10.153.897-10.157.743
MALFDVSGYKSYSELRGQIIIGTGVGHFIDRPNKQFVSVPGVDPGSGFLKSCSREGGVSIDL
NHGTTTTLAFTFRHGIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYSGGVNNLY
HMQEDGWIKVKCEDVSELIHRYRKGMF

>SalmonPSMB13a_#C NC_027326:10.158.350-10.163.366 incomplete sequence
MALS NVLEIPTSGFNFENVARNVALEGLLEGGHTKTLKPMKTGTIAGLVCKEGVVLGAD
TRATSGEVVADKMKAKIHYISPNISSCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
MAVRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVQYLAMGMEAEKELVRDAIHSGIM
SDLGSGNNIDICVITKQGVYIRPYQSESGTGTTSILTEKIVPLELEVQKTVQRM

>SalmonPSMB12a_#C XP_014032816.1 NC_027326:10.163.427-10.168.543
MERHLMDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGSYVSSKTINKLIQVHDRIFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAASVLEKELCYNNKVMQLQAGFITAGW
DRKKGPQVYTVALGGMLLSQPFTIGSGSGSTYIYGADAKYKPDMSKEECLQFAKNALALA
MGRDNVSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE"

>SalmonPSMB9a_#C XP_014032548.1 NC_027326:10.171.104-10.175.148
MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVITIDEKGAEKCILGNELPTFYDQ"

>SalmonTAP2a_#C XP_014032814.1 NC_027326:10.176.036-10.180.873 and TSA
GBRB01034973.1
MMLRMCAFAMAVGLCIDITTFCAPCFGESISETGPITFGTFGNVRLVWVAGIRLVLLLG
LTLLTLGSIKPVFKRWLAVHCFAPVYETGRRMLYGSSPERVYGSGLGSPSLWLLCTAA
AAAAALFWETTFPDSNGESNGKQKTQKARVLFMRVLYFYRPDTLLLVGAFIFLSLAVLCE
MFIPFYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMK
VELFGALVKQEI GFFETIKTGDITSRLSTDTLMGRAVALNVNLLRLTIKTVGMLSLMM
SLSWKLTLLMLMETPITGLLQSVHDNYYQRLSKEVQDSMARANEAAGETVGGIRTVRSFK
TEQHEAGRYNDRMLDTHNLKTRRDTVRVYLLLRRLTAVVMQVAMLYYGRLFIQRGQMST
GNLVSFILYQSDLADNIRTLYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTG
HVHFNNLSFSYPTQERKVLQGFSLRLPGQLTALVGPSSGGKSTCVSLLERFYQPQQGE
ILLDGLPLQSYQHLYLHRKIAMVQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAH
SFISQLEKGYDTDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQ
EALASCPSQTLVLIAHRLKTIERADQIILIDQGTVQEQGTHQELMDRKGSSYYKLKERLFT
EDDAPH

>SalmonPSMB8Faψ_#C No predicted ORF NC_027326:10189712-10189792

HKRNTGWIKVSQEDVGDLYHRFYNDKK

>SalmonPSMB10a_#C XP_014032774.1 NC_027326:10.580.777-10.585.553
MLHNSRPPQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILGA
DTRATDDMVVADKNCMKIHYIAPNIYCCGAGVAADAEVTTQMSSSNVELHSLSTGRPPLV
VMVTRQLKQMLFRYQGHIGSSLIVGGVDVTGAHLYSVYPHGSYDKLPFLTMTGSGAGAAIS
IFEDRYRPNMELEEAKKLVRDAIAAGIFCDLGSGSNVDLCVITQAGVQYLRSDQPAQKG
KKEGQYKYKPGTTAVLTCTVTPLPLDVVDESIQLMDAQ"

A.salmon Ib region Haplotype Ib_#A; BAC 8I14 (EF427379) deduced amino acid sequences

BAC sequence reference:see Lukacs et al.[1, 2].

>SalmonTAPBPb_#A ABQ59655.1 EF427379:39.450-49.132
MTNISTILQSFLAFSYFMHVYGASCPVLECFVQEKPGGGFHAAMSQEKSLLYINTDPDS
EETRSKQGLSTDHDRVYVTDPAATLCSSSMHPPEGSVQKPQCEINPFMPQPSTVQWTV
LTDSAHSPYIQLADWYTAALQGFDGQLRLSNVMRAPATKEPKVLLSVSSRTFMIRSRLG
EPVVLDCGFWDASSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTDRAETQEEGAGLDFT
ALHETGNASLILQEAQVRHSGTYICTVYLPYLLAQVAVELEIVEPPSLSIFPSPLPLSMP
GQVVTVQCEASGFYPLSLEFHWVLTGPGGRVSPLGQGSVTGHRQGPSTYSQTSRVELDS
AKLDLGRGGEVTCVAVLRGGTRRASVTLNITGVSAPSIEDYMAVAVALGIYGLIKVVS
TFSSGSDVAYTQEKVK

>SalmonUCA_#A EF427379:60.574-70.123
MKGCIILMFLGIVYHREAFGVTHSLKYFYTASSKVPNFPEFVVVGMVDGVQIDHYDSNSQR
MVPKQDWMNKQTDAAEYWERETGIAFDSQQVFKDDVNILKQRFNQSGGVHVLQYIYGCSWD
DETEQRDGFWAAGGFGQLGYNGEDFLVYDMNTLTWKALKQQADVMDKWNDRDISRLVFWK
TYFSQTCIECLKKQVVNGKSSLRTVPSPVSLLOKTPSSPVTCHATGFYPSGVMVFWQKDG
QEQHEDVEHGEILHNDGTFQKSTHLRVTPEEWKNNKYQCVVQVTGIKEDFIKVLTESEI
QTNWDDPAPIIVPIIGGVVALLLVVVVVVVVVGVVIWKKKSKKGFVPASTSATDSDNSWK
GFQKT*

>SalmonPSMB8b_#A ABQ59648.1 EF427379:72.420-75.566
MALFDVSGYKSHAGLRGQILGTGVGHVDRPNQEFAPVGVDPGSGFLKSCSREGGVSIDL
NHGTTTTLAFTFRHGIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIIGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVVDSGYREDMTVEEAYELGRRGITHATHRDAYSGGVVNLY
HMQEDGWIKVCKEDVSELIHRYRKGMF

>SalmonPSMB13b_#A ABQ59647.1 EF427379:76.573-80.426
MALTNVETPASGFNFENVSRNVALEGLLEGGHTKAPKPMKTGTTIAGVVCKDGVVLGAD
TRATSGEVVADKMKAKIHYISPNNYCCGAGTAADTEKTTDMLSSNLTIFSMTSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
LEDRLFKNMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQESEYKDK
RQRRYKYRPGTTSILTEKIVPLELEVQETVQRMATA

>SalmonPSMB12b_#A ABQ59652.1 EF427379:80.282-87.669
MERHFMDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGSYVSSKTINKLIQVHDRIFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAAASVLKELCYNNEELQAGFITAGW
DRKKGPQVYTVALGGMLLSQPFTIGSGSTYIYGYADAKYKPDMSKEECLQFATNALALA
MGRDNVSGGVAHLVVITEEGVEHVVI PGDKLPKFHDE

>SalmonPSMB9b_#A ABQ59649.1 EF427379:88-882-93.220
MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVITIDEKGAEKCILGNELPTFYDQ

>SalmonTAP2b_#A ABQ59650.1 EF427379:95.267-99.916
MMRRTCVFTMAVGLCIDITTFCATGLGASISKTPISFDVFGNLVRLWVEAGIRLVLLFG
LSLLTLGSIKPVLRWLAVHCFLAPVYETGKLMHLHGSSPESPYGSLGGPSLWLLCTAAAA
AAALFWEKTFPDSKEESNGKEKTQKARALFMRVLYFYRPDTLLLVGAFIFLALAVLSETF
IPFYTGKVIDILASQYKWNDFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMKVE
LFGALVKQEISFFETIKTGDITSRLSTDTTKMARALALNVNVLRLTIKTVGMLSLMMSL
SWKLTLLMLMETPITGLLQGVYDNYLRLSKEMQDSMARANEAAGETVAGIRTVRSFNTE
RSEAGRYDHRLMDTHNLKTRRDTVRAVYLLKRLTALVMQVAMLYYGRLLFIQRGQMSTGN
LVSFILYQSNLGANIRTLIYIFGDMLNSVGAAGKVFYLDREPQVSTKGTLPETLTGHV

```
HFHNLSFSYPTRQGRKVLQGFSLRLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGEIL
LDGQPLHSYQHHLHHRKVAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSF
ISQLEKGYDTDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQVEA
LASCPSQTLVLIAHRLKTIERADRIILIDRGSVLEQGTQELMDRKGYYKLRERLFTED
DTS
>SalmonPSMB8Fb_#A EF427379:110.523-110.606
HMRETGWINVSDQEDVADLYHHLYNEKK
```

A.salmon lb region Haplotype lb_#B; BAC 438J08 (FJ969490) deduced amino acid sequences

BAC sequence reference: see Lukacs et al. [1, 2].

```
>SalmonUDA_#B FJ969490:15.583-29.784 ACY30371.1
MKGFI LMFMTCHLFEAFGVTHSLKHFTYASSKVTNFPEFMVVGMDGVQIDHYDSNIQR
MVPKQDWMNKQTEAEYWERETGIAFDSQQVFKDDVNILKQRFNQSGGVHVLQYIYGCSWD
DETEQRDGFQGLGYNGEDFLVYDMNTLTWKALKQQADVMRDKNRDISRLVFWKTYFSQT
CIECLKKQVVNGKSTLRTAPPSVSLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHED
VEHGEILHNDGTFQKSTHLRVTPEEWKNNKYQCVVQVTGIKEDFIKVLTESEIQTNWGD
PAPII VPIIGGVALLLVVVVGVVIWKKKSKKGFVPASTNDTDSVYSGKDLLKT
>SalmonPSMB8b_#B FJ969490:79.890-83.013 ACY30372.1
MALFDVSGYKSHAGLRGQILGTGVGHLVDRPNQEFAPVGVDPGFLKSCSREGGVSIDL
NHGTTTTLAFTFRHGVI VAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRI SVSAASKLLCNMMLGYRGMGLSMGSMIIGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVVDSGYREDMTVEEAYELGRRGITHATHRDAYS GGVVNLY
HMQEDGWIKVCKEDVSELIHRYRKGMF
>SalmonPSMB13b_#B FJ969490:84.043-87.912 ACY30373.1
MALTNVETPASGFNFENVSRNVALEGLLEGHTKAPKPKMTGTTIAGVVCKDGVVLGAD
TRATSGEVVADKMACAKIHYISP NMYCCGAGTAADTEKTTDMLSSNLTIFSM TSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
LED R FKP NMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGV DYIRPYQESEYKDK
RQRRYKYRPGTTSILTEKIVPLELEVQETVQRM DTA
>SalmonPSMB12b_#B FJ969490:87.748-95.135 ACY30374.1
MERHFMDSQIKGVSTGTTILAVTFNGGVIIIGSDSRASIGGSYVSSKTINKLIQVHDRIFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAASVLKELCYNNKEELQAGFITAGW
DRKKGPQVYTVALGGM LLSQPFTIGSGSTYIYGADAKYKPDMSKEECLQFATNALALA
MGRDNVSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE
>SalmonPSMB9b_#B FJ969490:96.348-100.683 ACY30375.1
MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETV VNRVMNKL SLLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ
>SalmonTAP2b_#B FJ969490:102.730-107.377 ACY30376.1
MMRRTCFTMAVGLCIDITTF CATGLGASISKTGPISFDVFGNLVRLWVEAGIRLVLLFG
LSLLTLGSIKPV LKRWLAVHCF LAPVYETGKLMLHGSSPESPYGSLGGPSLWLLCTAAAA
AAALFWEKTFPDSKEESNGKEKTQKARALFMRVLYFYRPDTLLL VGAFIFLALAVLSETF
IPFYTGKVIDILASQYKWNDFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMKVE
LFGALVKQEISFFETIKTGDITSRLSTDTTKMARALALNVNLLRTL IKTVGMLSLMMSL
SWKLTLLMLMETPITGLLQGVYDNYYLRLSKEMQDSMARANEAAGETVAGIRTVRSFNTE
RSEAGRYDHRLMDTHNLKTRRDTVRVYLL LKRLTALVMQVAMLYYGR LFIQRGQMSTGN
LVSFILYQSNLGANIRTLIYIFGDM LNSVGAAGKVFEYLDREPQVSTKGTLPETLTGHV
HFHNLSFSYPTRQGRKVLQGFSLRLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGEIL
LDGQPLHSYQHHLHHRKVAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSF
ISQLEKGYDTDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQVEA
LASCPSQTLVLIAHRLKTIERADRIILIDRGSVLEQGTQELMDRKGYYKLRERLFTED
DTS
>SalmonPSMB8Fb_#B FJ969490:117.938-118.021
HMRETGWINVSDQEDVADLYHHLYNEKK*
```

**A.salmon Ib region haplotype Ib_#C; Chr.14 Sally genome
NC_027313:50.800.000-59.500.000 deduced amino acid sequences**
Atlantic salmon genome assembly: see Lien et al.[3].

A.salmon Haplotype Ib_#C; Chr.14 Region NC_027313:50.800.000-51.100.000

```
>SalmonTAPBPb_#C_50_XP_013996968.1_NC_027313:50.968.347-50.981.043
PAATLCSSSLHPPEGSVQKPQCEINPFMPQPSTVQWTVPLTDSAHSPIYLQADWYTAALQ
GFDGQLRLSNVMRAPTATKEPKVLLSVSSRTFMIRSRLGEPVVLDCGFWDASSPLSGSG
FAVEWRYQFRGDGRLVLAYDGKTDRAETQEEGAGLDFTALHETGNASLILQEAQVRHSG
TYICTVYLPYLLAQVAVELEI
>SalmonUDA1_#C_50_NC_027313.1:50.992.799-50.986.338
VTHSLKYFYTASSGVPNFPEFPMVGMVDGVQIDHYDSNSQRMVMPKQDWMNKQTEAEYWER
ETGIAFDSQQVFKDDVNILKQRFNQSGGVHIVQKTYGCEWDDDETGTGGFTQHGYDGEDF
LGIDMKTFTWIAPKQQAIEITRLKWNHDQAGLAFWHNYLTQTCIEWLKKYVDYGKSTLRKV
PPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHEDVEHGETLPNDGTFQKSAH
LTVTSEEWKINQYQCVVQVTGIKEDFIKVLTESEIQTNW
>SalmonUCA2_#C_51_NC_027313.1:51.002.977-51.037.583
MKGFIILMFMGTCHLFEAFGVTHSLKYFYTASSGVPNFPEFPMVGMVDGVQIDHYDSNSQR
MVPKQDWMNKQTEAEYWERETGIAFDSQQVFKADVDILKQRFNQSGGVHVLQYIYGCSWD
DETEQRDGF*QLGYNGEDFLVYDMNTLTWKALKQQADVMRDKNRDISRLVFWKTYFSQT
CIECLKKQVNVNGKSTLRVPPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHEDV
ENGETLHNDGTFQKRTHLKVTSEEWKNNKYQCVVQVTGIKEDFIKVLTESEIQTNW
```

A.salmon Haplotype Ib_#C continued; Chr.14 region NC_027313: 59.000.000-59.498.000.

```
>SalmonTAPBPb_#C_XP_013997225.1_NC_027313:59.041.175-59.053.861
MTNISTILQSFLAFSYFMHVYGASCPVLECFWVQEKPGGGFHAAMSQEKSLLYINTDPDS
EETRSKQGLSTDHDRVYVTDPAATLCSSSMHPPEGSVQKPQCEINPFMPQPSTVQWTVPL
LTDSAHSPIYLQADWYTAALQGFQDGLRLSNVMRAPMATKEPKVLLSVSSRTFMIRSRLG
EPVVLDCGFWDASSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTDRAETQEEGAGLDFT
ALHETGNASLILQEAQVRHSGTYICTVYLPYLLAQVAVELEIVEPPSLSIFPSPLPLSMP
GQVTVQCEASGFYPLSLEFHWVLTGPGGRVSPGLQGQSVTGHRQGPSTYSQTSRVELDS
AKLDLGRGGEVTCVAVHRGGTTRASVTLNITGVSAPSIEDYAMAMVAVALGIYGLIKVVS
WTFSSGSDVAYTQEKVK"
>SalmonUCA_#C_NC_027313.1:59.088.400-59.091.609_XR_001320594.1
VTHSLKYFYTASSGVPNFPEFPMVGMVDGVQIDHYDSNSQRMVMPKQDWMNKQTEAEYWER
ETGIAFDSQQVFKDDVNILKQRFNQSGGVHVLQYIYGCSWDDDETEQRDGFWAAGGFGQLG
YNGEDFLVYDMNTLTWKALKQQADVMRDKNRDISRLVFWKTYFSQTCIECLKKQVNVNGK
STLRVPPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHEDVEHGEILHNDGTF
QKSTHLRVTPEEWKNNKYQCVVQVTGIKEDFIKVLTESEIQTNW
>SalmonPSMB8b_#C_NP_001117007.1_NC_027313:59.095.880-59.098.288
MALFDVSGYKSHAGLRGQILGTGVGHLVDRPNQEFAPVGVDPGFLKSCSREGGVSIDL
NHGTTTTLAFTFRHGIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIIGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVVDSDGYREDMTVEEAYELGRRGITHATHRDAYS GG VV NLY
HMQEDGWIKVCKEDVSELIHRYRKGMF"
>SalmonPSMB13b_#C_XP_013997221.1_NC_027313:59.099.410-59.103.327
MALTNVETPASGFNFENVSRNVALEGLLEGGHTKAPKPMKTGTIAGVVCKDGVVLGAD
TRATSGEVVADKMCACIHYISPNNMYCCGAGTAADTEKTTDMLSSNLTIFSMTSGRNPV
MAVNILQDMLFRYRGQIGASLILGGVDCGTNHLTYVGPYGSIDNVPYLAMSGDLAALGI
LEDRFKPNMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQESEYKDK
RQRRYKYRPGTTSILTEKIVPLELEVQETVQRMDDTA"
>SalmonPSMB12b_#C_XP_013997223.1_NC_027313:59.103.305-59.111.118
MDSQIKGVSTGTTLAVTFNGGVIIGSDSRASIGGSYVSSKTINKLIQVHDRIFCCIAGS
LADAQAVTKAAKFQISFHSIQMESPLVKAAASVLKELCYNNKEELQAGFITAGWDRKKG
PQVYTVALGGMLLSQPFTIGSGSTYIYGYADAKYKPDMSKEECLQFATNALALAMGRDN
```

VSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE"

>SalmonPSMB9b_#C NP_001117186.1 NC_027313:59.112.113-59.116.449
MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVITIDEKGAEKCILGNELPTFYDQ"

>SalmonTAP2b_#C NP_001117161.1 NC_027313:59.118.496-59.123.144
MMRRTCVFTMAVGLCIDITTFCATGLGASISKTGPISFDVFGNLVRLWVEAGIRLVLLFG
LSLLTLGSIKPVLRWLAVHCFLAPVYETGKMLMLHGSSPESPYGSLGGPSLWLLCTAAAA
AAALFWEKTFPDSKEESNGKEKTQKARALFMRVLYFYRPDTLLLVGAFIFLALAVLSETF
IPFYTGKVIDILASQYKWNDFLTAILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMVVE
LFGALVKQEISFFETIKTGDITSRLSTDTTKMARALALNVNVLRLTIKTVGMLSLMMSL
SWKLTLLMLMETPITGLLQGVYDNYLRLSKEMQDSMARANEAAGETVAGIRTVRSFNTE
RSEAGRYDHRLMDTHNLKTRRDTVRAVYLLKRLTALVMQVAMLYYGRLFIQRGQMSTGN
LVSFILYQSNLGANIRTLIYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTGHV
HFHNLSFSYPTQRQKVLQGFSLRLPGQLTALVGPSSGGKSTCVSLLERFYQPQGEIL
LDGQPLHSYQHHLHRKVAMVQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSF
ISQLEKGYDTDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQEA
LASCPSQTLVLIAHRLKTIERADRIILIDRGSVLEQGTHTQELMDRKGYYKLRLRERLFTED
DTS

>SalmonPSMB8Fb_#Cψ No predicted ORF NC_027313:59.133.717-59.133.797 pseudogene
HMRETGWINVSEDVADLYHHLYNEKK

>SalmonPSMB8b_#Cψ XP_013997205.1 NC_027313:59.496.446-59.497.368 pseudogene
MNTFDFDLTWDVIKSLFIFKTALKETNKIIEINHYLLGTMSGSAADCQYWERLLAKECRL
YKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKVQGGRE

>SalmonUGA_#C XP_013997204.1 NC_027313:59.497.676-59.507.931
MKTRLISAMKIYFVLLSCIHGALSVIHSLRYFYTSSSGISDFPEFVDMGMVNDQVISHYD
SITKRKVPKQSWMGKVFDQQYWDSTTEDLRGAEKVFKNNLQTAQKRFNQTGGMHISQDMY
GCEWDDDETGLTEGFHHIGYDGDLLVFDLKRATWIASVPQALHSMKWEGDPSSIESEKR
YLTQDCIVWLKKYLEYGKTTLQRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDG
QDHHEDVENGETLHNDGTGFQKRTHLKVTSSEWKNNKYQCVVQVTGIKEDFIKVLTESEI
QTNRGVNTIGSAPIIGVVVALLVVVVVVVGLVMWRRKSKKGFVPASTSDTDSSENSGKGA
QKI

>SalmonPSMB10b_#C XP_013997188.1 NC_027313.1: 59.678.647-59.684.589
MLNNSRPYQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILGA
DTRATDDMVADKNCMKIHYIAPNICCGAGVAADAETVTQMMSSNVELHSLSTGRPPLVV
TVTRQLKQMLFRYQGHIGSSLIVGGVDVTGAHLYSVYPHGSYDKLPFLTMGSGAGAAISI
FEDRYRPNMELEEAKKLVRDAIAAGIFCDLGGSGSNVDLCVITQAGVQYLRSDQPAQKKG
KEGQYKYKPGTTAVLTKTVTPLPLHVVDESIQLMDTQ

Other deduced Atlantic salmon amino acid sequences used in this study identified in the Atlantic salmon genome [3]

>SalmonPSMB7ps_ssa11 NC_027310.1:48.402.865-48.410.166 keep
TLSVCQTQLGGFSFENGKRLGLVEDGLVLGADTRATEGMIVADKNCISKIHYICPSSCCGA
GTAADTEMTTQIISSNLELHSLSTGRLPVATANRMLKQMLFRYTTH

>SalmonCANXa calnexin ssa05 NC_027304.1:27.384.003-27.403.278
XP_014054814.1
MELNVRCVLLAVALCSTLLLTVAAHQEEEEPIMELAEDMGVEDELEDLGLGEELLDGE
VEPEDADTPPGPPPAPKVITYKAPEPMGEHFFAESFDRGTLDVWLSKAKKEDIDEDIAKY
DGKWEVEDMKDSKLPDGLVLKSLRAKHHAISAQLLRPFIFDTKPLIIQYEVNFQQGIDC
GGAYVKLLSQTPDLNLDEFVDKTPYTIMFGPKCGEDYKLHFIHRHKNPKTGEEYEEKHAK
KPDADLRITYYTDKKTHTLYTLVVPDNSFEVLVDQTVVNSGNLLTDMTPPINPAEIEDPD
DHKPEDWDERPKIQDPDAVKPEDWDEDAPKQIPDEDAVKPDGWLDDQPEYTSDDPAVKPE
DWDEMDGWEAPQIPNALCETAPGCGAWQRPIMIDNPYKKGWKAPMIDNPYQGVWKPR
KIANPDFFEDLHPRMTPFNAVGLELWSMSSDIFFDNFFITNERHTAERWANDGWGLKKA
AEGAAEPGLVNQMMTAADERPWLWVVYVLTAVPLILIVVFFCTGKKAAPAAADYKKT
EPQPDVKEEEEKAEEDQVKEEKSQPAAAEKKSDAEDSPAEEKEEEEEEEEEEEEEEEEE

EA AVNEEEEEAA TDQVRDKTTTR

>SalmonCANXb ssa09 NC_027308.1:57.626.097-57.639.090 XP_014067456.1

MELKVRYVLLLVAVGLWSTLLLTVTVAHQDEEDEPIVEMGGDMDEDDMEELDHGEELLDDG
EVEADMPPGPPSVPKVYKVPPEPTGEHFFAESFDMGTLD SWVLSKAKKEDIDEDIKYDG
KWEVEDMKDGKLPDGKGLVLKSRAXHHAISAQLLRPFIFDTKPLIVQYEVNFQOQGIDCGG
AYVKLLSQTPDLNLDEFVDKTPYTIMFGPDKCGEDYKLHFI FRHKNPKTGEYEEKHAKKP
DADLRYYTDDKTHLYTLVVNPDNSFEVLVDQAVVNSGNLLTDMTPPINPAAEIEDPDDH
KPEDWDERPKIQDPDAVKPEDWDEDA PKQIPDEDAVKPDGWLDDSEYTS DPAVKPEDW
DEDMDGEWEAPQVPNALCETAPGCGAWQRP MIDNPSYKGKWKAPMIDNPNYQGVWKPRKI
ANPAFFEDLHPFRMTPFNVAWGLELWSMSSDIFFDNFFITNERHTADRWANDGWGLKKAEE
GALEPGLVNQMMTAADERPWLWVVYVLTVA VPLVLIIVFCCTGKKTAAADYKKTDEPQPD
VKEEEVVEKAEADQVKEEKSQPAAEKNSDAEDSPAEEKEEVNEEEEEDEEEEEDEVTEEVRE

Q

>SalmonB2m1 XP_014044855.1 NW_012376632.1:2.680-3.763 3,8 kb unplaced

MKSILSIVVLVLIYSAVESKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTWEPDM

>Salmonb2m2 NC_027306.1:57788614-57788364

YSRNPGEHGKNTLICHVSGFHPPDISIQLLKNGVEIPDAKQTDLA FEQGWFHLTKSVG
FTPDSGEEYTCRVRHLK NLKTYTW

>Salmonb2m3 NP_001117171.1 NC_027306.1: 57.826.485-57.828.480 ssa07

MKSILSIVVLVLIYSAVESKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW EADM

>Salmonb2m4 XP_014064343.1 NC_027306.1: 58.302.043-58.303.127 ssa07

MKSILSIVVLVLIYSAVESKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW

>Salmonb2m5 NW_012366138.1: 15-1.099 XP_014044161.1 unplaced

MKSILSIVVLVLIYSAVESKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW

>Salmonb2m6 NW_012360375.1:3.156-4.240 XP_014043232.1 unplaced

MKSILSIVVLVLIYSAVESKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW EADM

>Salmonb2m7 NW_012349147.1: 7.171-7.154 XP_014039773.1 unplaced

MKTVLSVIAFCVFLGFINAKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPARGEEYTCRVRHLK NLKTYTWESNM

>Salmonb2m8 NW_012349147.1:10.766-11.950 XP_014039774.1 unplaced

MKTVLSVIAFCVFLGFINAKESPPKVQVYSRNPNGFDKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW EADM

>Salmonb2m9 NW_012394841.1:962-1.615 XP_014045653.1 unplaced

MKTVLSVIAFCVVLVTINAKESPPKVQVYSRNPGEHGKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW

>Salmonb2m10 NW_012394112.1: 494-1.164 XP_014045626.1 unplaced

MKTVLSVIAFCVVLVTINAKESPPKVQVYSRNPGEHGKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTW GELLV

>Salmonb2m11 NW_012347820.1:16.383-18.779 NP_001232842.1

MKTVLSVIAFCVFLGFINAKESPPKVQVYSRNPGEHGKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NLKTYTWESNM

>Salmonb2m12 NC_027316.1: 57.631.152-57.632.033 XP_014009395.1 ssa17

MKTVLSVIAFCVFLGFINAKESPPKVQVYSRNPGEHGKNTLICHVSGFHPPDISIQLLK
NGVEIPDAKQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHLK NL

>Salmonb2m13 NW_012571237.1:861-1.007 Pseudogene

KQTDLA FEQGWFHLTKSVGF TPD SGEEYTCRVRHMK NLKTYTW GELLV

>SalmonCALR1a ssa14 NC_027313.1:15.482.204-15.485.137 XP_013995146.1

MLVSVLLMIALASAKPSVYFREQFEDDAWNTRWVSSHRSYDGKFVLTAGKFYGD AEKDK
GLQTSQDAHFYSSARFEPFSNQGKTLVIQFTVKHEQNIDCGGGYIKLFPADLDQADMHG
DSNYNIMFGPDICGPATKKIHVIINYKGKNHLIRKDIRCKDDEYTHLYTLILNPDNTYEV
KIDNKKVESGSLEEDWDILPPKKVKDPEAVKPDHWDERERMEDPDDKKPEDWDRPENIAD
PDAKQPEDWDDMDGEWEPPMVSNPDYKGEWKPR TIDNPDYK GKWLHP EIDNPDYSADSE
IYRFDSIGVIGLDLWQVKS GTIFDNFLITDDATLAEVGNETWGQTKDPEKKMKESQEEK
ERKKLEAEEMARKEETKDEPGEEEEEEEEEELEHEEEEEDEEEGETGAQEEEEESDSIKDE
L

>SalmonCALR1b ssa03 NC_027302.1:16.022.537-16.025.752 XP_014045728.1

MRGSMFLFSALIALASAEPSTLYFKEQFEDGDAWTTTRWVSSHRSYDGKFLVTPGKFYGDPE
 KDKGLQTSQDARFYSSARFEPFSNQGKTLVIQFTVKHEQNIDCGGGYIKLFPADLDQAD
 MHGDSKYNIMFGPDICGPGTKKVHVIIINYKGNHLLISKDVRCDDDEYTHLYTLILNPDNT
 YEVIKIDNKKVESGSLEEDWDILPPKKVKDPEAVKPDWDERERVEDPDDKKPEDWDRPEN
 IADPDAKKPEDWDNEMDGEWEPPMVSNPDYKGEWKPRKIDNPYKKGWVHPEIDNPEYSA
 DSEIYRFDSIGVIGLDLWQVKSITFDNFLITDDATLAEVGNETWGTQKDPEKKMKVSQ
 EEQERKKLEAEEMGRKEQTKDEPEEEEEEEEEKEEELEDREEEEGTGAEEEEEETDSIK
 DEL

>SalmonCALR1.2 ssa23 : NC_027322.1:26.361.479-26.364.218 XP_014024943.1
 MTTMLILLMTVLVASIFGESSVYFREEFEDGDAWKSRLVSKHRSYDGKFLVHTAGKFYGD
 VERSKGLQTSQDARFYSSARFESISNKDQTLVIQFTVKHEQNIDCGGGYIKLFPADLNQ
 EEMHGDSSTYNIMFGPDICGPGTKKVHVIFINYKGNHLLINKDIRCKDDEYTHLYTLIVNPD
 NTYEVKIDNKKVESGSLEEDWDFLPLKKIKDPDAEKPDWDEKENIDDPEDKKPEDWDVA
 ENIPDPDAKKPDWDEMDGEWEPPMVNPDYKGEWKPKQIDNPAYKKGWVHPEIDNPEY
 TADPEIYQYASIGVIGLDLWQVKSITFDNFLITNDPKLAEVGNETWGTQKDPEKKMKD
 RLEEEERKKREAEENVKKEEDEDEDEDKEERDEEEDYDDEEEEEETDSKLDDEL

>SalmonCALRL1 ssa14 NC_027313.1:41.509.871-41.518.596 XP_013996614.1
 MQVLGSFAIILSIFSVHSTVYFQEQFLDGDWAKTRWLDKHKADYGEWKLTAGNFYGD
 KDKGLQTSQDARFYAASARFEPFSNEGKTLVIQFTVKHEQKIDCGGGYKVFVPSLTDQAD
 MHGDSQYYIMFGPDICGYSTKKVHVIFINYKGNHLLIKKEVKCKDDELTHLYTLILNPNQT
 YEVIKIDNEKVESGTLLEEDWDFLPAKTIKDPEAKKPDWDDRPKMDDAEDAKPEDWDVAEN
 IPDPDAKKPDWDEMDGEWEPPVITNPEYKGEWKPKQIDNPYKGAWIHPEIDNPEYAA
 DSTIYKFDDISVLGLDLWQVKSITFDNFLVSDDVKEAEKFGAETWGTKEPEKKMKQEE
 DDKKRKEEDEKNKEQATEAEEEGEEEGEDEGEDEEETPEEGTEEEEAAPGKDEL

>SalmonCALRL2a ssa10 NC_027309.1:16.605.321-16.609.148 XP_014071687.1
 MRVAVAILAVFASVAVTIDATVYFKEQFQ
 DGDWAKSRWLVSSEHKSYDGWKLTAGKFYGDADKGLQTSQDARFYALSSRFEPFSNEG
 KSLVVQFTVKHEQKIDCGGGYKVFVPSLTDQADMHGDSQYYIMFGPDICGYSTKKVHVIF
 NYKGNHLLIKKEIKCKDDELTHLYTLILNPDQTYEVKINNEKVESGTLLEDDWDILPPKTV
 KDPEAKKPEDWDDRKIDDPDTDKPEDWEKPENIPDPDAKIPDDWDVMDGEWEPPMIPN
 PEYQGEWKAKQIDNPEYKGAWVHPEIDNPEYTADASIYKFDNIGVLGLDLWQVKSITFD
 NFLIGDDIKEAEFGNETWGTQKDPEKKMKDAQEEERKAREEEKSKDTADDEGEDEDED
 EPEEEDDDSPTEEEGEDPKKDKDEL

>SalmonCALRL2b ssa16 NC_027315.1:40.259.148-40.267.862 XP_014004651.1
 MRVAVAI FSVFASVAVTIDATVYFKEQFQDGDWAKSRWLVSCHKTDYGEWKLTAGKFYGD
 AEADKGLQTSQDARFYAMSSRFEPFSNEGKPLVVQFTVKHEQKIDCGGGYKVFVPSLTDQ
 AAMHGDSQYYIMFGPDICGYSTKKVHVIFINYKGNHLLIKKEIKCKDDELTHLYTLILNPD
 QTYEVKINNEKVESGTLLEDDWDILPAKTIKDPEAKKPEDWDDRPKIDDPDTDKPEGWEK
 ENIPDPDAKKPDWDDVMDGEWEPPVIPNPEYQGEWKPKQIDNPYKGTWVHPEIDNPEY
 TADTSIYKFDNIGVLGLDLWQVKSITFDNFLIGDDVKEAEFGNETWGTTEKEPEKKMKD
 AQEEERKAREEEKSKKDTADDEGEDEDEEDESKEEEDSPTEEGEEIIPMKDKDEL

>SalmonERp57a (PDIA3a) ssa23 NC_027322.1:41.701.278-41.708.848
 XP_014025621.1
 MLKLFFFVVLAGAALASDVIEFTDDDFDSKIGDHGMILVEFFAPWCGHCKKLAPEYEVAA
 TRLKGIVGLAKVDCTVHNVCQKYGVSGYPTLKIIFRDGEDAGPYDGPRTADGIVSHLKKQ
 AGPASVELKTEADFTKYVGDRDASVVGFFADGSPAKAEFLKSASALRESFRFAHTNSEE
 LLQKHSVEGEGIIILFRPSRLNNKFEEGSVKFSEDTFTNAKIKQFIQDNIFGMCPHMTDDN
 KDQMKGKDLLVAYYDVVYEKNPKGSNYWRNRVMKVAKGFLDQGNKLNFAVASKNSFSQDI
 AEMGLDASSGELPVVGIRTAKGDKYVMTEEFSSRDGKALERFLQDYFDGKLKRYLKSEPIP
 ENNDGPVKTVAENFDAIVNEEDKDVLEFYAPWCGHCKSLPKWKELGEKLSSDPNIVI
 AKMDATANDVPSQYEVGRGFTIFFAPAGQKMSPKKYEGGREVSDFISYLKKEATNPLVAQ
 EEETSKKKKKNEL

>SalmonERp57b (PDIA3b) ssa10 NC_027309.1:65.889.946-65.898.140
 NP_001161991.1
 MLKLFFFIVLAGAARASDVIEFSDDDFDSKIGDHGMILVEFFAPWCGHCKRLAPEFEVAA
 TRLKGIVALAKVDCTVQNNVCQKYGVSGYPTLKIIFKDGEDAGAYDGPRTADGIVSHLKKQ
 AGPSSIELKTEADFTKYVGDRDASVVGFFADGSPAKAEFLKSASALRESFRFAHTNSGE
 LLQKNGVEGEGIIILFRPARLSNKFEESVIKFSEDKFTNAMIKKFIQDNIFGMCPHMTDDN
 KDQMKDKDLLVAYYDVVYEKNPKGSNYWRNRVMKVAKSFLDQGKTLNFAVASKNSFSHDI
 SEMGLDASSGELPVVGIRTAKGDKYVMAEEFSSRDGKALERFLQDYFDGKLKRYLKSEPS
 ENNDGPVKTVAENFDAIVNNEEKDVLEFYAPWCGHCKSLPKWKELGEKLSSDPNIVI

AKMDATANDVPSQYEVGRFPTIFFAPAGQKMSPKKYEGAREVSDFISYLKREATNPLVAQ
 EEETSKKNIQIEL
 >SalmonERp57L1 ssa02 NC_027301.1:21.616.513-21.621.161 XP_014016726.1
 MGTGLGPFMRFLFALAQNVFVAASDVLELGDSDFFHYTVAEYETVLVEFFAPWCGHCQQL
 APEYETAATKLKGTVSLAKVDCTVNSETCGRFGVNGYPTLKIIFRNGEDFAAYDGPRADG
 IVSYMKKQAGPSSVPLHNGRDLDAFVNNFDASVVGFFSGVDSSQMAEFLKASSAMRDSHR
 FAHTTDLGLKKGHVESDVTVLFRPPRLNSKFEDSLVKSDEAVSTASLRQFIRDNVFGLC
 PHLTAENRENMRGRDLLVAYYDVLRLNIKGTNYWRNRVMKVATQFQSRGLSYAVANRAE
 FQEELEEEFGLGPSDGGELPLITIRNREGHKYSMQEEFTRDGKSLERFLEDYFAGKLRQ
 VKSEAASENNDBGPKVVVADNFEEIVNNPSKDVLEFYAPWCGHCKSLEPKYTELGELQS
 ADTHIVIAKMDATANDVPPYTDVQGFPTIFFVPAGQKDQPKRYEGGREVNDFLNYLKEEA
 THPLVLGTAREDL
 >SalmonERp57L2a ssa14 NC_027313.1:61.089.980-61.110.809 XP_013997083.1
 MASFLSLIPPFLSVLIFSGAAVARGDVLELGDADFDYLAEEHETMLVKFYAPWCGHCKK
 LAPDFETAATRLKGTVPLAKVDCTANPDTCGRFGVTGYPTLKIIFRNGEDASSYDGPRAD
 GIVHFMKKQAGPNSVTLRREADLEAFVNHFDASVVGFFSGPDGQLAEFLKAASVMREHF
 RFAHTIDMTLGLKHGVDTERVLLFRPPRLSSKFEEVSLHFTETITHTLRRFIRDNIIFGM
 CPHLTNENRDKLKGQDLLTAYYDLDYLQNPKGSNYWRNRVMKVGSQFASQGLSFAVANRR
 DFVDELEEEFGLGASDGGDLPFVTIRTRQGFKYTMREEFTRDGKSLERFLVDYFAGRLKR
 YIKSEPIPEKNKGPKVVVAESFEEIVNDPEKDVLEFYAPWCGHCKSLEPKYKELAEQL
 YSDPNIVIAKMDATANDVPQGFVDVQGFPTIYFAQASKKDQPKRYEGAHEVKDFIKYLRKRE
 ASHVPVVSQVREDL
 >SalmonERp57L2b ssa27 NC_027326.1:12.040.425-12.049.378 XP_014032674.1
 MASFLSLIPAFTLSVIFCGAVVARGDVLELGDADFDYLAEEHETMLVKFYAPWCGHCKK
 LAPDFETAASRLKGTVPLAKVDCTASPDTCGRFGVTGYPTLKIIFRNGEDSSSYDGPRAD
 GIVHYMKKQAGPNSVTLRSEADVEAFVNHFDASVVGFFSGPDQAFLAEFLKAAGAMRDHF
 RFAHTIDMTLGLKHGVDTERVLLFRPPRLSGKFEEVSLRFTETITHTLRRFIRDNIIFGM
 CPHLTNENRDKLKGQDLLTAYYDLDYLQNPKGSNYWRNRVMKVGYQFASQGLSFAVANRR
 DFVDELEEEFGLGASDGGDLPFVTIRTRQGFKYTMREEFTRDGKSLERFLEDYFAGRLKR
 YIKSEPIPEKNKGPKVVVAESFEEIVNDPEKDVLEFYAPWCGHCKSLEPKYKELAEQL
 YSDPNIVIAKMDATANDVPQGFVDVQGFPTIYFARADKKDQPKRYEGAREVKDFIKYLRKRE
 ASHIPVVSQVREDL
 >SalmonPDIA1 ssa02 NC_027301.1:45.673.995-45.687.434 XP_014032240.1
 MFKFLLCTLAVASRADIGEEDVLVLKKSNFEEALKAHPNILVEFYAPWCGHCKALVPE
 YAKAASMLKAEGSEIRLAKVDATEEADLAQEYGVRGYPTIKFFKGGDKESPKEYSAGRQA
 DDIVNWLKRTGPAATTLGEVAQAESMIAENEVAVIGFFKDAESEGAFLKAAEAVDDV
 PFGITSNDVFSKFEVSKDGVVLFKKFDEGRNTFDGELSKADLLAFIKANQLPLVIEFTE
 QTAPKIFGGEIKSHILMFVPAASDFNDKMAEFKKASEGFKGKILFIFIDSEVDDNQRI
 EFFGLKKEECPAIRLITLEDKTYRPESEAITADNIIAFCTLFTEGKLKPHLMSQDIPE
 DWDKNPVRVLVGKNFEEVVDPKKNVFVEFYAPWCGHCKQLDPIWTKLGEKYQDSADIVV
 AKMDSTANEIETVKVHSFPTLKFPPAGDEHKVVDYNGERTLEGFTKFLESQKDGKDGAPAG
 EGEEDDEGIDMEDLDEQSDSDSGDGGDHDEL
 >SalmonPDIA2.L1 XP_014038534.1 NW_012347234.1:27.515-33.804
 MRFCVVLAAVVLVLRVSWTQTAEDTSPEQTDAAEQETDLEKKEKTTEIEEEKNMVVLHIN
 NFQRLASENKFLLEFYAPWCGHCRQLEPVYAEAARVLKGEREEEREGEEGFGGLAKVDAV
 EENQLAEEFDVGSFPTIKLFTDGRNPNVDFTGKRTVQGIVQWMKRRSGPVAVALETTDA
 AAHINLHNVTVLGFFTSLESEEAKVFYSVAMEMVDMFEGVTTSPPEVFQKYEIENNRVVL
 FKKFDEGRVDSLVSSEEVKVGEEELTVFIRTNSLELVIEFNEQNADKIFGSKIHSLSLFI
 NSTVQEQKNLLPEYRTAAKDFKGVLFIIIDVTGPVSHVLKYFGLSEGDAVRIINTDT
 TKKFALIGQITAATLQTFQGVLDGNVKSLLSEVPEDWDKGPVKVLVGKNFEAVALEN
 NKNVFVEFYAPWCGHCKELAPVWEKLAEKYADRDIIIAKMDATTNEVEGVSVSGFPTLR
 YYPAGEDSKVVEYSGRDLETfamFLDNGGQLPKAEEEEDDGDEEEVKDDEEDVKDDEEA
 VKDDEEDDEEEVTDESSPPPANETSKDEL
 >SalmonPDIA2.L2 XP_014042341.1 NW_012355960.1:107-7.487
 SLESEEAKVFYSVAMEMVDMFEGVTTSPPEVFQKYEIENNRVVLFKKNADKIFGSKIHS
 LLFINSTVQEQKNLLPEYRTAAKDFKGVLFIIIDVTGPVSHVLKYFGLSEGDAVRIINTDT
 TKKFALIGQITAATLQTFQGVLDGNVKSLLSEVPEDWDKGPVKVLVGKNFEAVALEN
 ALENNKNVFVEFYAPWCGHCKELAPVWEKLAEKYADRDIIIAKMDATTNEVEGVSVSGF
 PTLRYPAGEDSKVVEYSGRDLETfamFLDNGGQLPKAEEEEDDGDEEEVKDDEEDVKD
 DEEDDEEEVKDDEEDVKDDEEAVKDDEEDDEEEVTDESSPPPANETSKDEL
 >SalmonTAPBPR ssa02 NC_027301.1:10.225.492-10.231.249 NP_001133983.1

MLEILLFGYLITCVSGQGADVVLSCSLVEEGSGMGGMGGGALFSRTPAMLVLRDLAVTP
 DLSPDTLTPFNPPAVPDPDNIILEAKVESPEIPEADLLHADCNEQEVTCETISRYFPRNA
 KEGSTEPAYFIGSLQIEGGGLSLTLILQTLPLDQSDR PALMQSKLELPLSQSGTLLTEVV
 FLVFSRSESHSAPLGGVALLDCGFRQQAPPPGWELGLEWRLQHRGSGRKVLEIRAGQTET
 EEGPAVHVEREGSSVDAALLVGQGNASLTARLKVSDGTYICTVSTGLYQAQQVIQLHV
 TQPPRVSLSEEKLVFRDEL PQKLSCHCKNYYPLDVQMEWISVSSTDSEPSVLSDQVSLSS
 HRQHS DRMTSISSHLT LHPSTFPFGTTVTCTRVTHPALDTPLSLSLTVETPEPDSYWMVLG
 FLVITVLFYQVMK

>SalmonTAPBPL1a ssa09 NC_027308.1:102.783.830-102.789.947 XP_014069540.1
 MSLNVNILLCLFLCGEVPGIQSFEQVPWLPCLVDESVTFNDEGHAETEQYQHRDAGLQFG
 HPGDSALNPNTITFLVTGSKVDMRKYIEGVVEHQLQCEIRRYSTESPQMRWPGLGAQEHD
 IWFTCTIRHTDGVFIIITSFLRHTPATPTPGQAHYLNWAAIADRAALTSTVMVLVLRSPS
 VRVGLVKQQSLHCQFDVDHKAADLTVEWRFQRRGERTTLFSHSSRSGQTEGSGVPLNAIG
 RGDASLTPLTKQSSEGTYVCHVSVPPPLFGSHDITLQIMESPRVSLNVDSTISLVDRGEQ
 KVVCEAEGYYPLDVEMEWFREPSGGGLLPEKLDTVLYSSHRHHQDGTYSLSAFFLLHASL
 HDSGSKYFCRVSHSSQRMPIRKSF T LNVTEYDSWNAALWFFSGFGFILVMVATL FVMLPR
 LSSARKANQRKPY

>SalmonTAPBPL1b ssa20 NC_027319.1:46.797.073-46.803.184 XP_014017660.1
 MSLNVNIVICLFLCAEVPGIQSLEQVPWLPCLVDERVKFNDEGHAETQYQHRNAGLQFG
 HPGDSALNPNAITFLVTGSKVDMRKYIEGAVVEHQLQCEIRRYSTEGIQVRWPGLGAQDH
 DIWFTCTLRHTDGLFVITSFLRHTPATPTPGQADYRNWAAIADREMLTSTVMVLVTRTP
 SVWVGLMKQSSLHCQFDVDHKAADLTVEWRLQRRSERTTLFSHSSRSGQREGGGVELKGI
 GRGDASLTPLTKQSSEGTYVCLVSVPPPLFGSHEIALHIMEPPRVSLNVD SVISLV DTER
 QKMVCEAVGYYP LDVEMEWFREPSGGGGGLLPEKLDTVLYSSHRHHQDGTYSLSAFFLLH
 ASLHDSGSKYFCRVSHSSRLMPIRKSF T LIVTEVDSWNFLWLFFGCGFILVMVATLCVML
 PRLSSARKANKRKP Y

>SalmonTAPBPL2 ssa07 NC_027306.1:12.926.171-12.930.490 XP_014062182.1
 MKTFTVLLCLIPYAGVLGFLQVQWLRCLRKDEYVWNTNDEGHIETNNSYRDAVLQFENSGN
 SALLSESITFLVAASKVDMRKVFVEGPVDQLQCDIHRYSKGMSRVRWPSLGGTGHDWFTC
 TLRHTAGLFNITSFLRVTPATTMSAHQPDFLSWLTIGVKEQISASAVMVLMTSPSVRVG
 LQEERTLHCQFAVDHKVPHLTVEWHLQRHGARTKLFSSSSSGQTEGSGVAVKGIAGGD
 ASLTVPITKVSSEGTYVCSVRVPPLNGSVDIALHII EPPNVSLTTKEEQIPRVVCEANGF
 YPLDVDIDLKFTSSGQRPEKLG NVLHSSHIQHHDGTHSVA AFVRLRPSQD SGCE TYSC
 RVSHVSLQGLYIHKSIIGPGCWTWIHLTLPGLVCLIFVIFIAVTRQRF F

>SalmonERAP1 XP_013998888.1 ssa15:2.525.189-2.544.613
 MRTLTVTILVLLHVSFAPSLAAQLPGDHDTNDKSSSLPPIATNGQFPFWIHMRLPETVS
 PIHYDLLVHPNLTS L DFTGEVQIQ LQVFEDTSTIILH SKDLQIAKAELLAPEGPGSLPVP
 LQVLEYPAFHQLALMSDVLLVRGGMYKVRLEFSANLSDSFHGFYKSSYRTTKGEVRFMAS
 TQFEATSARAAFP CFDEPAFKANFTIQIRRESRHIALSNMPKVKTVELPGGVLEDHFDTS
 VRMSTYLVAFIVSDFQSVSKTTS HGVKISVYAVPDKINQTD FALNAAVRLLD FYDDYFDI
 PYPLPKQDLAAIPDFQSGAMENWGLTTYREAGLLFDPNKSSASDKLGITMVIAHELAHQW
 FGNLVTMQWVNDLWLN EGFAKFMEFVSVNITNP ELQVNDFFLGKCFEAMEVDSLSSSHPV
 SSQVDTP TQIQEMFDDVSYDKGACILNMLRDFLTPEAFKIGIVRYLRRYSYQNTVNSHLW
 ESLTNICQSDDLDEGR LKDEGFCSQEKAQSGAPKWYSGDQLDVRAIMDTWTLQEGFPLVT
 VEVGRQVRLSQERYLKTDDPSQTHGFLWQVPLTYITSSSSTVHRFLKTRTDVLYLPEE
 VGWIKFNVDMSGYYMVHYEGEGWRSLTSLLLTNHRALSSNDRASLINNAFQLVSVGKVGL
 DTALDLSLYLSKETDIMPVTQGLGELVPLYKLMEKRDMEGLENQMKGYIVELFRGLIDRQ
 TWSDDGSVSQVRVLSYLLLF GSVRNHPPCVATATHLFNKWRASDGNMSLPSDVSLAVFAI
 GARDPEGWDFLFEKYRHSQHTSVKSRIKSALSISPLQHKLKRLMEQSLAGEVMKTQDLPH
 VVISVSRNP KGYKLAWDFLRHNWHTLVKKFDLGSHSISGMVTGV TNQYSTREMLDEVGRG
 FDSLSEETGSGRLCIQQTYESIEENIRWMDQHLPLKAWLDRQAQGARTETQGHEDL

>SalmonERAP2a XP_014026228.1 ssa24:15.056.418-15.064.045
 MVRFLVLALLSLAGVTQTSASPTQASEPPNPTEEQPPLGTGSLSPFWSHRLRPGYIVPLH
 YHLLHPNLTMLSYSGTVRIELQVQNNTNWVVLHSGKLRITATMLDQNLALHSDRVLPV
 LHNPTHEQTAFSPRVLSGGQKYFLFLEFGAELGEGFYGFYRSTYRTSAGETRNLASTHF
 EPTSARMAFP CFDEPSFKANYSISIRSLAHTALS NMPVEQTVVLD DGLMEDRFAVSVRM
 SSYLVAFIVCDFRSVSATTASGVKVSVYAAPEKWQQTHYALKA AVKLL E FYE KYFNIKYP
 LPKQDLVAIPDFQAGAMENWGLITFRETSLLYDPATSSASDRVWVTMVIAHELAHQWFGN
 LVTMEWVNDIWLNEGFA RYMEYISVNATYPKLRVEDYLVDT CFAAIGRDSL NSSRP ISSA
 AESPTQIAEMFDTVSYDKGACVLHMLRH YLTDQVFQSGIVRYLRRYSYSNAHNQDLWDSL
 ANTCSEEEFTSGEHCYSSRQA AKNAYLYAGEHLDLTMMNTWMLQTVPLVTVTRQGSRL

LLKQERFLRTAHPSDPAWPSLQQGFLWHIPLTYRTDTSTSIHRHLMTTLTDSVDVGEEVG
WVKVNVDMAGYYLVHYDGSWDNLIQLLKDNHTALSFMDRTHLIHNAFQLTTAGRLSLDK
ALDLIGYLRSESHTVPLLEGLGYLEAFYRMVERRDIPDVTQNLRTYILWYFRDVIDRQW
SDKGSVSEERRLRSELLSLACHLGLDLPCLKQAQRSFTHWLDSNSTLNLPAOVAETVYSVGA
QEDTGWASLLQTYTHSLSETHKRKILSALASSRDTNKLTRLLELGEVEGIVRTQDLDSLI
VMVARNPRGHHLAWSYVQKYWSTLVDFKQLGSFSIRNIIIGTTAQFSSTEELTEVRVFFE
SIEHQASQLRVTEVAMDNVQKNILWLQRLNLGTLRLSWLDQQID
>SalmonERAP2b XP_014016590.1 ssa20:28.634.401-28.645.787
MFWRFLFLALLSLAGVTQTSSSPTQASEPPNPTEEQPPPLNTGSLSPWWSHLRLPEYIVP
LHYHLLLPNLITILSYIGTVRIELQVQNNTNWVVLHSGKLRITTATVLDQNLHLSDQVL
PVLHNPTEHQVAIFSPRALTTGGQKYFLFLEFGADLGEFGYGFYRSTYRTSTGETRRLAST
HFEPTSARMAFPCFDEPSIKANYSISIRSPAHTALSNNMPVEQTEVLDDGLMEDRFAVSV
RMSSYLVAFIVCDFRSVSATTASGVKVSVYAAPEKWQQTHYALKAALKLEFYEKYFNIS
YPLPKQDLVAIPDFQSGAMENWGLITFRETSLLYDPTTSSASDRLWVTKVIAHELAHQWF
GNLVTMEWNNDIWLNEGAFATYMEYISVDTTYPKLRVEDYLLDTCFVAIGRDSLNSRPIS
SVAESPTQIKEMFDTVSYNKGACVLHMLRHYLTDQVFQSGIMRYLRRYSYSNARNQDLWD
SLANTCPEEEFTSGGHCHYNSQAANKAYLYAGEHLDLTMMNTWTLOTGVPLVTVARQGS
RLVLKQERFLRTTHPSDPAWPSLQQGYLWHIPLTYRTDTSTSIHRHLMTTLTDSVEVGEE
VGWVKVNVDMAGYYLVHYDGSWDNLIQLLKNNHTALSFMDRTHLIHNAFQLTTAGRLSL
DKALDLIGYLRSESHTVPLLLQGLAYLEAFYRMVERMDIPDVTQNLSTYILWYFRGVDRQ
TWSKGSVSEERRLRSELLSLACHLGLDLPCLQQAQRSFTHWLDSNSTLSLPADVTTETVFSV
GAQEDSGWASLLHIYTLSELSETHKHKILSALASSRDTNKLHRLLELGEVEGIVRTQDLDS
LIVMVARNPRGHHLAWSYVQKYWSTLVDFKQLGSFSIRNIIIGTTGQFFSTEELTEVRVF
FESIHEQASQLRVTVAMDNVQKNILWMLQRLNLGTLRLSWLNQHD
>SalmonPSME1a ssa29 NC_027328.1: 20.102.332-20.105.959 XP_014036554.1
MTSIDIRPESKKMVDDFCTRLTKEAETLVTSFFPQKIAEMEMLLKKSFTDGLAALKSPL
DIPMPDPAKEEAKRKKKEEKEAKEGKKEKDSKEDEADAGPACGPIPCNERVESLLKEIKP
QIQILKEKLNTVSMWVQLQIPKIEDGNNFGVAVQEKVFELLTNTRTKIEGFQTQISKYYS
ERGDVADKASKEPHVGDYRQLVHELDQYQYCELRIVVLEIRNTYAVLLDIINKNYDKIKK
PRGDCKALIY
>SalmonPSME1b ssa19 NC_027318.1: 74.398.814-74.401.718 XP_014015182.1
MTSIDIRPESKKQVDDFCTQLTKEAETLVTSFFPQKIAEMDMLLKTSLSTEGLAALKAPL
DIPIDPAKEEAKQKKKEEKEAKEGKKDKDSEKEDEADAGPPCGPIPCNERVESLLKEIKP
QIQLLKEKLNTVSMWVQLQIPKIEDGNNFGVAVQEKVFELLTNTRTKIEEFQTQISKYYS
ERGDVAKASKQPHVGDYRQLVHELDQYQYCELRIVVLEIRNTYALLFDIINKNYDKIKK
PRGDCKALIY
>SalmonPSME2a ssa29 NP_001135126.1 NC_027328.1: 17.457.855-17.464.336
MSRSSVLKIKSANAVKVENFRQSLYQQAEDLFSNYIPLKITQLDNLKEEDLSITDLSTL
HAPLDIPIPDPTPEDEEMETDKNDDDEKKKKAPKCGFIKNEKIVKLLDRVKPEILALR
ETIITVSCWIQHLPKIEDGNDFGVAIQEKILERIVAVKTKVDGFHTNINKYFSERGDAV
SKASKLTHVMDYRSLVHEKDEAVYSDIRVILLDIRGFYAEYDIISKNLKVTNPKGEEK
PSMY
>SalmonPSME2b ssa19 NC_027318.1: 77.929.964-77.936.657 XP_014015323.1
MSKSSVMKIKSVNAVKVEHFHSLYQQADDLFSNYIPLKISQLDNLKEDDLNIPDLSTL
QAPLDIPIPDPTAEDEEMETDKNDDKKKKKAPSCGLIKNEKIVKLLDRVKPEILSLR
ETIITVSCWIQHLPKIEDGNDFGVAIQEKILERIAAVKTKVDGFHTNINKYFSERGDAV
AKASKSTHVMDYRSLVHEKDEAVYSDIRVILLDIRGFYVELYDIISKNLKVTNPKGEEK
PSMY

Rainbow trout (*Oncorhynchus mykiss*) MHC Ia region Haplotype Ia_#A; BAC (AB162342) deduced amino acid sequences

BAC sequence reference: see Shiina et al.[4].

>TroutUBA*0501_#A BAD89550.1 AB162342.1:52.087-74.224
MKGFIILLLLGIGLLHTASAATHSLRYVYTATSGVPDFPEFVSLGIVDGMQIDYYDSNTRK
VVPKQDWMKTEGSDYWERQTQNSIGDEQTFKANIDVAKQRFNQTGGVHTFQLMYGCELG
DDGITRQDFQLGYDGADFLSLDKSTLTWTAANQKAVITKLKWDATGAEANFQKNYLENTC
IEWLKKYVNYGKDTLERKVRPSVSLQKTPSPVPTCHATGFYPSGVMVFWQKDGQEQHGD
VEHGEILQNDGTFQKSTHLTVTPPEEWKNNKYQCVVQLAGIEDDITKVLIESEIQTNFGK

```

TNRGSNDPITIGLIIGGVIALLVIIIVGVVWKKKNNKKGFPASTSDTDSSENSGKGIQKI
>TroutPSMB8a_#A pseudogene AB162342.1:37.694-38.613
LYKLRRNRQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDNATRLSGR
MFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYSEGVVNLYHMQEDGW
IKVKCEDVSELIHR*
>TroutPSMB13a_#A BAD89549.1 AB162342.1:33.054-36.174
MALSNVETPASGFNFENVCRNVALEGLLQGGQTKAPKPMKTGTTIAGVLCCKNGVVLGAD
TRATSGEVVADKMKCAKIHYSIPNIYCCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
MAVNILQDMLFRYRQIGASLILGGVDCTGNHLYKVGPYGSIDNVPYLAMGSGDLAALGI
LEDRFKANMEMEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQESEYKDK
RQRRYKYRPGTTSILTEKIVPLELEVQETVQRMDTA
>TroutPSMB12a_#A BAD89548.1 AB162342.1:23.244-32.906
MEKRFMDSQIKGVSTGTTILAVTFNGGVIIIGSDSRASIGGYVSSKTINKLIQVHDRIFC
CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAASVLKELCYNNKEELQAGFITAGW
DRKKGPQVYTVALGGMLLSQPFTIGGSGSTYIYGADAKYKPDMSKEECLQFATNALALA
MGRDNVSGGVAHLVVITEEGVEHIVIPGDKLPKFHDE
>TroutPSMB9a_#A PSMB9da BAD89547.1 AB162342.1:19.534-21.449
MLDESLEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSSYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ
>TroutTAP2a_#A BAD89546.1 AB162342.1:12.473-18.254
MLRTCAFAVAVVLCVDITTFCAPGFGESISKTGPKTFDTFGNVVRLWVVS GIRLVLLLGL
SLLTLGSIKPVFKRWLAHVHCFAPVYETGRLILYGGSTESACGSLGGPSLWLLCTAAAAA
AALFWEKTFPDSNGESNGKEKTQKARVLFMRVLHFYRPDTLLLVGAFIFLSLAVLCMFIF
PFYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTSRMKVEL
FGALVKQDIGFFETIKTGDITSRLSTDITLMARAVALNVNVLRLTIKTVGMLSMLMMSLS
WKLTLMLMETPVTGLLQSVHDNYLRLSKVEQDSIARANEAAAGETVAGIRTVRSFKTEQ
HEAGRYNDRMLDTHNLKTRRDTVRAYVLLLRRLTALVMQVAMLYYGRLFIQRGQMSTGNL
VSFILYQSDLADNIRTLIYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTGHVQ
FHNLSFSYPTQRERKVLQGFSLERLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGEILL
DGLPLQSYQHHLHRKVAMVQGEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSFI
SHLEKGYDTDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQEAL
ASCPSQTLLVIAHRLKTIERADQIILIDQGTVEQGTHTQELMDRKGSSYYKLRERLFTEDD
MSHCPKQC

```

Rainbow trout (*Oncorhynchus mykiss*) MHC Ia region Haplotype Ia_#B; genome (GCA_002163495.1) Chr.18 (CM007952.1) deduced amino acid sequences

Unpublished rainbow trout genome assembly GCA_002163495.1.

```

>TroutTAPBP_a_#B CM007952.1:33.137.898-33.141.781 NP_001118025.1
MANISTIIYKLSFLAFTYFIQVYGTSCPVLCEWVQEKPGRGGFPAAMIQEKSLLYINTD
PESEQTKSQQGPSADISHDRVYYITDPAAILCSSLHPPEGSVHKPQCEINPFMPQPSTV
QWAVPLTDSAHSPYIYLQADWYSAALQGLDGLALSSVMRAPTATKEPTVVLVSVSRTPLV
RSRLGEPVVLDCGFWAETSSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTDRAETKEKGA
GLDFTALHETGNASLILQEA EVRHTGTIYICTVYLPYLLAQVAVELEIVEPPSLSIFPSPL
PLSVPGQVVKVQCEASGFFPLSLEFYWELTGPDKVRPLGQGSVTGHRQGPNTYSQTTR
LELDSAKLDLGRGGEVTCVAVHPGGTTRASVTNLNIGVNGPSIEDSMAMVAVALGLYGLI
KIVSWTFSSGSDDTNSQAKVK
>Trout_UBA*3301_#B CM007952.1:33.177.839-33.203.608 (New trout UBA allele)
XP_021427673.1
MNVKMNGFISLVLGIGLLHTASAATHSLKYFYTAVSGDIDFPEFTAVGLVDKGQFMYFDS
STKTAVPKTEWMKREGADYWRQTQGLIGAHQTFKVNITLTKDRFNQSKSTGVHVFQWY
GCEWDDEAGVTEGFDQYGYDGEDFIAFDLKTTKWIAPTPQAVITKHKWDSNTARNEQNKN
YYTQICIEWLKKYVDYGKSTLMRTVPSPVSLQLKAPSSPVTCHATGFYPRDVMVSWQKDG

```

QDHEDVEYGEILQNDGTFQKSSHLTVTPEDRKNSKYQCVVQVKGikedFIGVLPDQDA
 ANVVPIIVGVVALLLVVAVVVGVVWRKRSKKGFPASRTESDLENQRRSKDQTNVQRG
 IVKN
 >TroutPSMB8a_#B pseudogene CM007952.1:33.205.845-33.206.762 XP_021427676.1
 LYKLNRQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDNATRLSGR
 MFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYS GG VV NLYHMQEDGW
 IKVCKEDVSELIHRYRKGMF
 >TroutPSMB13a_#B CM007952.1:33.208.283-33.212.648 XP_021427674.1
 MALSNVETPASGFNFENVCRALEGLLQGGQTKAPKPMKTGTTIAGVLCCKNGVVLGADTR
 ATSGEVVADKMC AKIHYISPNIYCCGAGTAADTEKTTDLSSNLTIFSMNSGRNPRVMA
 VNILQDMLFRYRGQIGASLILGGVDCTGNHLYKVGPGYGSIDNPYLAMGSGDLAALGILE
 DRFKANMEMEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGV DYIRPYQESEYKDKRQ
 RRYKYRPGGTTSILTEKIVPLELEVQETVQRM DTA
 >TroutPSMB12a_#B CM007952.1:33.213.355-33.218.282 XP_021427675.1
 TTILAVTFNGGVIIGSDSRASIGGYVSSKTINKLIQVHDRIFCCIAGSLADAQAVT KAAK
 FQISFHSIQMESPLVKAAASVLKELCYNNKEELQAGFITAGWDRKKGPQVYTVALGGML
 LSQPFITIGSGSTYIYGADAKYKPDMSKEECLQFATNALALAMGRDNVSGGVAHLVVIT
 EEGVEHIVIPGDKLPKFHDE
 >TroutPSMB9a_#B omy18:33.220.076-33.221.979 XP_021426203.1
 MLDESLEPGWLSEEVKTGTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKL SLLHDKIY
 CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
 WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSSYVYGFVDAEYRKAMSKEDCQQFV VNSLSLA
 MSRDGSSGGVAYLV TIDEKGAEKCILGNELPTFYDQ
 >TroutTAP2a_#B CM007952.1:33.223.235-33.229.012 XP_021427678.1
 MLRTCAFAVTVGLCIDITTFCAPGFGESISK TGPI TFDTFGNVVR LWVVS GIRLVLLLGL
 SLGSFTLGSIKPVFKRWLA VHCFLAPVYETGR LILYGGSPESACGSLGGPSLWLLCTAAA
 AAAALFWETTFPDSNGESNGKEKTQKARVLFMRVLFHFRPD TLLLVGAFIFLSLAALCEM
 FIPFYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTSRMKV
 ELFGALVKQDIGFFETIKTGDITSRLSTDTTL MARAVALNVNVLRLTIKTVGMLS LMMS
 LSWKLTLLMLMETPVTGLLQSVHDNYYQRLSKEVQDSIARANEAAGETVAGIRT VRSFKT
 EQHEAGRYNDR LMDTHNLKTRRDTVRAYVLLLRLTALVMQVAMLYYGR LFIQRGQMSTG
 NLVSFILIYQSD LADNIRTLIYIFGDM LNSVGAAGKVFEYLDREPQVSTKGTLPETLTGH
 VQFHNLSFSYPT RQERKV LQGFSLELRPGQLTALVGPSGGGKSTCVSLLERFYQPQGEI
 LLDGLPLQSYQHYYLHKKVAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHS
 FISHLEKGYD TDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEH MVQE
 ALASCPSQTLLVIAHRLKTIERADQIILIDQGT VQE QGTHQELMDRKG SYYKLRLERLFTE
 DDMSH
 >TroutPSMB8F_#B CM007952.1:33.247.717-33.255.385 NP_001158688.1
 MALLDVCGISDWKEDGFNAERASVDKVNHFNF AAQTPELAVPVGVDPAEFLRPLVDSEE
 GVDGVKINLEHGTTTLAFKFQHGVMVAVD SRASAGSYVSTQMFKKVIEINPFL LGTMSG
 AADC VYWERVLAKECRIYKLNRKERISVSAASKLLANMVVN YRGMGLSMGTMICGWDKKG
 PGLYYVDDNGLRLCGNM FSTGSGNTYAYGVMDSGYRYDLSP EAYDLAQRAIFHATHRDA
 YSGGT VNVYHMR ETGWIKVSQEDVGDLYHRFSNEKK*
 >TroutPSMB10a_#B 18:33.608.302 -33.611.763 XP_021427701.1
 MLHNSRPPQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGT TIAGIVFKDGVILGA
 DTRTTDDMVVADKNCMKIHYIAPNISCCGAGVAADA EVTTQMMSSNVELHSLSTGRPPLV
 VTVTRQLKQMLFRYQGHIGSSLIVGGVDVTGAHLYSVYPHGSYDKLPFLT MGGSGAGAAI
 SIFEDRYRPNAQLEEAKKLVRDAIAAGIFCDLGSGSNVDLCVITHAGVQYLRGYDQPAQK
 GKSRGQYTYKPGTTAVLT KT V TPLPLDIVDESIQLMDTQ

Rainbow trout (*O. mykiss*) MHC Ib region Haplotype Ib_#A; BAC sequence (AB162343) deduced amino acid sequences

BAC sequence reference: see Shiina et al.[4].

```
>TroutUDA*0105_#A BAD89551.1
MKGFI L MV LGIVHLHEAFGVTHSLTHFYTASSDIPNFPEFV FVGMVDGVQMVHYDSNIQK
VVPEQDWMKQTDAEFCEKERENLFYSQQSFKA EVATLKQHFNQSGGVHILQYMYGCSRDD
ETEQT EFGFQLGYNGENFLEYDMKTLTWKSLKQQPNFMQDEWNSDTSRLLFWKNYFTQTC
```

IECLKKQVDNRKSTLMRTVPPSVSLLQKSPSSPVTCHATGFYPSGVMVFWQKDGQGHQEG
 VENGETLPNNDRTFQKSTHLTVTPEEWKNNKYQCVVQVTGIKEDFIKVLTESEIQTNWKD
 PAPNIVPIIGGVALLLVVVVVVVVVIWKKKKKKGFVPASTNDTDSAYSSEGLLKTSVL
 SALLKQLKTVTK
 >TroutUCA*0101_#A BAD89552.1
 MKGFMLMFMGIGHLYEAFVTHSLTHFYTASSEVYNFPEFMVVGMDGVQMVHYDSNSQK
 AVPEQDWMKQTDAEFEWEKERENFFYSQQSFKAEVATLKQHFNQSGGVHILQYMYGCSRDD
 ETEQTEGFGQLGYNGENFLEYDMKTLTWKSRKQQANFMQDEWNSDISRLGFWKIYFSQTC
 IECLKKQVDNGKRTLRRTPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQKDGQDHED
 VEHGEILFNDDGTFQKSTHLTMHEERKNKNKYQCVVQVTGIKEDFIKVLTESEIQTNWKD
 PAPNIPIIIGGMVAPFLVVVVVGVVVIWKKKSKKGFVPASTSDTSDNSGKGLLK
 >TroutUEA*0101_#A BAD89553.1
 MISFCLFLWIPSSSADPHSLKYLYTAVSGDTDFPEFTVVGLLDDQQFVHFGSNTKTLVN
 DAWMNKTGKYNDLYNEPLINQYEGFKNLITFAKKQFNQTQSKGVHTIQONLYGCEWNDE
 TELQDYFHGYDGEDFISLDMKTVRWITSVQQADTIKQKWDNSKDLHYLKWYFTKECI
 DTLKKYVDFASSVLKKIVPPSVSLLQKTPSSPVTCHATGFYPSGVMVSWQKDGQEQHEDV
 LHGEILLNGDGTQKSAHLTIDSEKWENNNTCVVEHKENIIAIRLNQSVIKTNSVKPSE
 YVPIIIGLVVAVILLVLVPIITFIMWNRKSTASDDGSNPSMTQQNQIQLEVSLSLIKQE
 EAEKRGPLLKDSEAS
 >TroutPSMB8b_#A BAD89554.1
 MALFDVSGYKSHAGLRGQILGTGVGHLVDRPNQEFAPVGVDPSPGFLKSCSREGGVSIDL
 NHGTTTTLAFKFRHGIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
 LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDN
 ATRLSGRMFSTGCGSSYAYGVMDSGYREDMTVEEAYELGRRGITHATHRDAYS GG VVNLY
 HMQEDGWIKVCKEDVSELIHRYRKGMF
 >TroutPSMB13b_#A BAD89555.1
 MALSNVETPASGFNFENVS RNVALEGLLEGHTKAPKPMKTGTTIAGIVCKDGVVLGAD
 TRATSGEVVADKMCACIHYISPNNMYCCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
 MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
 LEDRFKPNMELEEAKQLVRDAIHSGIMSDLGSGNNIDICVITKQGV DYIRPYQESEYKDK
 RQRRYKYRPGTTSILTEKIVPLELEVQETVQRMDTA
 >TroutPSMB12b_#A BAD89556.1
 MDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGNYVSSKTINKLIQVHDRIFCCIAGS
 LADAQAVTKAAKFQISFHSIQMESPLLVKAAASVLKELCYNNKEELQAGFITAGWDRKKG
 PQVYTVALGGMLLSQPFTIGSGSTYIYGADAKYKPDMSREECLQFATNALALAMGRDN
 VSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE
 >TroutPSMB9b_#A BAD89557.1
 MLDESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKL SLLHDKIY
 CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
 WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSSYVYGFDVDAEYRKAMSKEDCQQFVVNTLSLA
 MSRDGSSGGVAYLV TIDEKGAEKCILGNELPTFYDQ
 >TroutTAP2b_#A BAD89558.1
 MLRTCVFAMAVGLCIDITTFCATGFGASIPKTGPISFDVFGSLLRLWVVGIRLVILFGV
 SLTLGSIKPVLRWLAVHCFLAPVYETGQLMLHGSSPESTHGS LGGPSLWLLCTAAAAA
 AALFWEKTFPDSSEESNGKEKTQKARVLFMRVLYFYRPDTLLLVGAFIFLSLAVLCMF I
 PFYTGKVIDILASQYKWNDFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTSRMKVEL
 FGALVKQEISFFETIKTGDITSRLSTDTTKMARAVALNVNVLRLTLIKIVGMLS LMMSLS
 WKLTLLMLMETPVTGLLQGVYDNYLRLTKEVQDSMARANEAAGETVAGIRTVRSFNTER
 SEASHYDHRLMDHTLKTTRDTRAVYVLLRLLTALVMQVAMLYYGR LFIQRGQMSTGNL
 VSFILYQSNLGTNIRTLIYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTGHVH
 FHNLSFSYPTQRERKVLQGFSLRLRPGLTALVGPSGGGKSTCVSLLERFYQPQQGEILL
 DGQPLHSYQHYYLHRKVAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSFI
 SQLEKGYD TDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEH MVQEAL
 TSCPSQTLLVIAHRLKTIERADQIILIDQGTVLEQGT HQELMDRKGGYKLRERLFTEDD
 TSH

>TroutUDA_#B CM007948.1:28.624.919-28.630.481 XP_021416252.1
MSLTHSLTHFYTASSDIPNFPEFVFGMVGDVGMVHYDSNIQKVVEQDWMKQTDAEFCE
KERENLFYSQQSFKAEVATLKQHFNQSGGVHILQYMYGCSRDETEQTEGFGQLGYNGEN
FLEYDMKTLTWKSLKQQPNFMQDEWNSDTSRLLFWKNYFTQTCIDCLKKQVDNRKSTLRR
TVPPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHEDVENGETLPNDGTFQKS
THLTVTPKEWKNNKYQCVVQVTGIKEVFIKVLTESEIQTNWKDPAPNIVLHIGGVVALLL
VVVVVVGVVIWKKKKKKGFVPASRSS

>TroutTAPBPb_#B CM007948.1:28.615.561-28.617.272 XP_021416251.1
MDVYGTSCPVLCEWFWQEKPGGGFHAAMSQEKSLLYINTDPDSEETRISKQGPPTDHDREVY
YVTDPAATLCSSSLHPPEGSVQKSQCEINPFMPQFPSTVQWTVPLTDSAHSPIYLQADWYT
AALQGFDFGQLRLSNVLRAPTATKEPKVVLVSRSRTFMIRSRLGEPVVLDCGFWDASSPL
SGSGFAVEWRYQFRGDGRLVLAYDGKSDRFAETQEEGVGLDFTALHETGDASLILQEAQV
RHSGTYICTVYLPYLLAQVAVELEIVGEER

>TroutUCA CM007948.1:28.598.666-28.609.318 XP_021416249.1
MKGLMLMFMGIGHLYEAFGVTHSLKHFTYASSDIPNFPEFVFGMVDDVQMVHYDSNSQK
AVPKQDWMKQTDAEFWEKERENFFYSQQSFKAEVATLKQHFNQSGGVHILQYMYGCSRDD
ETEQTEGFGQLGYNGENFLEYDMKTLTWKSRKQQANFMQDEWNRDISRLGFWKIYFSQTC
IDCLKKQVDNGKSTLMRTVSLQKTPSSPVTCHATGFYPSGVMVSWQKDGQDHHEDVENG
ETLPNDGTFQKSTHLLTVTPKEWKNNKYQCVVQVTGIKEDFIKVLTESEIQTNWKDPAPN
IVPIIGGMVALLLVVVVVGVVIWKKKSKKGFVPASRSNKLKNITYNTYNITHSIYIEQ
KYKRNSLKDFAELKFI

>TroutPSMB8bps_#B chr.14 CM007948.1:28.592.051-28.592.188
LYKLNRNKQRISVSAASKLLCNMMLRYRGMGLSMGSMIVGWDNKKVQG

>TroutUEA_#B CM007948.1:28.580.630-28.582.037 XP_021416247.1
PHSLKYLYTAVSGDITDFPEFTVVGLLDDQQFVHFGSNTKTLVNDAEWMNKTGKYNDLYN
EPLINQYEGFKNLITFAKKQFNQTQSKGVHTIQNLYGCEWNDTELQDYFHHYGYDGEDF
ISLDMKTVRWITSVQQADTIKQKWDNSKDLHYLKWYFTKECIDTLKKYVDFASSVLKKI
VSLLQKTPSSPVTCHATGFYPSGVMVSWQKDGQEQHEDVLHGEILLNGDGTFFQKSAHLTI
DSEKWENNNTYTCVVEHKENIIAIRLNQSVIKTNSVKPSEYVPIIIGVGVAVILLVLVPI
ITFIMWNRKST

>TroutUFA_#B chr.14 CM007948.1:28.552.026-28.552.825 No predicted orf
VSWMMTADFLSLDKSTRWTAAANQKAAIS*LKWDATGDNANYWKNYLENTCIEWLKKYVN
YGKDTLRSVSLQETPSSPVTCHATGFYPSGVMVFWQKDGQGHQEGVENGETLPNNDGT
FQKSTHLKVTPPEWKNNKYQCVVQVTGIKEDFIKVLTESETQPIWR

>TroutPSMB8b_#B chr.14 CM007948.1:28.546.790-28.549.183 XP_021416243.1
MALFDVSGYKSHAGLRGQILGTGVGHLVDRPNQEFAPVPGVDVSPSGFLKSCSREGGVSI
DLNHGTTTTLAFKFRHGVIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYW
ERLLAKECRLYKLNRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVD
DNATRLSGRMFSTGCGSSYAYGVMDSGYHEDMTVEEAYELGRRGITATHRDAYSGGVVN
VYHMQEDGWIKVKCEDVSELIHRYRKGMF

>TroutPSMB13b_#B CM007948.1:28.542.425-28.549.187 XP_021416241.1
MALSNVETPASGFNFENVSRNVALEGLLEGHTKAPKPMKTGTTIAGIVCKDGVVLGAD
TRATSGEVVADKMKAKIHYISPNNMYCCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
LEDRFKPNMELEEAKQLVRDAIHSGIMSDLGSGNNIDICVITKQGVDIIRPYQESEYKDK
RQRRYKYPGTTTILTEKIVPLEME

>TroutPSMB12b_#B CM007948.1:28.539.879-28.542.248 XP_021416245.1
MDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGNYVSSKTINKLIQVHDIRFCCIAGS
LADAQAVTKAAKFQISFHSIQMESPLVKAAASVLKELCYNNEELQAGFITAGWDRKKG
PQVYTVALGGMLLSQPFTIGSGSTYIYGYADAKYKPDMSREECLQFATNALALAMGRDN
VSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE

>TroutPSMB9b_#B CM007948.1:28.534.744-28.538.439 XP_021416244.1
MLDESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY

CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
 WDKRGGGQVYVTLNGLLSRQPFVAVGGSSSYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
 MSRDGSSGGVAYLVTIDEKGETFVVP
 >TroutTAP2b_#B 14:28.526.642-28.532.384 XP_021416240.1
 MLRTCVFAMAVGLCIDITTFCATGFGASIPKTGPISFDVFGSLRLRLVWVAGIRLVILFGV
 SLLTLGSIKPVLRWLAVHCFAPVYETGQLMLHGSSPESTHGSLLGGPSLWLLCTAAAAA
 AALFWEKTFPDSSEESNGKEKTQKARVLFMRVLYFYRPDTLLLVGAFIFLSLAVLCKYFI
 EMFIPFYTGKVIDILASQYKWNDFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTSRM
 KVELFGALVKQEISFFETIKTGDITSRLSTDTTKMARAVALNVNVLRLTIKIVGMLSML
 MSLSWKLTLLMLMETPVTGLLQGVDNYLRLTKEVQDSMARANEAAGETVAGIRTVRSF
 NTERSEASHYDHRLMDTHLTKTRRDTVRAVYVLLRRLTALVMQVAMLYYGRFLFIQRGQMS
 TGNLVSFILYQSNLGTNIRTLIYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLT
 GHVHFHNLFSYPTQRERKVLQGFSLRLRPGQLTALVGPSSGGKSTCVSLLERFYQPQQG
 EILLDGQPLHSYQHHLHRKVAMVQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANA
 HSFISQLEKGYDTDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHNV
 RHPSTHKLECFSLPSQTLLVIAHRLKTIERADQIILIDQGTVLEQGTQELMDRKGYYK
 LRERLFTEDDTSH
 >TroutUGAp_#B CM007948.1: 28.280.757-28.280.987
 VPASLSLLQKTHSSPVTCHATGFYPSGVMVFWQKDGPDHEDVEYGETLPNKDGTGFKNT
 HQWSSGKATINWVFSSQYTGSNRQPSRPIQSDE
 >TroutUGA_#B CM007948.1: 28.275.281-28.276.500 MHCI
 MCYMSVRHSLRYFYTTSSIEIPAFPEFVDMGMVNDQVISHYDSITKRKVPKQSWMETFFDQ
 QYWDSTTENLKSASIEFKTNIQIAQKRNFQTGGIHIQSODMYGCTWDDTGTGTFEGHHIGY
 DGQDLLTFDVKTATWIALVPQALHSMKWEMDLSGIKSKKSYLTQDCIEWLKKYLDYGKT
 TLQRTVPPSVSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQEQHEDVENGETLPNDGDT
 FQKSTHLTVTPEEWKNNKYQCVVQVTGIKEDFIKVLTEFEIQTNWGDPAAPNCIPIIGGVV
 ALLLVVVVGVVVIWKKKSKGF
 >TroutPSMB10b_#B CM007948.1:28.118.268-28.125.740
 MLNNSRPCQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILGA
 DTRATDDMVVADKNCMKIHYIAPNIYCCGAGVAADAETTTQMMSSNVELHSLSTGRPPLV
 VTVTRQLKQMLFRYQGHIGSSSLIVGGVDVTGAHLYSVYPHGSYDKLPFLTMSGAGAAIS
 IFEDRYRPNMELEEAKKLVRDAIAAGIFCDLGSNSVDLCVITQAGVQYLRGYDQPAPGT
 TAVLTKTVTPLPLDVVDESIQLMDTQ

Coho salmon (*O. kisutch*) MHCIa region; Genome GCA_002021735.1 LG17 NC_034190: 25.300.000-26.000.000 deduced amino acid sequences

Coho salmon genome assembly unpublished.

>CohoTAPBP NC_034190:25.979.910-25.984.738 XP_020361139.1
 MANISTIIKLSFLAFTYFIQVYGTSCPVLECFVQEKPGRGGGFPAAMIQEKALLYINTD
 PESEQTKSQRPASADISHRVYIITDPAAILCSSLHPPEGSVHKPQCEINPFMTQPTV
 QWAVPLTDSAHSPIYLQADWYSAALQGLDGLALSSVMRAPATATKEPTVVLVSSTPLV
 RSRLGEPVVLDCGFWAEASSPLSGSGFAVEWRYQFRGDGRLVLAYDGKTRFAETKEKGA
 GLDFTALHETGNASLILQEAQEVHRTGTIYICTAYLPYLLAQVAVELEIVEPPSLSIFFPSPL
 PLSVPGQVVKVQCEASGFFPLSLEFYWELTGPDGKVRPLGQGSVTGHRQGPNTYSQTTR
 LELESAKLDLGRGGEVTCVAVHPGGTTRASVTNLNIGVNGPSIEDSMAMVAVALGLYGLI
 KIVSWTFSSGSDDTNSQEKVK
 >CohoUBA NC_034190:25.919.304-25.940.726 XP_020361135.1
 MKGIILLVLGIGLLHTASAVTHSLKYFYTASSEVPNFPEFVVVGMVDGVQMVHYDSNSQR
 AVPKQDWNKAADPQYWERNTGNFKGSQQTFKANIDIVKQRFNQSGGVHIVQMMYGCEWD
 DETGVTEGFQHGVDGEDFIAFDLKTTKWIAPMPQAVITKHKLDSNTANNEYKKNYYTQT
 CIDWLKKYVDYDKSTLMRTVCPSVSLFQKTSSSLVTCHATGFYPSDVMVSWQKDGQDHHE
 DVEYGETLPNDGDTFQKSSHLTVTPEDRKNSKYQCVVQVKGIKEDFIKVLPLDAANVVP
 IIVGVVALLLVVAVVVGVVWKRKSKKGFVPTSTSDTSDNSGRAAPQI"
 >CohoPSMB8a NC_034190:25.910.797-25.913.925 XP_020361133.1

MALFDVSGYKSHSELRGQIIIGTGVGHFIDRPNQQFSVPVGVDPSPGFLKSCSREGGVSIDL
 NHGTTTTLAFTFRHGIVAVDSRASAGSYIESKEANKVIEINPYLLGTMSGSAADCQYWER
 LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDN
 ATRLSGRMFSTGCGSSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYSGGVVNLY
 HMQEDGWIKVCKEDVSELIHRYRKGMF"

>CohoPSMB13a NC_034190: 25.904.598-25.910.466 XP_020361130.1
 MALSNVLEIPTSGFNFENVARNVALEGLLEGGQTKTPKPMKTGTTIAGLVCKEGVVLGAD
 TRATSGEVVADKMKAKIHYISPNIYCCGAGTAADTEKTTDLLSSNLTIFSMNSGRNPRVV
 MAVNILQDMLFRYRGQIGASLILGGVDCSGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
 LEDRFKPNMEMEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQESEYKDK
 RQKRYKYGPGTTSILTEKIVPLELEVQETVQQMDTA

>CohoPSMB12a NC_034190: 25.901.088-25.904.637 XP_020361131.1
 MERHLMDSQIKGVSTGTTILAVTFNGGVIIGSDSRASIGGYVSSKTINKLIQVHDIRFC
 CIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAASVLKELCYNNKEELQAGFITAGW
 DRKKGPQVYTVALGGMLLSQPFTIGGSGSTYIYGADAKYKPDMSKEECLQFATNALALA
 MGRDNVSGGVAHLVVITEEGVEHIVIPGDKLPKFHDE"

>CohoPSMB9a NC_034190: 25.895.415-25.899.072 XP_020361128.1
 MLEESSEPGWLSEEVKTGTTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
 CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
 WDKRGGGQVYVTLNGLLSRQPFVAVGGSGSAYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
 MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ"

>CohoTAP2a NC_034190: 25.890.101-25.894.811 XP_020361127.1
 MTKMMLRTCVFAMAVGLCIDITTLAFGASISKGTGRTSDAFGNVVRLLWVAGIRLVLLLG
 LSLTLGSIKPVFKRWLAVHCFLAPVYETGRMLMYGGSPERVYRSLGLGSPSLWPLCTVA
 AAAAALFWETTTPDSNGESNGKQKKQKARVLFMRVLIIFYRPDILLVGAFFLSLAVLCE
 MFIFPYTGKVIDILGTQYKWNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMK
 VKLFGALVQEI GFETIKTGDTISRLSTDITLMGRAVALNVNVLRLTIKTVGMLSMLM
 SLSWKLTLMLMDTPITGLLQSVHDNYYQRLSKEVQDSIARANEAAAGETVGGIRTVRSFK
 TEQHEAGRYNDRMLMDTHNLKTRRDTVMVAYLLLRRLTAVVMQVAMLYYGRLFIQRGQMTT
 GNLVSFILYQSDLDNIRTLIYIFGDMLNSVGAAGKVFEYLDREPQVSTKGTLPETLTG
 HVQFHNLSFSYPTQRERKVLQGFSLRLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGE
 ILLDGLPLQSYQHLYLHRKIAMVQEPVLFSGSVKDNIAYGLADCSLERVQEAARRANAH
 SFISQLEKGYDTDVGERGGQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQ
 EALASCPSQTLVLIAHRLKTIERADQIILIDQGTVQEQGTHQELMDRKGSSYYKLKERLFT
 EDDVPH

>CohoPSMB8Fa NC_034190.1:25.878.698-25.879.229
 RIYKLRN*ERISVSAASTLLANMVVNYRGMGLRGTMICGCDQGPGLYYVDDNGLRLGNML
 TTGSGNYGMMDSGSQYDLVPEAYDLAQRAIFHATHRDAYSGGTVN

>CohoPSMB10a LG17 NC_034190.1: 25.541.569 -25.544.041 XP_020361094.1
 MLHNSRPPQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILAA
 DTRATDDMVVADKNCMKIHYIAPNIYSCCGAGVAADA EVTTQMMSSNVELHSLSTGRPP
 LVVTVTQRQLKQMLFR*ES-MNESVTMGK*DLYSGHSHNMAAISIFEDRYRPNLEEAKKL
 RDAIAAGIFCDLGSNSVDLCVITQAGVQYLRGYDQPAQKKGKQYKYKPGTTAVLTKT
 VTPLPLDIVDESIQLMDTQ

Coho salmon (*O. kisutch*) MHCIb region; LG14, NC_034187 22.700.000-23.500.000 deduced amino acid sequences

>CohoTAPBPb NC_034187:23.335.122-23.343.564 XP_020356015.1
 MTNISIIILQSFLAFSYVMHVYGASCPVLECFVQEKPGGFHAAMSQEKSLLYINTDPDS
 EETRSKQGPPTDHDVYYYVTDPAATLCSSSLHPPEGSVQKSQCEINPFMPQPSTVQWTV
 LTDSAHSPYIYLQADWYTATLQGFQGLRLSNVLRAPTATKEPKVVLVSVSRTFMIRSRLG
 EPVVLDCGFWDASSPLSGSGFAVEWRYQFRGDGRLVLAYDGKSDRFAETQEEGSGLDFT
 ALHETGDASLILQEAQVRHSGTYICMVYLPYLLAQVAVELEIVEPPSLSIFPSPLPLSMP
 GQVVTVQCEASGFYPLSLEFRWMLTGPDGKVRPLGQGSVTGHRQGPSTFSQTSRLELDS
 AKLDLGGVETCVAVHHGGTRRASVTLNIAGVSAPIEDYAMAMVAVALGIYGLIQVVS WTF
 >CohoUDA NC_034187:23.304.841-23.315.855 XP_020356011.1
 MKAFMLMFMGIGHLHEAFVTHSLEHFYTASSELHNFPEFMVVGMDGVQMVHYDSNIQK
 AVPEKDWKQTDTEFWEKEKENFFYSQKSFKAEVANLQHLNQSGGVHILQYMYGCSRDD

ETEQTGEGFQGLGYNGENFLEYDMETLTWKSARKPQANFMQDEWNSDISRLGFWKNYFSQTC
IDCLKKQVDNGKSTLMRTVPSPVSLQKTPSSPVTCHATGFYPSGVMVFWQKYGKYQHEG
VEHGEILFNDDDETQKSTHLLTVMPEEWKNNKYQCVVQVTGIKKDFIKVLTESEIQTNWKD
PAPNIVLYIGGVVAFLLVVVVVVVGVIIWKKKSKKGFVPASTSNTDWDNSG
>CohoPSMB8b3 NC_034187.1:23.299.625-23.299.750
SKLLCNMMLGYRGMGLSMGSMIVGWDNKVQG
>CohoPSMB8b2 NC_034187.1:23.289.363-23.296.651
ALKEANKVIEINPYLLGTMSGSAADCQYWERLLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLS
>CohoUEA NC_034187:23.276.214-23.279.219 XP_020356010.1
MISFCLLFLWIPTMSADPHSLKYIYTAVSGDTEFPEFTVVGLLDDQQFVHFGSNTKTLVN
DAEWMNKTKGYYRELYNETLINQYEGFKNLITFAKKQFNQTQSKGVHTIQONLYGCEWNDE
TELQDYFHYYGYDGEDFISLDMKTVRWITSVQQADTIKQKWDNSKDLHYLKWYFTKECI
DTLKEYVNFASSVLKKIVPPSVSLQKTPSSLVTCHATGFYPSGVMVSWQKDGQEQHEDV
LHGEILPNGDGTQKSAHLTIDSEKWNNNYTCVVEHKENIIAIRLNQSVIKTNSVKLLE
TVPIIIGGVVAVILLVTFTITRIFIMWNRKSTASDDGSNPSMTQQNQIQLEVSLSLIKQE
RSREKRTFIEGQ
>CohoPSMB8b NC_034187:23.256.442-23.259.352 XP_020356008.1
MALFDVSGYKSHAGLRGQILGTGVGHLVDRPNQEFAPVGVDPGSGFLKSCSREGGVSIDL
NHGTTTTLAFKFRHGIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQYWER
LLAKECRLYKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDN
ATRLSGRMFSTGCGSSYAYGVMDSGYREDMTVEEAYELGRRGITHATHRDAYS GG VV NLY
HMQEDGWIKVCKEDVSELIHRYRKGMF
>CohoPSMB13b NC_034187:23.251.688-23.255.606 XP_020356007.1
MALSNVETPASGFNFENVSRNVALEGLLEGHTKAPKPMKTGTIAGVVCCKDGVVLGAD
TRATSGEVVADKMAKIHYSIPNMYCCGAGTAADTEKTTDLSSNLTFISMNSGRNPRVV
MAVNILQDMLFRYRGQIGASLILGGVDCTGNHLYTVGPYGSIDNVPYLAMSGDLAALGI
LEDRFKPNMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVDIIRPYQSESEYKDK
RQRRYKYPGTTTILTEKIVPLELEVQETVQRMDDTA
>CohoPSMB12b NC_034187:23.248.913-23.251.681 XP_020356009.1
MDSQIKGVSTGTTLAVTFNGGVIIGSDSRASIGGYVSSKTINKLIQVHDRIFCCIAGS
LADAQAVTKAAKFQISFHSIQMESPLVKAASVLKELCYNNKEELQAGFITAGWDRKKG
PQVYTVALGGMLLSQPFTIGSGSTYIYGADAKYKPDMSREECLQFATNALALAMGRDN
VSGGVAHLVVITEEGVEHVVIPGDKLPKFHDE
>CohopSMB9b NC_034187:23.243.647-23.248.046 XP_020356006.1
MLDESSEPGWLSEEVKTGTIIAIEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
CALSGSAADAQTIAEMVNYQLDVHSIEVGEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDKKGQGVYVTLNGLLSRQPFVAVGGSGSSYVYGFVDAEYRKAMSKEDCQQFVVNTLSLA
MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ
>CohoTAP2b NC_034187:23.236.849-23.242.173 XP_020356005.1
MMLRTCVFAMAVGLCIDITTFCATGFGASISKTPISFDLFGNLLRLWVAGIRLVILFG
VSLTLGSIKPVLRWLTVHCFAPVYETGQLMLHGSSPESTHGLGGPSLWLLCTAAAA
AAALFWEKTFPDSNEESNGKEKTQKARVLFMRVLYFYRPDTLLLVGAFIFLSLAVLCMEF
IPFYTGKVIDILASQYKWNDFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTSRIKVE
LFGALVKQDISFFETIKTGDITSRLSTDTTKMARAVALNVNVLRLTIKIVGMLSLMMSL
SWKLTLLMLMETPITGLLQGVYDNYLRLTKEVQDSMARANEAAGETVAGIRTVRSFNTE
RSEAGHYDHRMLMDTHNLKTRRDTVRVYVLLRRLTALVMQVAMLYYGRLFIQRGQMSTGN
LVSFILYQSNLGTNIRTLYIIFGDMNSVGAAGKVFEYLDREPQVSAKGTLPETLTGHV
HFHNLFSFSYPTQERKVLQGFSLRLPGQLTALVGPSSGGKSTCVSLLERFYQPQQGEIL
LDGQPLHSYQHYYLHRKVAMVGQEPVLFSGSIKDNIAYGLADCSLERVQEAARRANAHSF
ISQLEKGYDTDVGERGGQMSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQEA
LASCPSQTLVLIAHRLKTIERADQIILIDQGTVLEQGTHTQELMDRKGYYKLRERLFTED
DTSH
>CohoUGA NC_034187:22.955.555..22.971.186 XP_020355985.1
MLLYHFSPRCILRHSRLRYFYTTSSIEIPAFPEFVDMGMVNDQVISHYDSITKRKVPKQSWM
EPFFNQYWDSTTENLRSAESVFKTNIQIAQKRFNQTGGIHIHSQDMYGCTWDDDETGFTEG
FHHIGCDGQDLLAFDVKTATWIASVPQALHSMKMWEMDLSSIESKSYLTQDCIEWLKKY
LDYGMTTLQRTVPPSVSLQKTPSSPVTCHATGFYPSGVMVSWQKDGQEQHEDVENGETL
PNDDGTFRSTHLKVMPEEWKNNKYQCVVQVTGIKKDFIKVLTEFEIQTNWGDPAPIIIP
IIGGVVALLLVVVVVVVVVVGVVIWKMMSKKGFVPARTSDTDSSENSGERSPDDLKNFCNIR
ELKRTATEITNRNFPRL
>CohoPSMB10b NC_034187:22.819763-22.825.755 XP_020355973.1

MLNNSRPCQPQSAGFSFENTRRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILGA
DTRATDDMVVADKNCMKIHYIAPNIYCCGAGVAADA EVTTQMMSSNVELHSLSTGRPPLV
VTVTRQLKQMLFRYQGHIGSSLIVGGVDVTGAHLYSVYPHGSYDKLPFLTMSGAGAAIS
IFEDRYRPNMELEEAKKLVRDAITAGIFCDLGSNSVDLCVITQAGVQYLRGYDQPAQKG
KREGQYKYKPGTTAVLTCTVTPPLD VVDESIQLMDTQ

Northern pike (*Esox Lucius*) genome GCA_000721915.3 MHC la region LG10 (NC_025977.3) deduced amino acid sequences

Northern pike genome assembly: Rondeau et al.[5].

```
>PikeTAPBP NC_025977.3:7.006.557-7009.875 XP_010899738.2
MAXISTIYKLSIIAFIYFTNVYGTSCPVLECFVQERPGRGGGFPAAMSQEKSLLYINTD
PEWKRPEQQRPSTDINXDRLYITDPAATFCSSSLHPPEGSVQKPQCEINPFMPQPSSV
QWTVSLTDSAHSPIYLQADWYSASLQGLNGQLSLSSVMRAPTATKEPAVLLSVSSRTPVV
RSRLGERVVLDGCFWAEASSPLSGSGFTVEWRYQFRGDGRLVLAYDGKNDRLAETQEKGA
ELDFTALHERGNASLILQEAQVRHSGTYICTVYLPYILXQVAVELEIVEPPSLSI FXPPL
PLSVPGQVVTVQCEASGFSPLSLELHWELMETDGNVRPLGQGSVSGHRQGPDTYSQTTR
LQLDSAKLDLGRGGEVTCVAVHPGGTRRARVTNLNVIGVSAPSIEDSMAMVAVALCLYGLI
KVISWTFISSGSEDAEKKAK
>PikeUBA1 NC_025977.3:7.021.021-7.034.658 XP_019905630.1
MSGVQTTHSLRYVYTATSGIPDFPEFVTVGLVNGEPISYYSII RTETPRQDWMXKAVDP
EYWSRNTQLSIGTEKTFKDNIDVVNQRFNQTMGVHINQLMYGCEWDDDETGVTEGFDQFGF
DGEDFISLDLMTLTWIAKPKQAVISKRLRDNINIVHSNNYLKNYFTQECIDWLKKFLDYGK
SSLMRTVPSPVSLQLKTPSSPVTCHATGFYPSEVIVTWKKDQGEQHEDEVEGETLHNDG
TFQKSVHLTVKPEEWKNIKYQCVVQVSGINEDFIKVLTEDQIQTNCRKTNRPDTSTGLII
GVVVGVSILVIMAAVFGVVIWKNKSKKVSDDGSGNSNNSSPKP
>PikeUBA2 NC_025977.3:7.043.107-7.060.011 XP_010899737.2
MKVILSILGIGFLPTVSAVTHSLKYFYTGVSXINFPEFTTVGLVDNGQFIYFDSNTKT
AVPKTEWMKREGPDYWDSETQTGINLHQVFKNISIQSTMDRFNQSMSTGVHVQVMYGEW
DDETEATEGFDRYGYDGEDFLLDLKTLSWIAPKSAFITKNKFD SNKDTNLYLRKYFTQ
TCVDWLKKYVDYGKSSLMRTVPPTVSLQLKTPSSPVTCHATGFYPSRVNVTWKKNGQEQH
EDVEMGEILPNDGTFQTSVRLNVKPEERKNNKYQCVVQVAGINEDFIKDLTEDTTSIGP
IIGVVVAGLLVIVA AVIIGVVIWKKKSKKGFVPANTSDDGSGNSNNNTAPKA
>PikePSMB8F NC_025977.3:7.084.970-7.085.777 XP_019905769.1 GATF01015276.1
match
SVSAGSKLLANMVVNRYRGMGLSMGTMICGCDKKGPGLYYVDDNGMRLCGNMFSTGSGNSY
AYGVMDSGYRYDLSPVEAYDLAQRAIFHATHRDAYS GGTVNMYHMKETGWVKVSQEDVGD
LYHRFNENK
>PikePSMB8 NC_025977.3:7.095.708-7.109.763 XP_010866033.1
MALFDVSGYKSHLELRGNTLGTGLRHLNFGTTNKEFAVPAGMDPSGFLKSCSHEGQGGVS
IDLNHGTTTTLAFKFRHGVIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQY
WERLLAKECRLYKLRNKKRISVSAASKLLSNMMLGYRGMGLSMGSMIIGWDNKGPGLYYV
DDNGTRLSGQMFSTGCGNSYAYGVIDSYREDMTVEEAYELGRRGITHATHRDAYS GGVV
NLYHMQEDGWIKVKEDVSELIHRYRKGFM
>PikePSMB13 NC_025977.3:7.111.771-7.114.933 XP_010866031.1
MALPNVLETPPTGFNFENVFRNVALEGLLEGGRTKAPKPMKTGTTIAGVVCKDGVVLGAD
TRATSSEVVADKMC AKIHYIAPNIYCCGAGTAADTEKTTDLLSSNLTIFSMDSGRNPRVV
MAANILQDMLFRYRQIGAHILGGVDCTGSHLYTVGPYGDIDKVTYLAMSGNLAAALGI
LEDRFKPNMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGV DYL RPYQESEFKDK
RQMRKYKHPGTTPILTEKIVPLQMEVVQETVQRMDTA
>PikePSMB12 NC_025977.3:7.115.084-7.119.208 XP_010866032.1
MERHYMDSKIKGVSTGTTILAVTFNGGVIIGSDSRASMGGNYVCSKTINKLIQVHDRIFC
CIAGSLADAQAVTKTAKFQISLHSIQMESPLVKAAASVMKELCYNNKEELQAGFITAGW
DKKKGPQVYTVALGMLLSLPFTTIGSGSGSTYIYG YTD AKYKPDMSREECLQFATNALALA
MGRDNVSGGVAHLVVITEGGVEHVVIPGDKLPKFHDE
>PikePSMB9 NC_025977.3:7.121.547-7.131.354 NP_001297802.1
MLEESTESEWQSEVKTGTTIIAVEFNGGVVLGSDSRVSAGETV VNRVMNKLSLLHDKIYC
ALSGSAADAQTIAEMVNYQLDVHSIEIGEDPQVRS AATLVKNISYKYKEELSAHLIVAGW
```

DRRGGGQVFTLNLGLLSRQPFVAVGGSGSSYVYGFVDAEYQKGMSEDCQQFVVNTISLAM
 SRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ"
 >PikeTAP2 NC_025977.3:7.133.087-7.143.465 XP_010866026.3 polished with TSA
 sequence GATF01026336.1
 MKMIIKTCLLAIALGLDITIFSAPGFGASISKIGQITYTFGYIVSVWIVAVIRFVFLVSV
 SVTTLESIKPIFKRWLAACHLLVPVYETGRLILHGRSLESSCGPLGGPRLWLVTVAAT
 AAFWETTFPDSNGESNEKEKKQKARVLFRLVVRFCKHDSLILICAFIFLSLAVLCEMFI
 PIYTGKVIDILGTKYMWNDERTVIIIFMGLCSLGSSFSAGCRGGLFMCAINSFTCRIKVQL
 FGALVKQEIGFFETIKTGDVTSRLSTDAPKMGRAVALNVNVLRLTAIKTVGMLSLMMSLS
 WKLTLLMLMETPITGLLQSIYDNYQRLDKVQDSLARANETAGETVAGIRTVRSFKTEQ
 IEAGRYDNRLMDTHNIKTRRDTVRAYVLLLRLLIALVTQVAMLYFGRLVQVQGMQMTTGNL
 VSFILYQSDADNIRTLIYVYGDMNSVGAAGKVFEYLDREPHVSTKGTLQPETLTGHIC
 FHNLSFSYPTRQEHKILQDFSLELKPGKLTALVGPSGGGKSTCVSLLERFYQPQQGEILL
 DGQPLQSYQHHLHRKVMVGVQEPVLFSGSIKDNIAYGLKDCSLERVQEAARRANAHGFI
 SKLEKGYTVDVGERGNQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMQEAL
 ASCPSQTVLVIAHRLKTIERADQIILIDQGTVLEQGTGTHKVLMMKGSYYKLRERIFTEDD
 VPK
 >PikePSMB10 NC_025977.3:7.488-404-7.493-157 NP_001291088.1
 MLNTRSPCQPQSGGFSFEDSRNAVLEGNLSELGYSSPKARKTGTTIAGIVFKDGVILGA
 DTRATDDMVVADKNCMKIHYIAPNIYCCGAGVAADAETVTQMMSSNVELHSLSTGRPPLV
 VTVTRQLKQMLFRYQGHIGSSLIVGGVDVTGAHLYSAYPHGSYDKLPFLTMSGGAGPAIS
 IFEDRFRPNMELEDAKQLVRDAIVAGIFCDLGSGSNVDLCIITQAGVQYLRAVDQPAQKG
 KKEGQYKYKPGTTAVLTCTVTPLPLDIVDESQVQLMDTQ

Esox Lucius unplaced genomic scaffold Un_scaffold0763 NW_017859271.1 deduced amino acid sequences

>PikeMHCI.3_NW71 NW_017859271.1:13.681-29.329 XP_012986808.1
 MKVFILLIMIGIGLLHSVSAATHSLCYVYTATSGIPDFPEFVTVGLVNRPIISYYDSIIRT
 ETPRQDWMANNEGSDYWESQTQISNGTEQTFKANIDILKQRFNQTGGVHVQNMYGCEWD
 DETGVTGDFDQYGYDGEDFIVFDLKRILTIAAPKPAFINKNKWDNDRGWIEQTKNYLTQT
 CIEWLKKYVDYKGSSLMRTVPPSVFLLQRTSPSPVTCHATGFYPSRVMVTWMKDGQEQYE
 DVEVGETLQNDGTFQKSVHLTVKPEEWKKNKYQCVVQVSGIKEDFIKDLTEDQIQNTTG
 FVPANTSDDGSNSSNNTAPKA
 >PikeMHCI.4_NW71 NW_017859271.1:69.714-85.450 XP_012986807.1
 MKVILLILIGIGLLPTVSAASHSIRYVYTATSGIPDFPEFVNLGILDGMQIDYYDSNTKK
 VVPKQDWMKAVGSDYWDRETQNSIGREKSFKANIDTAKQRFNQTGGVHVQMMFGCEWD
 DETGVTRGFNQGGFDGEDFISFDLKTLTWIAAPKPAVITKNKWDNNKAGNEYWKNYLTQE
 CIDWVKKYVDYKSSLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVHVTWKKDGDHHE
 DVEMGETLQNDGTFQKSVRLNVKPEEWKKNKYQCVVQISGTEKDFIKDLTEDELVSNWG
 KTNRGVNDQTSIGPIIGVVVAGLLVIVAIAVIGVVIWKKNRKGFVPANTSDDGSSTNSAP
 EA
 >PikeMHCI.5_NW71 NW_017859271.1:149.722-177.971 XP_010866362.2
 MKVFILLILIGXGHLHTLSAFAHSLKHTCTAVSGDTHLPEFTALGLLDNVQVAYFDSNTKT
 AVPKTEWIRREGPDYLNLSLTNLGINAHQRLKAHLQIFKDQFNQSMSTGVHTLQVMMGCEW
 DDETGVTEGYMQFGYDGEDFISFDLKTLTWITSKPQAVITKDRWDNDKADNEYWKNYLTQE
 ECIDSLKNFLDYKSSLMRTVPPSVSLLQKTPSSPVTCHATGFYPSSEVIVTWKKNGQDHH
 EDVEMGETLHNDGTFQKSVHLTVKPEEWKKNKYQCVVQVSGINEVFIVKVLTEDEIETNW
 GKTQK"
 >PikeMHCI.6_NW71 NW_017859271.1:189.068-206.991 XP_012986804.1
 MYFFVLLILIIIGHLHTVSAVTHSLKYFYTGVSVDINFPEFTTVGLVDNVTFIYFDSNTKT
 AVPKTEWIKRVGPGYVWSQTKTCINLHQFFKNNIQVAKDRFNQSMSTGVHVLQWMYGCDW
 DDETGVIIQGFQYGYDGENFISLDLKTLSWIAAPKPAFITKNKWDSDKADNEYLKHFTL
 ECIEWLKKYLQYEETSQMRTVPPSVSLLQKTPSSPVTCHATGFYPSSEVNILWQKDGPFOH
 EDVETGHTLPLNGDGTQKSVLLTVKPEEWKKNKYQCVVQVSGINKDFIKDLTESEIQTNR
 DTSIGHIITVVVAGLLAIGAAVIGVVRKSKKSGKFVPNTSDDGSNSSNKSSSNV
 >PikeMHCI.7_NW71 NW_017859271.1: 215.029-249.012 XP_012986801.1
 MKVIVLLILIGIGLPTVSTVTHSLKYFYTGVSVDINFPEFTTVGLVDNQGFIYFDSNTKTA
 VPKTEWMKREGPDYWDSTQILIGSQPVFKNNIQVAKDRFNQSMSTGVHVVMYMMYGCEWD
 DETGVIEGFDQGHYDGEDFISFDLKTLLKVVAPTQAFISKNKLDYNEAGNEGEKHLYLTQE

CIEWLKKYVDY GKSSLMRTVPPSVSLLQKTPSSPVTCHATGFYPSGVHVAWKKGQDHHE
 DVEMGETLQNDGTFQKSVLLTVKPEEWKNNKYQCVVQVSGINKDFIKDLTESEIQTNRE
 PTSIGPIIGAVAGLLVIAVVVIGVVIWKKSKKGFIPANTSDDGSNSSNNTAPKA
 >PikePSMB8_NW71 NW_017859271.1:275.209-276.945
 MAIFCVSGYNHSQLRGEITGKGFSSVLSL*PSGFLKSYSHEGQGGGSIDLNHGTTTLAQ
 FRHGVLTTLTHGVIVAVDSRASAVSYIASKEANKVIEINPYLLGTMSGSAADCQYWERLLA
 KECRAEKLNRKKRISVSAASKLLSNMMLGYR

Esox lucius unplaced genomic scaffold Un_scaffold1132 NW_017859580.1 deduced amino acid sequences from pike genome [5]

>PikePSMB8_NW80 XP_010866810.3 NW_017859580.1:27.821-34.344
 MALFDVSGYKSHLELRGKTLGTGLRHFNFGTNNQEFAPVAGMDPSGFLKSWRHEGQGGVS
 IDLNHGTTTLAFKFRHGVIVAVDSRASAGSYIASKEANKVIEINPYLLGTMSGSAADCQY
 WERLLAKECRLYKLNRKKRISVSAASKLLSNMMLGYRGMRLSMGSMIIGWDNKGPGLYYV
 DDNGTRLSGQMFSTGCGNSYAYGVIDSGYREDMTVEEAYELGRRGITHATHRDAYS GGVV
 NLYHMQEDGWIKVCKEDVSELIHYYKKGFM
 >PikePSMB13_NW80 XP_010866807.1 NW_017859580.1:21.631-26.326
 MALPNVLETPPSGFNFNDVFRNVALEGLLEGGRTKAPKPMKTGTIAGVVCKDGVVLGAD
 TRATSSEVVADKMCACIHYIAPNIYCCGAGTAADTEKTTDLLSSNLTIFSMDSGRNPRVV
 MAANILQDMLFRYRQIGAHILGGVDCTGSHLYTVGPYGDIDKVTYLSMGSGDLAALGI
 LEDRFKPNMELEEAKELVRDAIHSGIMSDLGSGNNIDICVITKQGVYIRPYQESEFKDK
 RQRRYKYHPGTTPILTEKIVPLQMEVVQETVQRMDDTA
 >PikePSMB12_NW80 XP_012988529.1 NW_017859580.1:16.159-21.594
 MERHYMDSKIKGVSTGTTILAVTFNGGVIIIGSDSRASMGGSYVCSKTINKLIQVHDRIFC
 CIAGSLADAQAVTKTAKFQISLHSIQMESPLVKAASVMKELCYNNKEELQAGFITAGW
 DKKKGPPQVYTVALLGMLLSLPFTIGSGSTYIYGYTDAKYKPDMTREECLQFATNALALA
 MGRDNVSGGVAHLVVITEGGVEHVVIPGDKLPKFHDE
 >PikePSMB9_NW80 XP_010866806.1 NW_017859580.1:12.295-15.497
 MLEESSEPGWPSEEVKTGTTIIAVEFDGGVVLGSDSRVSAGETVVNRVMNKLSSLHDKIY
 CALSGSAADAQTIAEMVNYQLDVHSIEIGEDPQVRSAAATLVKNISYKYKEELSAHLIVAG
 WDKRGGGQVFTLNLGLLSRQPFVAVGGSGSSYVYGFVDAEYQKGMKEDCQQFVVNTISLA
 MSRDGSSGGVAYLVTIDEKGAEKCILGNELPTFYDQ
 >PikeTAP2_NW80 XP_019900791.1 NW_017859580.1:4.666-11.424
 MIKACVLAFALISIDIILTTFSAPGSGSSKTGPITCDAFGNLVNVWIVAVIRLVLLVTVS
 VSTLGSIKPIFKRWLTVHCLLAPVYVTVGRILILHGSSLESAYGPLGGPRLWLVTVA AAAA
 ALFWETTFPDSNGESNGKEKKEKKQKALVLFMKVLRFYKPDILFLTGAFFLSLAVLCQM
 CIPFYTGKVIDILGAQYKWKDFRTAIIIFMGFFSLGSSFCAGCRGGLFMCAINSFTCRMKV
 QLFGLTVLQKEIGFFETNKLGDITSRLSTDITLMGRAVALNVNVLLRTAIKTVGMLSMTS
 LSQKLTLLMLMETPITGLLQSIYDSYYQRLAKEVQDSLARANETAGETVAGIRTVRSFKT
 EQSEAGRYDDRLMDTHNIKSRRDTRAVYLLLRLLIALVMQVAMLYFGRFLFIQQGQMTTG
 NLVSFILIYQSDLDADNIRTLIYIFGDMNLNSVGAAGKVFEYLEREPQVSTNGTLQPETLTGH
 IFFHNLSFSYPHKKERKILQDFSLELKPGKLTALVGPSSGGKSTCVSLLERFYQPQQGEI
 LLDEQPLQNYQHLYLHRKVAMVGQEPVLFSGSIKDNXAYGLKECSLERVQEAAXRANAQG
 FIRXLEKGYTVDVGERGNQLSGGEKQRIAIARALIREPQVLILDEVTSALDTESEHMIQE
 ALASCPSQTVLVIAHRLKTIERADQIILIDQGTVLEQGT HKVLMEMKGSYYKLRLRFTV
 DDDPK

Deduced amino acid sequences from Spotted gar (*Lepisosteus oculatus*)

>Spotted gar TAPBP GFIM01016833 No genomic match *Lepisosteus oculatus*
 MSDTATIYRLTLLALWGLAAVCGSRCPVLECWYVQEKAGQGGGFPAAMTQEKSLLYVRTE
 PDPVGLQONPPSDIDPKRIIYVTDAGASLCSPLRAAEGAVEKPKQCEVSPFLPQPALVRW
 AAPLTAAGSSPAHLQADWLSAALRGLQGQLVLSLLRAPASDQASVVLVSVFSRRLVVRT
 RLGA PALLD CGFWARPGSPLWGAGFAVEWRYQFRGEGRLVLAYDGRGDRLEEAEPGAELD
 AA EAHRTGNVSLRLGGA EVRHAGTYICTVYLPYLLAQVAMELEVVEPPALLLSPLWAL
 PGHSAPVQCEASGFFPLALELQWEFLPAGGAGEALALEGATVSGHRQGRAGTFTQTSRLE
 LD LGRLRLEHGGQLVCVAKHEGGTRRPGVT LN IAGVGGPTVEDSMAMVAVALLLYGLIKI

VSWTLKTTDSSVPPEEPEKKQE*
 >Spotted gar TAPBPR ENSLOCG00000006844 LG26:11.560.611-11.567.557
 XP_015193320.1 *Lepisosteus oculatus*
 MRLAFGVLVLASHLMSSVSGSVADVVLACSLVEEGGGMGRLAGAGALFTRDEATLVLRLD
 LPVTADES LDTVTPFAPPAPDPENLIFEATVTSLEIPEADSLHADCNEQEVTCEISRYF
 PRGPEGTEPSPEDAFFIGSLQLEGGGVSLTLVLRTQPIPAEQEEDGGRPLRQSKLDLPLS
 PSGTILNEVVVVFTRLQSLTAPLGGSALIDCGYKEATPSQEVALEWRLQHKGHGRRILQ
 LRAGREGEEPTVHPEREGASVEPGLVLEEGNVSLTLTNVKVADEGTYYICTVSGSVYQAQQ
 VIQLHITQTPSVTLTPNQLVFQDDTPQRVICHCDHYYPLDVQVEWFSLSPSASEPVPISN
 GVIYFSSHRQHS DGTYSLSAYISVSPSEDLTGATFSCIVSHHSLSEPITASITVSAPEESQ
 LWSMVGGLVLSFVLFGLFGLFMLLR
 >Spotted gar TAPBPL LG22: 10.745.093-10.749.329 ENSLOCG00000006827
 GFIM01040944.1 *Lepisosteus oculatus*
 MITTAVLLCVLICSGRLRCVRSESLAEAGLLACELVEERVSSGLSGAPETQFVNRKAALLL
 ARPGREADTAQRLQEAQRAGVVTFLATGSAVDVTRHVPGGTEQLQCEIRRYSTGGIQVPW
 KGTAGPDSAWFTCTLRHSANEFIVTAFLLYPSAPQPIGDADTLNTSVAMVVLTRTPTV
 QAGLLQDARLDCGFAVDHAQPDVTVEWRFRHRGEQRKLFYSRRSGRAETEVP GATAAAR
 DVSRGDASLRLSAVKVKSEGTYVCSVYVPPLYGSHDVQLQIMESPRVSLSVATALALQEE
 EEQKVVCETITGYPLDVQVDWLHEQGGGGRLPQVIKNILFTSHRHHADGTYSISAFFILR
 PSLADHDARYTCRVSHVSLRTPIRKSFTLTVTTERLRIGSILFWIFVAFIVMMLLLLLLVIL
 LPYLHSARREAKKKKPY*
 >Spotted gar ERp57 LG3: 45.308.360-45.321.947 ENSLOCG00000014242
 MRRDNMIRHGMNGWQTVPCTSLVLYLGVPRTPSPCFLAVAVSASDVLEFTDDDFESRIADH
 DLVLVEFFAPWCGHCKRLAPEYESAATRLKGIVPLAKVDCTANSNVCNKFVSGYPTLKI
 FRGGEESGAYDGPRTADGIVSHLKKQAGPASVEIKTVEEFKFIGDRDASVIGFFAEGGS
 TTQAEFLKSASALRESYRFAHTNSEELLQKHGIDGEGIVLFRPPRLSNKFEESSVKYSED
 KYTSGKIKKFIQDNILGICPHMTEDNKKDLKGDLLVAYFEVDYKKNPKGTNYWRNRVMK
 VAKSFLDQGGKLSFAVANKNSFSHEISEFGLDSSSGELPVVGIRTAKGDKYVMQEEFSRD
 GKALERFLQDYFDGKLRYLKSEPVPESNDGPVKVLVAENFEEIVNDDSKDVLIEFYAPW
 CGHCKSLEPKYKELGEKLSDDPNIVIAKMDATANDVSPYEVGRGFTIYFSPAGKKQSPK
 KYEGGREVSDFISYLLKKEASNPLVIQEEEEKSKKKKSELSCLLERMKEQCLVDVN
 >Spotted gar ERp57L LG24:5.612.549-5.621.013 FGENESH predicted
 MARLSVRLCPAALLALAAAGAVAVAGGDVLELGDADF DGAQAQHETVLVEFFAPWCGHC
 QKLAPEYEEAATRLKGAVPLAKVDCTAHSETCSRFGVNGYPTLKI FRNGQESSAYDGPRT
 ADGIVSYMKKQAGPSSVELRTEEDLDSFVSHFDGSSVVGFFLGPEADLAEFLLRAASSMRD
 SYRFAHTSAESLRQRHGVQGEAVLLFRAPQLASKFEESVVRHRGAVRPDALRRFIRENIF
 GMCPHLTHENREQLMRQDLLTAFYELDYLRNPKGSNYWRNRVMMVASRFAGRGLSFAVAD
 RRDFPDELEEFGLGAEAGELPVAAVRTVLGQKYVMREEFTRDGKALERFLEDYFAGRLKR
 YLKSEPVPESNDGPVKVVVAETFDEIVNDASKDVLIEFYAPWCGHCKNLEPRYTELGQQL
 SGDPNIVIAKMDATANDVPEGYDIQGFPTIYFAPVGQKAAPRRYEHAHFHFCISRWLKTR
 QVCPL
 >Spotted gar PDIA4 LG9: 50.559.604-50.567.699 ENSLOCG00000013319
 MKVKKAFLLLALLGLAHLALLASCEESVKPESIRSEMTQRAVSHRLYFSSTTRDFLWVRK
 TGLVILKDSNDFSFIADKDTVLVEFYAPWCGHCKQFAPEYEEKIAETLKEHDPPIPAKVD
 ATAESGLSRFEVSGYPTIKILKRGQPVVDYDGERTEQAIVSRVKEVARPDWTPPPEATLV
 LTKDNFDNVNEADIILVEFYAPWCGHCKRLAPEYEKAAKELSARTPPVPLAKVDATAEA
 ELATRFDVSGYPTLKI FRKGKAFEYNGPREKYGIVDYMIEQAGPPSKQVQALKQVQESVK
 DGDDVFIIGVFASDEDAAYEIIYQEACNNLRDDYKFHHTFSDEIAKFLKASPGQVIMMQPE
 KFQSKYEPKSSTFQIKDSTTASEVKDFFTQHALPLVGHKQSNDAKRYTKRPLVVVYGV
 DFSFDYRVATQFWRSKVLEVAQDFPEYTFIAIDEEDYAEELKALGLAESGEEVNAGILGE
 GGKKYAMEPEEFDS DALRDFVTAFKKGKLKPIIKSQPVPKNNKGPVKTVVGKTFEEIVLD
 PKKDVLIEFYAPWCGHCKKLEPEYLELGKKYKNEKGLVIAKMDATANDVLHESYKAEGFP
 TIYFAPSGDKQHPIKFEDGKRDLEGLSKFLEQHATKLSQKRDEL

Deduced TAPBP-like amino acid sequences from other species

>Frog TAPBPL XP_018100952.1 tapasin-related protein-like [*Xenopus laevis*]
 MDSRLYLIIILNLLIVCNDIQSFEDLTSRKSRKLS CMYVDMRNKTL LGYNDIMKQKAFLVL
 SSRKQLSKEYFDSEGITFILKDSPVNTMKFLKEHTLDLVCEVNPYFTDNTQIIWPGIHVI

ENMEEAWFVSRFKHPNDKFQFTLFFTHISQKSLEEQDGSQQLQAFPNAEIRYVLPVFMST
 QTPHVQTRLQEDALLNCDFSIDHHADVNIWVWLLKKGQEMKLLTYKGSEKSVVYHVKGI
 AMYVDEVPKGSASLVVKNVDLEKEGLYTCSVSVNSLFGDQVIHLEVVEPIVNLNVKSLL
 LTEGQEQKLVCESKYYPLDVNIEWLREGRDQTLPTVLKNVLYSSHKHNNDDGTYSLSF
 FIFTASQLDNGAVLTCRVEHMSLKHPKRSVKVRVEESPSYQEFFLVILIIIFLSVLLTL
 ILHLNRGRSSSKKETMTDKLAAFDANVSEMEFAQSECFYSF
 >Chicken TAPBPL Merged ESTs BU342879.1+ BU369515.1+ BX257449.3 (*Gallus gallus*)
 MSGDVLSLMAQRTLELPPCGHGSPQDGGGAQCSNHKDPKQHCFCIPRLRRILQLPCMFMEMK
 KNIPLSSTETVLLNIRLLLVPAGAGGAQRLQTESLPAGNLPSPFIVQEPSVNILQHTDED
 INMLDCRISPYFTANTQILWPGRKVKAHGLDTWFTCTIEHAAGKYTATAFLVQGHEDREH
 QTPGQLHQGIAEQTRVSAVLAARTRLPVRVTAALGKDVVLDCAFAADPRAAVAVRWALRKK
 GRHERYIATSGGRAEMFPQEPGRNASLLIRVELSDEGTICTVEAAALVLEQAIQLQIT
 EKPTVTNVNSLSLVEGEQKLVCDVRNYPADIHVQWLREPQSAGQLPDTVPNVLTSSH
 LRSSNGTYSFSRFFLLTATLRDNGHTYTCRVEHSSLQAPIRRSITVAVREATSTTWLLLL
 LLLGLTGCLVASLHHLHQVRSTTKPKPY
 >Turtle TAPBPL XP_005298961.1 tapasin-related protein-like [*Chrysemys picta bellii*]
 MNFRIYTAVYCFCLCTGILKHNGVATAGSTFRQSRRLSCLFETTETIPLLSEFTYIRQNA
 WLLLSGAERPHQENPPAADTLTFIVQESSLDILQHVEDTDKLECKINRYFTDNTQILW
 PGLQTQLHKLDSWFTGTIKHLADKFTVTVFLVQSSASKEAENRVPEDVSQGNAERFYLSG
 VFLVRTGPSLIQSGLNKDVLLSCAFSVDHHTDVTIKWVLQKQGGHKKLIFAYNGFTRQVE
 HEDKRAEMFLAELPKGNASLLLRVTGVRDAGTYSVSVSSLLWAQNIQLEIVEKPTVT
 NADVLSLVEGDEHKLVCIDISHFYPHDADAQWLQEPMEQGMPLHVVTHTVSSSHRHNDSGT
 YSFSSYFLLKASLRDDGRRYTCRVEHQSLKHPIKSLTVKVTESTSATWFLVILLVLLAG
 CLVVTLCYLYRGMGPNKPKPY
 >Kiwi TAPBPL XP_013817376.1 tapasin-related protein-like [*Apteryx australis mantelli*]
 MNYRTYLVAGCFLCIGIAKNLNGIATAGSSFRKSRQLACVFETSKIIPLTKELEIIRLNV
 RLLLSGTESEEAQRSHSENHPQDNIPSFFVKESSLDILQHANEDINTLDCRISAYFTANT
 QIIWPGREIRASSLDSWFICTIKHTAEKYMTTAFLVQTHKENENHSQGQLSPGIAEQSHV
 SAVFLVHTRPSVVRSAKSDVLFDCAFSIDHQADVTIQWVLHRKGGHKKLMFTYNGSSKQ
 VEHMADRAEMFLEEIPKGNASLLLRNVEMRDEGTYSVSVSSLTGEQAIQLQIEEKPTV
 IVNVNSLSLVEGEQYKLICDIRNYYPLDAQWLRELKGSRKVPDVVKNVLSNHRQSSN
 GTYSFSRYFLLTASLEDNEHTYTCRVDHQSLQAPIRRSVIVQVRGTSIIWVLLLLLGLSV
 CLAVVLHYFYKVKSTAKPKPY
 >Opossum TAPBPL XP_007485846.1: tapasin-related protein-like [*Monodelphis domestica*]
 MRWITLEGTRSPGNVLGMLGRGMLALRLLSPGLLWILGWHMVESGMTLGKAGQLPCLLEV
 ALDPSETLNGERKQWHIQLVLVSGSESQSLSHQGNATFLLQDRAGVLLPHMEDVDKLEC
 KVSRYFTANTQILWPGLPPRQAQLPTWHLVTISHTDAQFKASTFCFQSAESAQGTGPAIF
 SGVFSLYTKASRVQVELYGTVLLACSFVMDHAPEALDVRWALRQKGGKQREILKYDGKQR
 QVTHLHEQVEAFPSEIPGGDASIRLNNVAVNDQGDYFCVSTAGLYWELSIEVAVVEAPK
 VRFSPQTPILTLEEGEEELKLVCDVRHYFPLEAHVFWLRERLEGRMMPEGVKNVLFSSHRK
 NGDETFSSSYLLKASLQDDGFRYTCLVEHEGLRFPIKSIIRVTEKSGSSWGLLIIL
 IILMVVIFVLLRYLHQVKSMNKTTPY

Zebrafish gene sequences

Zebrafish (*Danio rerio*) Haplotype 19D amino acid gene sequences from McConnell et al.[6].

>ZebrafishUGA Q6PH40 NP_956879.1 Major histocompatibility complex class I
 UXA2 gene
 MQQVLLFLLGAHLAYAGTHSLRYFYTAVSGDIDFPEFTMVGLVDGGQFIYFDSKKMEAVP
 KTEWIRQNEGADYWDINTQRLIATHQAFKNNIQVAKERFNQSQGVHTFQVMYGCELEDDG
 STRGYWQYGYDGEDFLSLDKSTLTWTATKPAVITKNKWDADNADRQYTKSYLENECIEW
 VKKYVDYGKDTLERKDAPEVFMLQKDPSSPVVCQATGFYPSNIMMTWQKNKEEHFEDVDV
 GATLTNADGTQKQVTLSVKPEEWKNNKEAYRCVVQHVGAKNVDVITVKDIRSNEGSDNT

IAIIVGCVAALLAVIAGLIYWRRSNGYGKASSKDTDSEQSDPRAVEVK
 >ZebrafishPSMBF_19D NP_001017791.1
 MALMDVCLSELKQSCFTANAWQSEPTNCGPSHYTFGLNANELAVPTGMDPEQFLGPLTE
 GDSDIKIEFLHGTTTLAFKFQHGVMVAVDSTRASAGSYIDTKDFKKVIEINPYLLGTMSGN
 AADCYVWERRLAKECRIYKLRNKTRISVAAASKLLANMVAEYRGMGLSMGTMVCGWDQRG
 PGLYYVSSSGTRLAGDMFSTGSGCNYAYGVMDSGHRWDLTVPEAYDLAERAIYHATHRDA
 YSGGVVNMVYHMQKEGWIKVSQNDVGLHKKYVAERKK
 >ZebrafishPSMB9b_19D NP_571753.1
 MQNGGEHEGTNGGVGMGTIIIAVEFDGGVVVGSDSRVSAGASVNNRVMNKLSPHDKIYC
 ALSGSAADAQTIAEIVNYQLDVHSIEVEDDPLVCSAATLVKNISYKYKEELSAHLIVAGW
 DRREGGQVYATLSGLLTRQPFVGGSGSFYIYGFVDAEYRAGMTKKECQEFVINSLSLAM
 GRDGSSGGVAYLVTIDSESVEEKCILGNQLPTFYDPDTPVEPSKAVKV
 >ZebrafishPSMB13b_19D
 MFATVLKTPPEPGFCFENSLRNEALDSLKGGKPKALKTGTTIAGVVFKGDVVLGADTRAT
 SNKVVADKMCEKIHYPNIYCCGAGTAADTQKITELISSNLTIFSMNSGRNPRVIMAVS
 VLQDMLFRYHGQIGANLLGGVDCTGGHLYDIGPYGDMKVFPFLAMGSGNLPAMGILEDG
 FKSNDMLSAKHLVRDAIYAGVMNDLGSNGHNDLCVITRDGVYIRPYQESQHIKREGQ
 YKYKPGTTAVLTEKLHHIEMDLIEESVQKMETS
 >ZebrafishTAP2d_19D
 MELKMQHAKALLVDLAMSCLVHYSTNAFIRINSNVFLIAQLWLVSQVWIGLRCFLRGSW
 REAAVQRCVAVSCLCPMYESGQTLLINTQPENWSGCLSCPGKTIISAAATVLACLFWEV
 SFSDTNRKAESSESVEKKEKNRFLMRVVRYSKPDVALLSGAFVFLSLAVICDMCIPFYT
 GLVIDILGEHYQPNFSMAIFLMGLLSLSSGLRGGLFMCTLSRLNKRVRMLFNL
 VKQEIGFFEDKKTGDLTSRLSVDTKLMSQSVAMNVNILLRSLIKSVGILYLMLSLSWKLT
 LLTFIEAPLIAIAQKIYNTHYEQLSKDVQDSVARANETAGEAVAGVRTVRSFCMEPSEAG
 RYDERLTDTHNLKTRRDTVRAYVLLVRLKSLGMQVLMLYCGRQLIKSGAMSTGNLVSFI
 LYQGDLSGYIARTLVYMSDMLNSVGAAKVFYEYLDKPLVDTDGSLHPKTLAGQVHFKNL
 TFFYPSRPDQAALKDFSLKPGQMTALVGMSSGGKSTCVSLERFYQPQQGQILLDQGP
 LQKYQHLYLHQKVMVQDPVLFSGSVRDNIAYGLTDYDQKKVEDAAKEANAHDFICRLE
 KTYDTDVGERGCLLSAGQKQRIAIARALIRQPQILILDEVSSSLDTESEKMQDALARRP
 NQTLVLIAHRLKTIERADQIIIVIDKGEVLETGTHQELMERKGNYYKLRERLFSEEKEADE
 KEIKQG
 >ZebrafishTap2e_19D
 METADSRHLKTIIFRLSLSRPDLLPICGAFLFLVAAVIGDMFMPAFTGKIIDALNSTFDH
 GTFTSAIFFMGLTSLGGSFSAGCRGGLFMILTISRLTKRVRERLFSFVKQDIAFFEKTTK
 GDLTTCSSDASLMSRSLAANVNILLRSLIMTIGIYCFMIQLCWPLALLSAFESPITITA
 EKIYNKYQGLVSSVKKSIVKSNQVAGESVYNIRAVRSQGAETVEQRLYDTSLEETHYFK
 IHRDSVRATYLLFMRLFQLCMRVLILWYGHQMITSGQMTPGNLVSFIFYQMEIGGHVQTL
 VELHIDLIESLESADKVFEMDHKPSVLSGDLAPDQLKGHVMFKNISFSYSSCPDKKVLE
 NVSFEMKPGTITALVGVSOGGKSTCVALLKHLYPEQSGDIFLDGRPLKEYDPKYFHQKVA
 VVSQEPPELFARSIQKNITYGLDSYSEDMVQTAVKQANINSFVLSLEDGYNTEVGERGGHL
 SVGEKQRIAIARALIRQPQVLILDEVTSLLDTENEKMIQNALACFATQTRLIVAHRLKTI
 EKADQIIIVIDKGSVSEKGTVELMKKKGIYYKLRESRFNEDSLET

Zebrafish (*Danio rerio*) haplotype 19B amino acid gene sequences from McConnell et al.[6].

>Zebrafish PSMB8a_19b NP_571467.3 D. rerio
 MALLDVSGYKYNASQFGFKQTLDRSNHYSFGTKCQEFVAVPVGVDPKFLKSCSCEDGV
 CIDLNHGTTTTLAFKFRHGVIIVAVDSRASAGKIASKEANKVIEINPYLLGTMSGSAADCQ
 YWERLLAKECRLYKLRNKQRIISVSAASKLLSNMMLGYRGMGLSMGSMICGWDKQGPGLYY
 VDDNGTRLSGRMFSTGCGNSYAYGVVDVSGYREDMTVEEAYELGRRGIAHATHRDAYSAGGV
 VNLYHMQEDGWIKVCKEDVSELIHRYKKGMF
 >Zebrafish PSMB9a_(19b) NP_571466.1 D. rerio
 MSEELFPEPGWLSEEVKTGTIIIAVTFDGGVVIGSDSRVSAGESVNNRVMNKLSPHDKI
 YCALSGSAADAQTIAEIVNYQLDVHSIEVEDDPLVCSAATLVKNISYKYKEELSAHLIVA
 GWDKKGQVYATLSGLLTQKQPFVGGSGSFYINGFVDAEYKKNMTKRECQEFVNNALTL
 AMGRDGSSGGVAYVVTIDKDGTEEKCVLGNELPKFFDE
 >Zebrafish PSMB12_(19b) NP_571751.1 D. rerio
 MDRHHPYSQVNGVSTGTITLAVKFNGGVIIIGSDSRASMGESYVSSKTINKLIQVHDRIFC

CIAGSLADAQAVTKMAKFQLSFHSIQMESPLVKAASIMRELCYSNKEELRAGFITAGW
 DRKKGPQIYVVSLLGMLLSQPFTIGGSGSTYIYGYVDAKFKPDMTLEEATQFSTNALALA
 MGRDNVSGGVVHLVVITEAGVKHIVVPGDELPHFDE
 >Zebrafish PSMB13a (19b) NP_571752.1 D. rerio
 MALTSHVLEPSLCGFNFENATRNIVLENGAEEGKIKPPKALKTGTTIAGVVFKDGVVLGA
 DTRATSDEVVADKMKAKIHYYIAPNIYCCGAGTAADTEKTTDMLSSNLTIFSMNSGRNPRV
 VMAVNIIQDMLFRYRGMIGANLILGGVDCTGSHLYTVGPYGSMDKVPYLAMGSGNLAAMG
 ILEDRFKVNMDLQAKALVSDAIQAGIMCDLGSNNIDLCVITKEGVYIRPHKESPYNY
 KRQAKYKYKSGTTPILTCTVKNLELDLVHETVQMMETAASS
 >ZebrafishTAP2a (19b) NP_001006594.1 D. rerio
 MRKVLVFAFCLCFDILIVLILDLAETLLKSPTSSLDYNLFRQWAESGLRCTLTYAGSTLS
 KDAADPLIRRWITVHCFIGSVYETGRLAMFNRYRLDKFTVWLVGTVAALACLFWEITLP
 DTNEESNGKERKQKARVLFIRVIRLYRPDYILLFGAFVFLALAVLCMFIPLYTGEVIDI
 LGSHYQWDNFRSAIIFMGLFSLGSSFSAGCRGGLFMCAINSFTCRVKVQLFGSLIRQDIG
 FFETIKTGDISRLSTDITLMGRAVALNVNVLRLTLVKTGLMGLYLMVSLSWKLTLLMLME
 TPLTGLLQNIYDTHYQKLSKEVQDSMAQANDAAGEAVSGIRTVKSFKTELGEAHRYDGRL
 METHNLKTRRDTVRAIYLLIRRMTELGMKVAMLYYGRFLFIQYQGMSTGNLVSFILIYQDDL
 GDNIRTLIYIFGDMLSVGAAGKVFEYQDRKSEVSIDGNLMPKDLKGHVKFQKLTFSYPR
 RPDHNVLKDFSLELKPQGITLVGMSGGKSTCVSLLERFYQPQQGTILLDGKPLQDYQH
 KYLHSHKAMVGPVLFSGTVRDNIAIYGLQGCSMERVKEAASKANAHSFISKLEKGYDTD
 VGERGNLLSGGEKQRIAIARALIREPQVLILDEVTSLLDTESEQMVQQALSCCPTQTLLV
 IAHRLKTIERADQIVVIDSGELVEKGTHEELMEKKGSYYKLRERLFSDDKTTKQEKESD
 TVKTQ

Other zebrafish deduced amino acid sequences from McConnell et al.[6]

>Zebrafish TAPBP_CG2 (Danio_rerio) TSA GDQH01003123.1
 MSDISTVFKISVIAFTLFGHYGSACPVLCEWVQEKPGHGGGLSTPMSQEKSLMFIRTE
 AYSEETKAELHPPADISSSRVYVTDPAFTCSAALNPPKGSVNKPKCEINPFMPHASMV
 RWASALTDASQSPVYLQADWFSVAAQGLDEQLTSLNIMRAPASKEPEVILSVSSKTPVV
 RCRLGEPVLLDCGFWDPSPLHSGSFSIEWRYQFRGEGRLVLAYDGKNDRAETSESGA
 EMDITGLYQTGNASLILEESQVRHSGTYICTVYLPHELLAQVAVDLEIVEPPSLSLPSPL
 PLLVPGQVLSVQCEASGFAPHTLDLGEFTGADGKSLSLGQGSVTGHRRASDGTFSQSSR
 LELEDSTKLRLARGGEISCAVHGGGTRRASAALNVIGVAAPSIEDSMAMVAVALLLYGM
 KFLSWTFSSSDSGDSELNDKKEK*
 >Zebrafish Psmb7_21 NP_001039029.2 ENSDARG00000037962 Chr.21: 8.249.235-
 8.259.556
 MATVSVCQYQPGGFSFENCRRNALLEADITKLGFSSPAARKTGTTICGIVYKDGVLGAD
 TRATEGMIVADKNCISKIHYISPNIYCCGAGTAADTEMTTQIISSNLELHSLSTGRLPRVA
 TANRMLKQMLFRYQGYIGAALVLGGVDCTGPHLYSIYPHGSTDKLPYVTMGSGSLAAMAV
 FEDRYRPDMEEEDAKSLVRDAIAAGIFNDLGSNSNIDVCVITKGKVDYLRPHDIANKKGV
 RTGSYRYKHGTTGVLSKAVTPLNLDVVEESVQTMDS
 >Zebrafish PSMB7/10_4 Chr.4 NP_001315481.1 Chromosome 4: 75.630.955-
 75.639.988
 MLNTSTKTLTGGSFENFTRNAVLEANLSEKGYSAFNARKTGTTIAGLVFKDGVILGADT
 RATDDMVVADKNCMKIHYIAPNIYCCGAGVAADAETVTTQMSSNVELHSLSTGRPPLVAM
 VTRQLKQMLFRYQGHIGSSLIVGGVDVNGAQLYSVYPHGSYDKLPFLTMGSGAASAI SVF
 EDRYKPNMELEEAKQLVRDAITAGIFCDLGSNSNIDLCVITDKKVDYLRTYDQPVHKNQR
 GGTTRYKPGTTAVLSKTVTPLTLDVVDESVMVMDTE

Previously published Medaka (*Oryzias latipes*) sequences

Sequences originate from HdrR[7], HN1 [8], cab [9].

>HdrR_PSMB9 BA000027:216010-218097 BAB83845.1
 MLGEAEPQWISEEVKTGTIIAIEFNNGGVVLGSDSRVSAGDSVNVNRMNKLSPHDKIYC
 ALSGSAADAQTIAEMVNYQLDVHSLIEDDPQVRSAAATLVKNISYKYKEELSAHLIVAGW
 DRRDGQVVFATLGGLLTRQPFPAIGSGSSVYGFVDAEYRRGMTKEECQKFVVNTLALAM
 NRDGSSGGVAYIVTIDEHSTDEKIVILGNLPTFFDQ
 >HdrR_PSMB12 BA000027:222454-226075 BAB83846.1

MEKHFTDSRVKGVSTGTTILAAVFDGGVVIGSDSRASIGGEYVSSKTINKVIQVHDRIFC
CMAGSLADAQAVTKTAKFQLSFHSIQMESPLVISAASVLKQLCYNNKEELQAGFITAGW
DKKKGPQVYVVSLLGMMISQPVTIGGSGSTYIYGVDKAKYKVNMTREECLQFATNALALA
MGRDNVSGGVANLVVITKEGVEHIVIPGDKLPRFSDE
>HdrR_PSMB13 BA000027:226185-228725 BAB83847.2
MALSNVLDSPAAGFNFDNAARNAAFEGLFEGGQTPKPLKTGTIAGVVFKDGVVLGADTR
ATSSEVVADKMKCAKIHYSNPNIYCCGAGTAADTEKTTDLLSSNLTVFSLNSGRNPRVMA
VNILQDMLYRYHGQIGANLILGGVDCTGNHLYTVGPYGSVNKVPYLAMGSGDLAALGILE
DRFKHDLELEKAKELVRDAIHAGIMSDLGSGNNIDICVITKQGVVDYIRPFQESEYKETRK
PKYKYRPGTTPVLTKKVPLKLEVVEEIQQMDTA
>HdrR_PSMB8A BA000027:230073-234656 BAB83848.1
MALAAVCGGQSSSEHFGQLFSGKQARLFDRPNHFSFGTKIQEFAPVPGNEPSGFLRSCNR
EEGVRIDLNHGTTTLAFKFRHGVIIVAVDSRASAGNYLASNDVNKVEINPYLLGTMSGSA
ADCQYWERLLAKECRLYRLRNNHRISVAAASKLLCNMMLGYRGMGLSVGSMICGWDKEGP
GLYYVDDNGTRLSGRMFSTGCGNSYAYGVVDSGYKEDMTVEEAYELGCRGIAHATHRDAY
SGGSVNMYHMREDGWIKVCKEDVSELIHRYREGMF
>HN1_PSMB9 AB183488:235856-237954 BAD93261.1
MLGEEAEPQWISEEVKTGTIIAIEFNNGGVVLGSDSRVSAGDSVVNRVMNKLSPHDKIY
CALSGSAADAQTIAEMVNYQLDVHSLEIDEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDRRDGGQVFATLGGLLTRQPFAIGGSGSSYVYGFVDAEYRRGMTKEECQKFVVNTLALA
MNRDGSSGGVAYIVTIDEHSTDEKVLGNDLPTFFDQ
>HN1_PSMB12 AB183488:241217..245420 BAD93262.1
MEKHFTDSRVKGVSTGTTILAAVFDGRGVV-SDSRASIGGEYVSSKTINKVIQVHDRIFC
CMAGSLADAQAVTKTAKFQLSFHSIQMESPLVISAASVLKQLCYNNKEELQAGFITAGW
DKKKGPQVYVVSLLGMMISQPVTIGGSGSTYIYGVDKAKYKVNMTREECLQFATNALALA
MGRDNVSGGVANLVVITEAGVEHIVIPGDKLPRFNDE
>HN1_PSMB13 AB183488:245530-251792 BAD93263.1
MALSNVLDSPAAGFNFDNAARNAAFEGLFEGGQTPKPLKTGTIAGVVFKDGVVLGADTR
ATSSEVVADKMKCAKIHYSNPNIYCCGAGTAADTEKTTDLLSSNLTVFSLNSGRNPRVMA
VNILQDMLYRYHGQIGANLILGGVDCTGNHLYTVGPYGSVNKVPYLAMGSGDLAALGILE
DGFQHDLELEKAKKLVRDAVHSGIMNDLGSGNNIDICVITAEGVDYIRPFQSESEFKDRRE
IKYKYRQGTTPVLTKKITPLKMEVVQETTQKMETV
>HN1_PSMB8A AB183488:253199-259873 BAD93264.1
MALANVCGLKNYSEQSGQMFASTRRLIDRPNHYSFGTKIQEFAPVPGEEPVGFLKSCNTEG
GARFELHHGTTTSLFKFKHGVIIVAVDSRASTGSYIATCEYNKVEINPYLLGTMSGSAAD
CKYWERLLAKECRLYRLRNNHRISVAAASKLLCNMMLGYRGMGLSVGSMICGWDKEGPGI
YYVDDNGNRLSGRMFSTGSGSNYAYGVLD SGYKEDMTVEEAYELGRRGIVHATHRDSYSG
GVVNMYHIQEDGWIKVCKDDVSELLHHYKKGMF
>cab_PSMB9 AB450971:56863-59061 BAJ07258.1
MLGEEAEPQWISEEVKTGTIIAIEFNNGGVVLGSDSRVSAGDSVVNRVMNKLSPHDKIY
CALSGSAADAQTIAEMVNYQLDVHSLEIDEDPQVRSATLVKNISYKYKEELSAHLIVAG
WDRRDGGQVFATLGGLLTRQPFAIGGSGSSYVYGFVDAEYRRGMTKEECQKFVVNTLALA
MNRDGSSGGVAYIVTIDEHSTDEKVLGNDLPTFFDQ
>cab_PSMB12 AB450971:63356-67035 BAJ07259.1
MEKHFTDSRVKGVSTGTTILAAVFDGGVVIGSDSRASIGGEYVSSKTINKVIQVHDRIFC
CMAGSLADAQAVTKTAKFQLSFHSIQMESPLVISAASVLKQLCYNNKEELQAGFITAGW
DKKKGPQVYVVSLLGMLISQPVTIGGSGSTYIYGVDKAKYKVNMTREECLQFATNALALA
MGRDNVSGGVANLVVITEEGVEHIVIPGDKLPRFSDE
>cab_PSMB13 AB450971:67055-69667 BAJ07260.1
MALSNVLDSPAAGFNFDNAARNAAFEGLFEGGQTPKPLKTGTIAGVVFKDGVVLGADTR
ATSSEVVADKMKCAKIHYSNPNIYCCGAGTAADTEKTTDLLSSNLTVFSLNSGRNPRVMA
VNILQDMLYRYHGQIGANLILGGVDCTGNHLYTVGPYGSVNKVPYLAMGSGDLAALGILE
DRFKHDLELEKAKELVRDAIHAGIMSDLGSGNNIDICVITKQGVVDYIRPFQESEYKETRK
PKYKYRPGTTPVLTKKVPLKLEVVEEIQQMDTA
>cabPSMB8 AB450971:70934-75613 BAJ07261.1
MALAAVCGGQSSSEHFGQLFSGKQARLFDRPNHFSFGTKIQEFAPVPGNEPSGFLRSCNR
EEGVRIDLNHGTTTLAFKFRHGVIIVAVDSRASAGNYLASNDVNKVEINPYLLGTMSGSA
ADCQYWERLLAKECRLYRLRNNHRISVAAASKLLCNMMLGYRGMGLSVGSMICGWDKEGP
GLYYVDDNGTRLSGRMFSTGCGNSYAYGVVDSGYKEDMTVEEAYELGCRGIAHATHRDAY
SGGSVNMYHMREDGWIKVCKEDVSELIHRYREGMF

References

1. Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, Koop BF, Grimholt U: **Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon.** *BMC genomics* 2010, **11**:154.
2. Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS *et al*: **Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*).** *BMC genomics* 2007, **8**:251.
3. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A *et al*: **The Atlantic salmon genome provides insights into rediploidization.** *Nature* 2016, **533**(7602):200-205.
4. Shiina T, Dijkstra JM, Shimizu S, Watanabe A, Yanagiya K, Kiryu I, Fujiwara A, Nishida-Umehara C, Kaba Y, Hirono I *et al*: **Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry.** *Immunogenetics* 2005, **56**(12):878-893.
5. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, Lemon C, Bird NH, Koop BF: **The genome and linkage map of the northern pike (*Esox lucius*): conserved syntenies revealed between the salmonid sister group and the Neoteleostei.** *PloS one* 2014, **9**(7):e102089.
6. McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, Yoder JA, Andrade J, de Jong JL: **Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution.** *Proceedings of the National Academy of Sciences of the United States of America* 2016, **113**(34):E5014-5023.
7. Matsuo M, Asakawa S, Shimizu N, Kimura H, Nonaka M: **Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*).** *Immunogenetics* 2002, **53**:930-940.
8. Tsukamoto K, Hayashi S, Matsuo M, Nonaka M, Kondo M, Shima MI, Asakawa S, Shimizu N, Nonaka M: **Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*.** *Immunogenetics* 2005, **57**:420-431.
9. Nonaka MI, Nonaka M: **Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization.** *Immunogenetics* 2010, **62**(5):319-332.

Additional file 3. Text S2. Alignment of deduced PSMB8 amino acid sequences.

Sequences can be found in Additional file 2: Text S1 and phylogeny of bona fide sequences including accession numbers is shown in main text Figure 2. Aligned sequences are as follows: OnmyPSMB8F= rainbow trout genome Ia region, SasaPSMB8F= Genbank ACI66984.1, DarePSMB8F_19D= NP_001017791.1, SasaPSMB8a_#C =Atlantic salmon genome Ia region, SasaPSMB8a_#A=Atlantic salmon Ia region BAC 92I04, SasaPSMB8a_#B=Atlantic salmon Ia region BAC 714P22, OnkiPSMB8a= coho salmon genome Ia region, SasaPSMB8b_#A= Atlantic salmon Ib region BAC 8I14, SasaPSMB8b_#B= Atlantic salmon Ib region BAC 439J08, OnmyPSMB8a/omPSMB8aψ= rainbow trout BAC and genome, SasaPSMB8b_#C= Atlantic salmon genome Ib region, OnmyPSMB8b_#B= rainbow trout genome Ib region, OnmyPSMB8b_#A= rainbow trout Ib region BAC, OnkiPSMB8b= coho salmon genome Ib region, medaka OrlaPSMB8_Hd-rR= BAB83848., medaka OrlaPSMB8_HN1= BAD93264.1, medaka OrlaPSMB8_cab= BAJ07261.1, EsluPSMB8_NW71=Northern pike genome unplaced scaffold NW_017859271.1, EsluPSMB8_NW80= Northern pike genome unplaced scaffold NW_017859580.1, EsluPSMB8 = Northern pike genome chr.10, DarePSMB8a_19b= NP_571467.3, OnkiPSMB8b2/3=coho salmon genome, OnmyPSMB8bψ=rainbow trout genome Ib region, HosaPSMB8= NP_683720.2.

```

      *      20      *      40      *      60      *
OnkiPSMB8a      : MALFDVSG---YKSHSEL--RGQII GTG-VGHFIDRPNQQFSPVGVGD--PSGFLKSCSREG--GVSIDLNH : 62
OnmyPSMB8a_#Aψ  : ----- : -
OnmyPSMB8a_#Bψ  : ----- : -
SasaPSMB8a_#A   : .....--..Y...--.....K..... : 62
SasaPSMB8a_#B   : .....--..Y...--.....K..... : 62
SasaPSMB8a_#C   : .....--..Y...--.....K..... : 62
SasaPSMB8b_#A   : .....--..AG..--...L...LV...E.A..... : 62
SasaPSMB8b_#B   : .....--..AG..--...L...LV...E.A..... : 62
SasaPSMB8b_#C   : .....--..AG..--...L...LV...E.A..... : 62
SasaPSMB8b_#Cψ  : -----M- : 2
OnmyPSMB8b_#A   : .....--..AG..--...L...LV...E.A..... : 62
OnmyPSMB8b_#B   : .....--..AG..--...L...LV...E.A...VS... : 64
OnmyPSMB8b_#Bψ  : ----- : -
OnkiPSMB8b      : .....--..AG..--...L...LV...E.A..... : 62
OnkiPSMB8b2     : ----- : -
OnkiPSMB8b3     : ----- : -
EsluPSMB8_NW71  : ..I.C...--..N...Q...E.T.K...SVLSL--.....Y.H..QG.G..... : 52
EsluPSMB8_NW80  : .....--..L...--..KTL...LRHFNFGTT..E.A..A.M.--.....WRH..QG..... : 65
EsluPSMB8       : .....--..L...--..NTL...LRHLNFGTT..KE.A..A.M.--.....H..QG..... : 65
OrlaPSMB8_HdrR  : ...AA.C.GQSSSE.FGQLFS.KQARLFD RPNHFSFGTKIQEFA.P.GNE.....R..N..E--..R..... : 70
OrlaPSMB8_HN1   : ...AN.C.--LKNYSEQSGQMFATRRLID RPNHYSFGTKIQEFA.P.GEE.V.....NT..--ARFE.H.. : 68
OrlaPSMB8_cab   : ...AA.C.GQSSSE.FGQLFS.KQARLFD RPNHFSFGTKIQEFA.P.GNE.....R..N..E--..R..... : 70
DarePSMB8a      : ...L...YK-.N.A.QFGFKQTLLDRS-NHYSFGTKC.E.A.....--..K.....C.D--..C..... : 66
OnkiPSMB8a.F    : ----- : -
OnmyPSMB8Fa_#B  : ...L..C.IS-DWVKEDGFNAERASVDKVNHFNFAAQTPELA.....PAEFLRPLVD.E..VD..K.N.E. : 71
SasaPsmb8F      : ...L..C.IS-DWVKEDGFNAERASVDKVNHFNFKFAAQTPELA.....PAEFLRPLVD.E..VD..K.N.E. : 71
EsluPSMB8F      : ----- : -
Dare8F_19D      : ...M..CLSS-ELKQ.CFTANAWQSEPTNC.PSHYTFGLNANELAVPTGMDPEQFLGPLT..DSDIK.EFL. : 71
SasaPSMB8Fa_#A  : ----- : -
SasaPSMB8Fa_#B  : ----- : -
SasaPSMB8Fb_#A  : ----- : -
SasaPSMB8Fb_#B  : ----- : -
HosaPsmb8A      : ...L..C.APRGQRPESALPVAGSGRRSD-PGHYSFSMRSP ELALPRGMQ.TE.FQ.LGGD.ERN.Q.EMA. : 71

```

		80	*	100	*	120	*	140	
OnkiPSMB8a	:	GTTTLA	----	FTFRHGVIVAVDSRASAGSYIESKEANKVIEINPYLLGTMSGSAADCQYWERLLAKECRL	:	128			
OnmyPSMB8a_#Aψ	:	-----							1
OnmyPSMB8a_#Bψ	:	-----							1
SasaPSMB8a_#A	:		-----		A				128
SasaPSMB8a_#B	:		-----		A				128
SasaPSMB8a_#C	:		-----		A				128
SasaPSMB8b_#A	:		-----		A				128
SasaPSMB8b_#B	:		-----		A				128
SasaPSMB8b_#C	:		-----		A				128
SasaPSMB8b_#Cψ	:	---,FD	---	DLTWD	..KSLFIFKT	---	L..T..I..H		60
OnmyPSMB8b_#A	:		---	K		A			128
OnmyPSMB8b_#B	:		---	K		A			130
OnmyPSMB8b_#Bψ	:	-----							1
OnkiPSMB8b	:		---	K		A			128
OnkiPSMB8b2	:	-----				AL			39
OnkiPSMB8b3	:	-----							-
EsluPSMB8_NW71	:		QFRHGV	LT		V	A		124
EsluPSMB8_NW80	:		---	K		A			131
EsluPSMB8	:		---	K		A			131
OrlaPSMB8_HdrR	:		---	K		N.LA	NDV		136
OrlaPSMB8_HNI	:		S	---	K.K		T	ATC.Y	134
OrlaPSMB8_cab	:		---	K		N.LA	NDV		136
DarePSMB8a	:		---	K		K	A		132
OnkiPSMB8aF	:	-----							1
OnmyPSMB8Fa_#B	:		---	K.Q	..M		VSTQMFK	F	137
SasaPsmb8F	:		---	K.Q	..M		VSTQMFK	F	137
EsluPSMB8F	:	-----							-
DarePSMB8F_19D	:		---	K.Q	..M		DT.DFK	N	137
SasaPSMB8Fa_#A	:	-----							-
SasaPSMB8Fa_#B	:	-----							-
SasaPSMB8Fb_#A	:	-----							-
SasaPSMB8Fb_#B	:	-----							-
HosaPsmb8A	:		---	K.Q	..A		SALRV	C	137
		*	160	*	180	*	200	*	
OnkiPSMB8a	:	YKLRNKQRISVSAASKLLCNMMLGYRGMGLSMGSMIVGWDNKGPGLYYVDDNATRLSGRMFSTGCGSSYAYG	:	200					
OnmyPSMB8a_#A	:		R						73
OnmyPSMB8a_#Bψ	:		R						73
SasaPSMB8a_#A	:								200
SasaPSMB8a_#B	:								200
SasaPSMB8a_#C	:								200
SasaPSMB8b_#A	:				I				200
SasaPSMB8b_#B	:				I				200
SasaPSMB8b_#C	:				I				200
SasaPSMB8b_#Cψ	:					VQ	GRE	-----	108
OnmyPSMB8b_#A	:								200
OnmyPSMB8b_#B	:								202
OnmyPSMB8b_#Bψ	:			R		VQ	-----		46
OnkiPSMB8b	:								200
OnkiPSMB8b2	:								70
OnkiPSMB8b3	:	-----				VQ	-----		31
EsluPSMB8	:		K		S		I		203
EsluPSMB8_NW71	:	E		K		S	-----		150
EsluPSMB8_NW80	:		K		S		R		203
OrlaPSMB8_HdrR	:	R	NH	..A		V	C	KE	208
OrlaPSMB8_HNI	:	R	NH	..A		V	C	KE	206
OrlaPSMB8_cab	:	R	NH	..A		V	C	KE	208
DarePSMB8a	:			S		C	KQ		204
OnkiPSMB8aF	:	IYKLRNE		T	A	VVN		RG-T	67
OnmyPSMB8Fa_#B	:		E		A	VVN		T	209
SasaPsmb8F	:	N		E		A	VVN		209
EsluPSMB8F	:	-----				G		A	63
DarePSMB8F_19D	:		T		A	VAE		T	209
SasaPSMB8Fa_#A	:	-----							-
SasaPSMB8Fa_#B	:	-----							-
SasaPSMB8Fb_#A	:	-----							-
SasaPSMB8Fb_#B	:	-----							-
HosaPsmb8A	:	Y		GE		S	CQ		209

```

      220          *          240          *          260          *          280
OnkiPSMB8a      : VIDSGYREDMTVEEAYELGRRGITHATHRDAYSGGVVNLYHMQEDGWIKVCKEDVSELIHRYRKGMF- : 267
OnmyPSMB8a_#Aψ : .....E.....- : 134
OnmyPSMB8a_#B : .....- : 140
SasaPSMB8a_#A : .....- : 267
SasaPSMB8a_#B : .....- : 267
SasaPSMB8a_#C : .....- : 267
SasaPSMB8b_#A : .V.....- : 267
SasaPSMB8b_#B : .V.....- : 267
SasaPSMB8b_#C : .V.....- : 267
SasaPSMB8b_#Cψ : ----- : -
OnmyPSMB8b_#A : .M.....- : 267
OnmyPSMB8b_#B : .M...H.....V.....- : 269
OnmyPSMB8b_#Bψ : ----- : -
OnkiPSMB8b      : .M.....- : 267
OnkiPSMB8b2     : ----- : -
OnkiPSMB8b3     : ----- : -
EsluPSMB8_NW71  : ----- : -
EsluPSMB8_NW80  : .....H.K.....- : 270
EsluPSMB8       : .....- : 270
OrlaPSMB8_HdrR  : .V...K.....C...A.....S...M...R.....E.....- : 275
OrlaPSMB8_HN1   : .L...K.....V.....S...M...I.....D...L.H.K.....- : 273
OrlaPSMB8_cab   : .V...K.....C...A.....S...M...R.....E.....- : 275
DarePSMB8a      : .V.....A.....K.....- : 271
OnkiPSMB8a.F    : MM...SQY.LS.P...D.AQ.A.F.....T.....- : 105
OnmyPSMB8Fa_#B  : .M...Y.LS.P...D.AQ.A.F.....T...V...R.T...SQ...GD.Y...FSNEKK- : 276
SasaPsm8F       : .M...Y.LS.P...D.AQ.A.F.....T...M...R.T...SQ...GD.Y...FYNDKK- : 276
EsluPSMB8F      : .M...Y.LS.P...D.AQ.A.F.....T...M...K.T.V...SQ...GD.Y...FNNENK- : 130
DarePSMB8F_19D  : .M...H.W.L.P...D.AE.A.Y.....M...KE...SQN...GD.H.K.VAERKK : 277
SasaPSMB8Fa_#Aψ : -----KRNT...SQ...GD.Y...FYNDKK- : 27
SasaPSMB8Fa_#Bψ : -----KRNT...SQ...GD.Y...FYNDKK- : 27
SasaPSMB8Fa_#Cψ : -----KRNT...SQ...GD.Y...FYNDKK- : 27
SasaPSMB8Fb_#Aψ : -----R.T...N.SQ...AD.Y.HLYNEKK- : 27
SasaPSMB8Fb_#Bψ : -----R.T...N.SQ...AD.Y.HLYNEKK- : 27
SasaPSMB8Fb_#Cψ : -----R.T...N.SQ...AD.Y.HLYNEKK- : 27
HosaPsm8a       : .M...PNLSP...D...A.AY...S...M...K...V...EST...D.L.Q...EANQ- : 276

```

Additional file 4: Figure S2. Teleost MHCI data

Table of Contents		Page
S2a	Phylogeny of MHCI alpha 1 domain amino acid sequences from selected species.	2
S2b	Phylogeny of deduced MHCI alpha 2 domain sequences from selected species.	3
S2c	Phylogeny of deduced MHCI alpha3 domain sequences from selected species.	4
S2d	Alignment of selected MHC class I amino acid sequences	5

Figure S2a. Phylogeny of MHCI alpha 1 domain amino acid sequences from selected species.

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (1). The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Red font shows MHCI genes linked to PSMB8F sequences and colour boxes show alpha 1 domain lineage categories (3). Sequence references not shown in figure can be found in Additional file 2: Text S1.

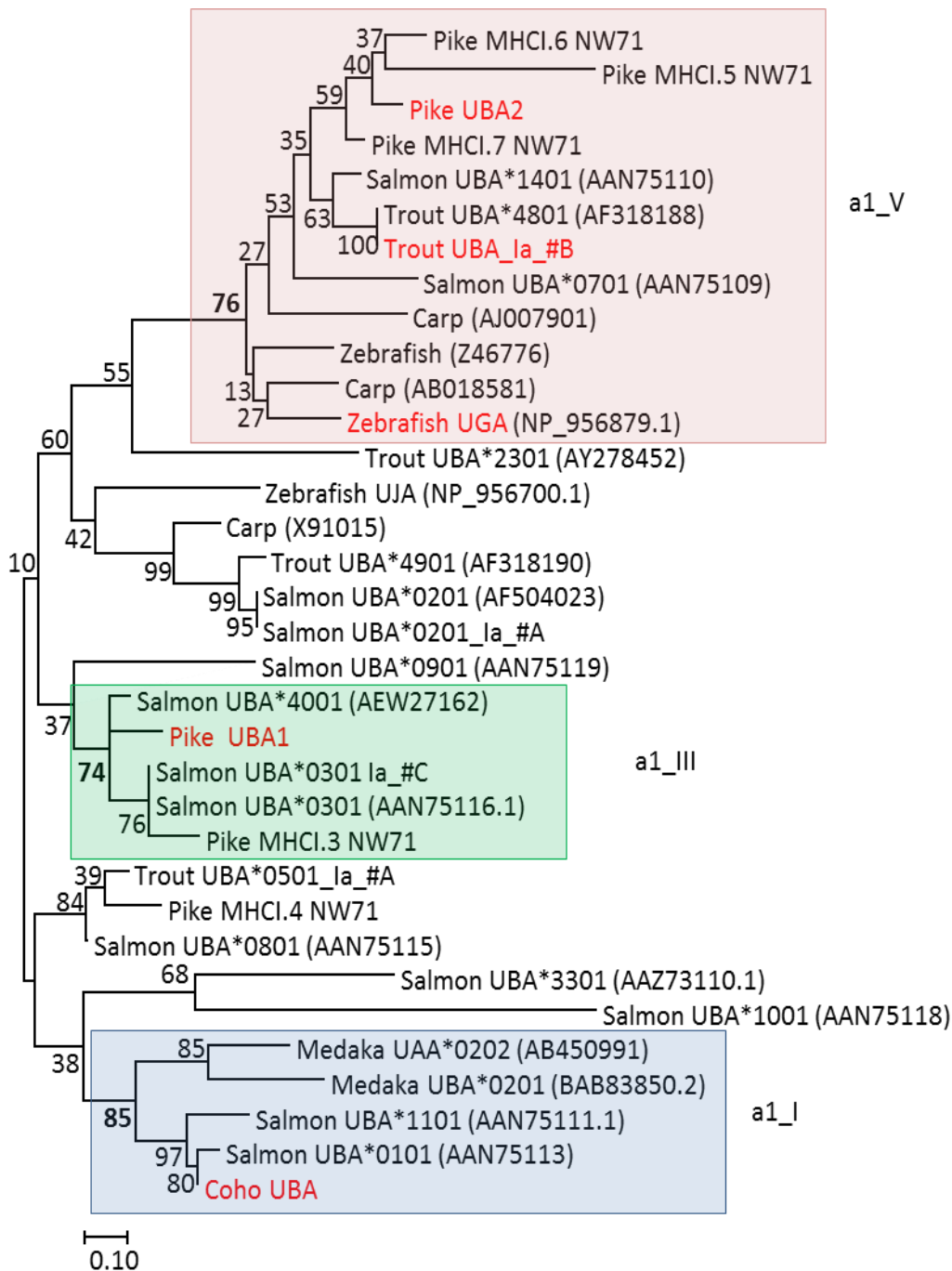


Figure S2b. Phylogeny of deduced MHCI alpha 2 domain sequences from selected species.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. w/freq. model (1). The percentage of trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Red font shows MHCI genes linked to PSMB8F sequences and colour boxes show alpha 2 domain lineage categories (3). Sequence references not shown in figure can be found in Additional file 2: Text S1.

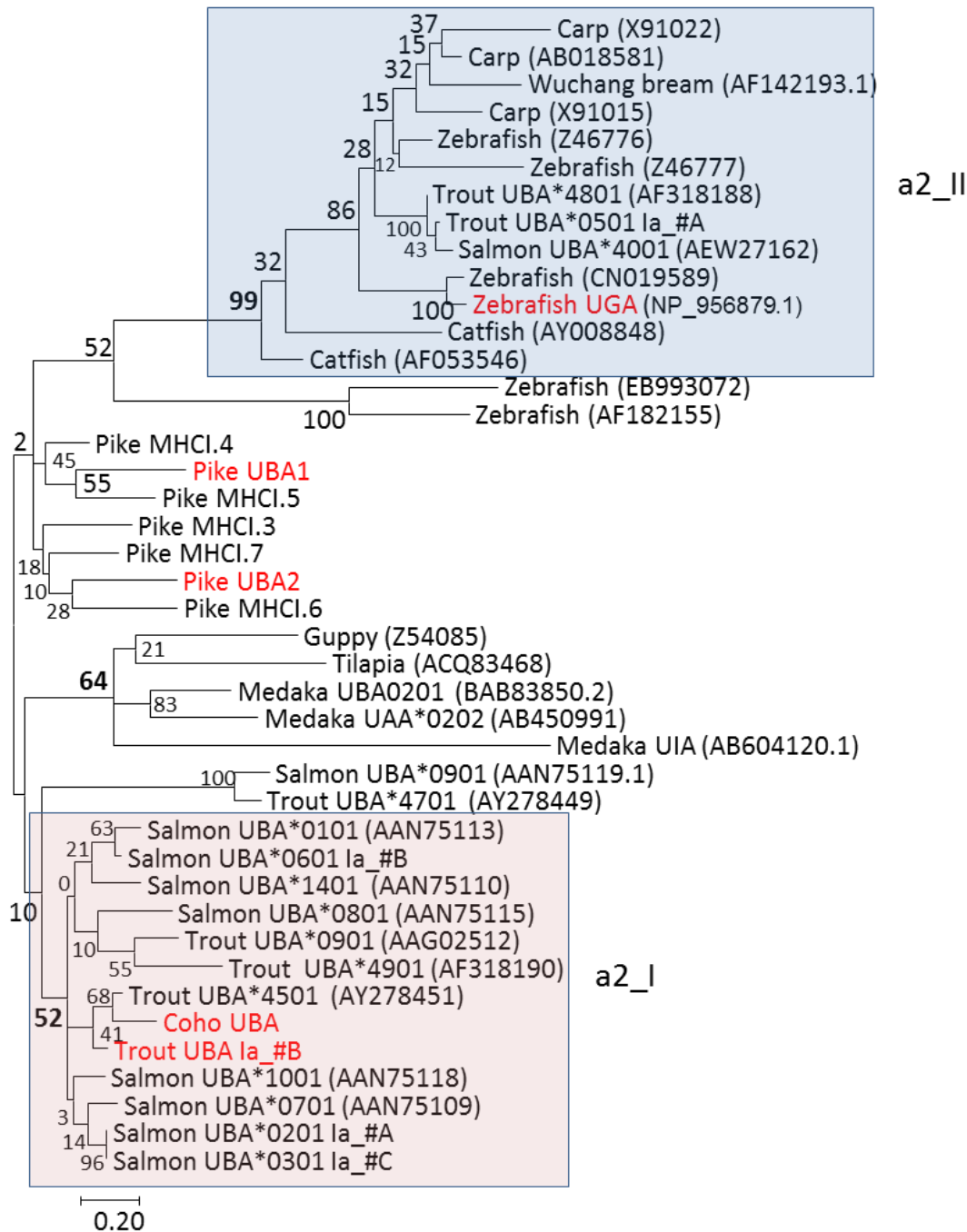


Figure S2c. Phylogeny of deduced MHCI alpha3 domain sequences from selected species.

Deduced MHCI alpha 3 domain sequences from selected species were used in the unrooted phylogenetic tree. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. w/freq. model. The percentage of trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Red font shows MHCI genes linked to PSMB8F sequences. Sequence references not shown in figure can be found in Additional file 2: Text S1.

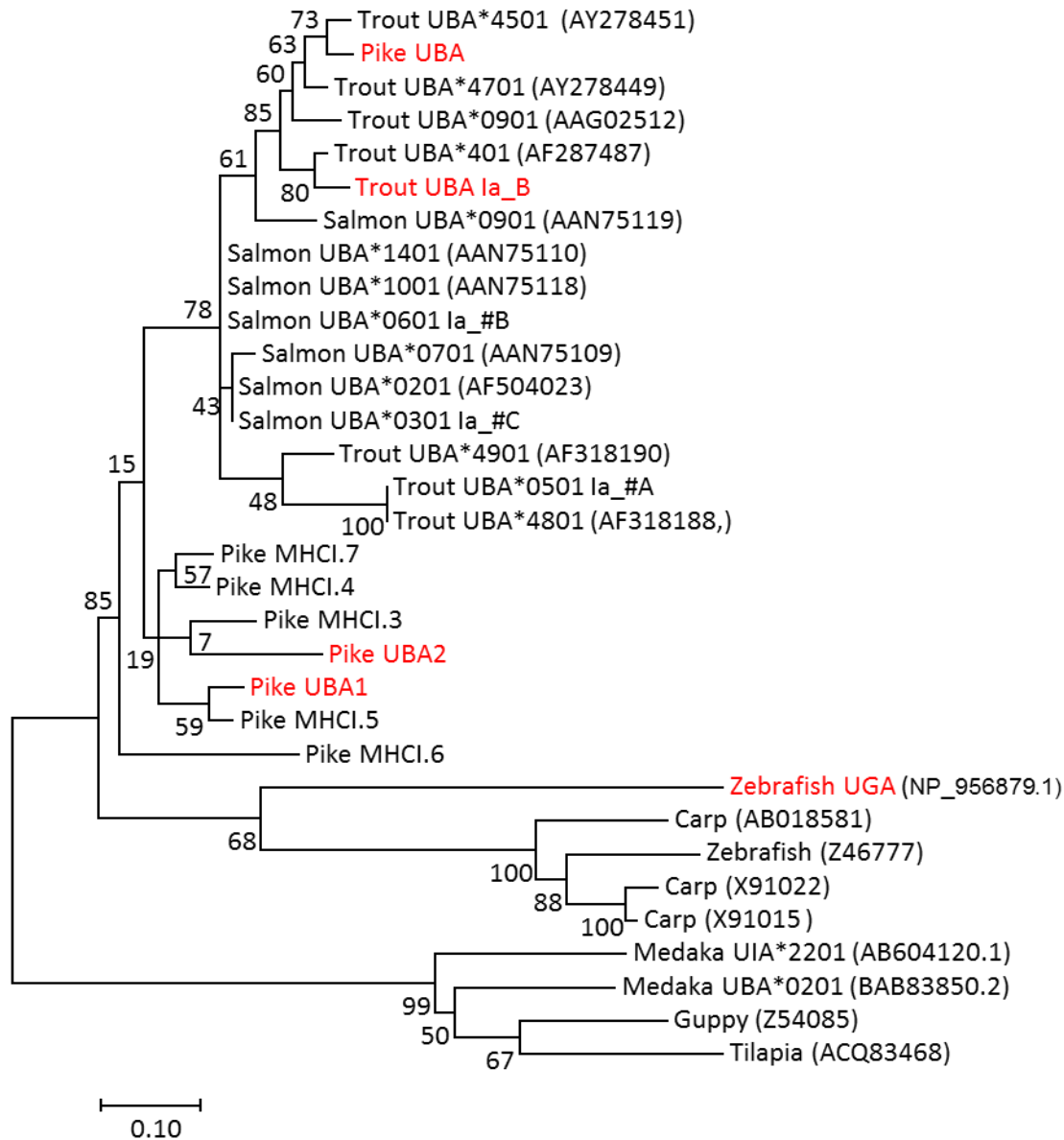


Figure S2d. Alignment of selected MHC class I amino acid sequences

Aligned sequences mostly originate from the haplotypes described in main text Figure 1, Additional file 1: Figure 1 and Additional file 2: Text S1 and are as follows: *onmyUBA*3301*= rainbow trout genome haplotype Ia_#B, *onmyUBA*0501*= rainbow trout BAC haplotype Ia_#A, *sasaUBA*0201*= Atlantic salmon BAC 92104 haplotype Ia_#A, *sasaUBA*0601*= Atlantic salmon BAC 714P22 haplotype Ia_#B, *sasaUBA*0301*= Atlantic salmon genome Ia region haplotype Ia_#C, *sasaULA*= Atlantic salmon genome Ia region haplotypes Ia_A and C, *onkiUBA*= coho salmon genome, *dareUGA*= zebrafish haplotype 19D (4), *esluUBA1/-2*= the Northern pike genome Chr.10, *esluMHCI.3-1.7*=unplaced genomic scaffold NW_017859271 and HLA-A2= AAA76608.2. Numbering according to the mature HLA-A2 sequence. Sequence references not shown in figures S2a-d can be found in Additional file 2: Text S1.

	Leader sequence	Alpha 1 domain	
<i>OnmyUBA*Ia_#B</i>	: -----	ATHSLKYFYTAVSG-DIDFPEFTAVGLVDKGQ	: 31
<i>OnmyUBA*0501_#A</i>	: ----MK-GFILLLLGIGLLHTASA.....R.V...T..-VP.....VSL.I..GM.		: 50
<i>SasaUBA*0201_#A</i>	: ----MKSCILLLLFLGIVLLHTVSA..NT.Q.....T..-IDN....VTM.I.NGH.		: 51
<i>SasaUBA*0601_#B</i>	: -----		: -
<i>SasaUBA*0301_#C</i>	: ----MKC-FILLLLG-IALHSSSA.....R.V...T..-IP.....VT....NGEP		: 49
<i>SasaULA_#A/C</i>	: ----MKC-FILLLLS-ISLHAASA.M...R.V...T..-MP.....VT....NGEP		: 49
<i>OnkiUBA</i>	: ----MK-GIILLVLGIGLLHTASAV.....S.E-VPN...VV..M..GV.		: 50
<i>EsluUBA1</i>	: -----MSGVQT...R.V...T..-IP.....VT....NGEP		: 36
<i>EsluUBA2</i>	: ----MK-VIILSILGIGFLPTVSAV.....G...-X.N....T....N..		: 50
<i>EsluMHCI.3</i>	: ----MK-VFILLIMGIGLLHSVSA.....C.V...T..-IP.....VT....NREP		: 50
<i>EsluMHCI.4</i>	: ----MK-VIILLILGIGLLPTVSA.S..IR.V...T..-IP.....VNL.IL.GM.		: 50
<i>EsluMHCI.5</i>	: ----MK-VFILLILGXGHLHTLSA.F...HTC.....-THL.....L..L.NV.		: 50
<i>EsluMHCI.6</i>	: ----M-YFFVLLILIIIGHLHTVSAV.....G...-..N....T....NVT		: 50
<i>EsluMHCI.7</i>	: ----MK-VIVLLILGIG-LPTVSTV.....G...-G.N....V....N..		: 49
<i>DareUGA</i>	: -----MQQVLLFLLGAHLAYAG...R.....-.....M.....G..		: 47
<i>HLA-A2</i>	: MAVMAPRTLVLLLSGALALTQTWAGS..MR..F.S..RPGRGE.R.I...Y..DT.		: 56
		1 * 20 *	
			Glycosylation site
<i>OnmyUBA*Ia_#B</i>	: FMYFDS--STKTAVPKTEWMKR-EGADYWDRQTQGLIGAHQTFKVNIQTLKDRFNQ		: 84
<i>OnmyUBA*0501_#A</i>	: ID.Y..--N..RV...QD..AKT..S...E....NS..DE....A..DVA.Q....		: 104
<i>SasaUBA*0201_#A</i>	: IDHY..--I..R.IQ.A..ISGAVDP...KTN..IYA.TETV.VN..NVA.S....		: 105
<i>SasaUBA*0601_#B</i>	: -----		: -
<i>SasaUBA*0301_#C</i>	: IS.Y..--IIRRET.RQD..AKT..S...ES...VS..SE....A..DVA.Q....		: 103
<i>SasaULA_#A/C</i>	: IS.Y..--IIRSET.RQD...EAVDP...N.N..TS..DE....A..DVA.Q....		: 103
<i>OnkiUBA</i>	: MVHY..--NSQR...QD.VNKAADPQ..E.N.GNFK.SQ....A..DIV.Q....		: 104
<i>EsluUBA1</i>	: IS.Y..--IIR.ET.RQD..XKAVDPE..S.N..LS..TEK...D..DVPNQ....		: 90
<i>EsluUBA2</i>	: .I....--N.....-..P...SE..TG.NL..V..NS..STM....		: 103
<i>EsluMHCI.3</i>	: IS.Y..--IIR.ET.RQD..ANN..S...ES...ISN.TE....A..DI..Q....		: 104
<i>EsluMHCI.7</i>	: .I....--N.....-..P...S...I...SQPV..N..VA.....		: 102
<i>EsluMHCI.6</i>	: .I....--N.....I..-V.PG..VS..KTC.NL..F..N..VA.....		: 103
<i>EsluMHCI.5</i>	: VA....--N.....IR..-..P..LNSL.NLG.N...RL.AHL.IF..Q....		: 103
<i>EsluMHCI.4</i>	: ID.Y..--N..KV...QD..AKAV.S...E..NS..REKS..A..D.A.Q....		: 104
<i>DareUGA</i>	: .I....--KKME.....IRQN.....IN..R..AT..A..N..VA.E....		: 101
<i>HLA-A2</i>	: .VR...DAASQRME.RAP.IEQ-..PE..GE.RKVKAHS..HR.DLG..RGYY..		: 111
		40 * 60 * 80	

Alpha 2 domain TAPBP dependency (5)

```

OnmyUBA*Ia_#B : SKSTGVHVFQWMYGCEWDDDEAGVTEGFDQYGYDGEDFIAFDLKTTKWIAPTPQAVI : 140
OnmyUBA*0501_#A : TG--...T..L....LG.D-.I.R.DF.L...A..LSL.KS.LT.T.ANQK... : 157
SasaUBA*0201_#A : T---...N.K.....T.....D.....L.....LT..... : 158
SasaUBA*0601_#B : -----...N.....E.W.....KS..... : 51
SasaUBA*0301_#C : TG--...N.K.....T.....D.....L.....LT..... : 157
SasaULA_#A/C : TG--...Y.N.....L.....L.....SL..... : 157
OnkiUBA : .G--...IV.M.....T.....E.H.....M..... : 158
EsluUBA1 : TM--...IN.L.....T.....F.F.....SL..M.LT...K..... : 144
EsluUBA2 : .M.....V.M.....TEA...R.....LLL...LS...KS..F... : 159
EsluMHCI.3 : TG--...V.N.....T.....D.....V...RLT...K...F... : 158
EsluMHCI.7 : .M.....V.M.....T..I.....H.....S...L..V...F... : 158
EsluMHCI.6 : .M.....L.....D...T..IQ.....N..SL...LS...K...F... : 159
EsluMHCI.5 : .M.....TL.V.M.....T.....YM.F.....S...LT..TSK... : 159
EsluMHCI.4 : TG--...A.M.F.....T...R..N.Q.F.....S...LT...K..... : 158
DareUGA : .Q--...T..V.....LE.DG-S.R.YW.....LSL.KS.LT.T.TK..... : 154
HLA-A2 : .E-A.S.TV.R....DVGSDWRFLR.YH..A...K.Y..LKEDLRS.T.ADMA.QT : 166

```

* 100 * 120 * 140

Alpha 3 domain

```

OnmyUBA*Ia_#B : TKHKWDSNTARNE-QNKNYTQICIEWLKKYVDYGKSTLMRTVPPSVSLLQK---A : 192
OnmyUBA*0501_#A : ..L...ATG.EAN-FQ...LENT.....N...D..E.K.R.....---T : 209
SasaUBA*0201_#A : ..L.....Q...-YR...L..T.....L.....---T : 210
SasaUBA*0601_#B : ..L...D..Q...-HR.....---T : 103
SasaUBA*0301_#C : ..L.....Q...-YR...L..T.....L.....---T : 209
SasaULA_#A/C : ..L...N.M.QIQ-.D.H.L.TS.....L.....---T : 209
OnkiUBA : ....L.....N...-YK.....T..D.....C.....F...---T : 210
EsluUBA1 : S.LRL.N.IVHSNNYL...F..E..D...FL.....S.....---T : 197
EsluUBA2 : ..N.F...KDT.L-YLRK.F..T.VD.....S.....T.....---T : 211
EsluMHCI.3 : N.N...NDRGWI.-.T...L..T.....S.....F...R---T : 210
EsluMHCI.7 : S.N.L.Y.E.G.-GE.H.L.E.....S.....---T : 210
EsluMHCI.6 : ..N...DK.D...-YL.H.F.LE.....LQ.EETSQ.....---T : 211
EsluMHCI.5 : ..DR..NDK.D...-YW...L..E..DS..NFL...S.....---T : 211
EsluMHCI.4 : ..N...N.K.G...-YW...L..E..D.V...E..S.....---T : 210
DareUGA : ..N...AD-NADRQYT.S.LENE...V.....D..E.KDA.E.FM...---KD : 206
HLA-A2 : ....EAA-HVA.QLR-A.LEGT.V...RR.LEN..E..Q..DA.KTHMTHHAVSD : 220

```

* 160 * 180 *

```

OnmyUBA*Ia_#B : PSSPVTCHATGFYPRDVMVSWQKDGQDHHEDVEYGEILQNDGTFQKSSHLTVTPE : 248
OnmyUBA*0501_#A : .....SG...F.....EQ.G...H.....T..... : 265
SasaUBA*0201_#A : .....SG.....H..T..... : 266
SasaUBA*0601_#B : .....SG.....T..... : 159
SasaUBA*0301_#C : .....SG.....H..T..... : 265
SasaULA_#A/C : .....SG.....T..... : 265
OnkiUBA : S..L.....S.....T.P..... : 266
EsluUBA1 : .....SE.I.T.K...EQ....V..T.H.....V...K... : 253
EsluUBA2 : .....SR.N.T.K.N.EQ....M...P.....T.VR.N.K... : 267
EsluMHCI.3 : .....SR...T.M...EQY...V..T.....V...K... : 266
EsluMHCI.7 : .....SG.H.A.K...M..T.....VL...K... : 266
EsluMHCI.6 : .....SE.NIL...PFQ....T.HT.P.G.....VL...K... : 267
EsluMHCI.5 : .....SE.I.T.K.N.....M..T.H.....V...K... : 267
EsluMHCI.4 : .....SG.H.T.K.....M..T.....VR.N.K... : 266
DareUGA : .....V.Q.....SNI.MT...NKEE.F...DV.AT.T.A.....TVT.S.K... : 262
HLA-A2 : HEATLR.W.LS...AEITLT...R..E.QTQ.T.LV.TRPAG.....WAAVV.PS- : 275

```

200 * 220 * 240 *

Connecting peptide/ transmembrane region		
OnmyUBA*Ia_#B	: DRKNSK--YQCVVQVKGIKEDFIGVLDP-----QDAANVVPIIVGVV	: 288
OnmyUBA*0501_#A	: EW..N.--.....LA..ED.ITK..IESEIQTNFGKTNRGSN.PITIGL..G..I	: 319
SasaUBA*0201_#A	: EW..N.--.....T.LQ...K..TESEIKTN-----WNDP.I.L..GV..	: 312
SasaUBA*0601_#B	: EW..N.--.....T.V...K..TESEIKTN-----WNDP.I.L..V..	: 205
SasaUBA*0301_#C	: EW..N.--.....T.LQ...K..TESEIKTN-----WNDP.I.L..GV..	: 311
SasaULA_#A/C	: EW..N.--.....T.V...K..TESEIKTNWGNT-----NIGF.P-----	: 307
OnkiUBA	:K.....L.....	: 306
EsluUBA1	: EW..I.--.....S..N...K..TEDQIQTNCRKTNR---PTSTGL..GV..	: 304
EsluUBA2	: E...N.--.....A..N...K..TE-----TTSIG...GV..	: 306
EsluluMHC1.3	: EW..KN.--.....S.....K..TEDQIQTN-----TGFPV-----	: 303
EsluMHC1.7	: EW..N.--.....S..NK...K..TESEIQTNR-----EPTSIG...G-A..	: 311
EsluMHC1.6	: EW..N.--.....S..NK...K..TESEIQTNR-----RDTSIGH...TV..	: 312
EsluMHC1.5	: EW..K.--.....S..N.V...K..TEDEIETN-----WGKTQK-----	: 305
EsluMHC1.4	: EW..N.--.....IS.T...K..TEDELVSNWGKTNRGVN.QTSIG...GV..	: 320
DareUGA	: EW..N.EA.R...HV.A.N.V.VTVK.IRSNEG-----SDN---T.AII.G	: 306
HLA-A2	: -GQEQR--.T.H..HE.LP--KPLT.RWEPSSQP-----TIPIVGIIAG	: 314
	260 * 280 *	
Cytoplasmic region		
OnmyUBA*Ia_#B	: ALLLVVAVVV--GVVIWRKRSKKGFPASRTESDLENQRRSKDQTNVQRGIVKN	: 341
OnmyUBA*0501_#A	: ...VII.V-----K.KN.....TSDT.S..SGKGIQKI-----	: 360
SasaUBA*0201_#A	:K.K.....TSDT.SD.SG.AAQM.-----	: 356
SasaUBA*0601_#B	:L.....K.K.....TSDT.SD.SG.AAQM.-----	: 249
SasaUBA*0301_#C	:K.K.....TSDT.SD.SG.AAQM.-----	: 355
SasaULA_#A/C	: -----NTSDVGSNSSHNTAPKE-----	: 325
OnkiUBA	:T.TSDT.SD.SG.AAP.I-----	: 350
EsluUBA1	: GVSIL.IMAA.F.....KNK...VSDGG.NSSNNSPKP-----	: 343
EsluUBA2	: .G...I..A.II-.....K.K.....NTSDDGSNSSNNTAPKA-----	: 351
EsluMHC1.3	: -----ANTSDDGSNSSNNTAPKA-----	: 321
EsluMHC1.7	: .G...IAV..-I-.....K.K.....I..NTSDDGSNSSNNTAPKA-----	: 355
EsluMHC1.6	: .G...AIG.A.I--...NRK.K.....PNTSDDGSNSSNK.SSNV-----	: 356
EsluMHC1.5	: -----	: -
EsluMHC1.4	: .G...I..A.II-.....K.NR.....NTSDGG--SSTN.APEA-----	: 362
DareUGA	: CVAA.ALLA-.IA.LIY..RSNGY.KASSKD.D.EQSDP.AVEVK-----	: 350
HLA-A2	: LV.FGA.ITGAVVAA.M..RK.SDRKGGSYSQAASSDSAQG.DVSLTACKV----	: 365
	300 * 320 * 340	

References:

1. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutations data matrixes from protein sequences. *Computer Applications in the Biosciences* 8:275-282.
2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. Kiryu I, et al. (2005) New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. *Fish & shellfish immunology* 18:243-254.
4. McConnell SC, et al. (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
5. Park B, Lee S, Kim E, & Ahn K (2003) A single polymorphic residue within the peptide-binding cleft of MHC class I molecules determines spectrum of tapasin dependence. *Journal of immunology* 170(2):961-968.

Additional file 5: Figure S3. PSMB9 and PSMB12 data

Table of Contents		Page
S3a	Phylogeny of deduced PSMB9 and PSMB12 amino acid sequences from selected species.	2
S3b	Alignment of deduced PSMB9 amino acid sequences	3
S3c	Alignment of deduced PSMB12 amino acid sequences	5

Figure S3a. Phylogeny of deduced PSMB9 and PSMB12 sequences from selected species.

Salmonid haplotype and pike deduced amino acid sequence references can be found in Additional file 2: Text S1 while remaining references are shown in the figure. Zebrafish sequences originate from McConnell et al(1). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel 2008 model (2). The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (3). Red font indicated gene sequences originating from PSMB8F gene containing haplotypes. Salmon is Atlantic salmon, Trout is rainbow trout and Pike is Northern pike.

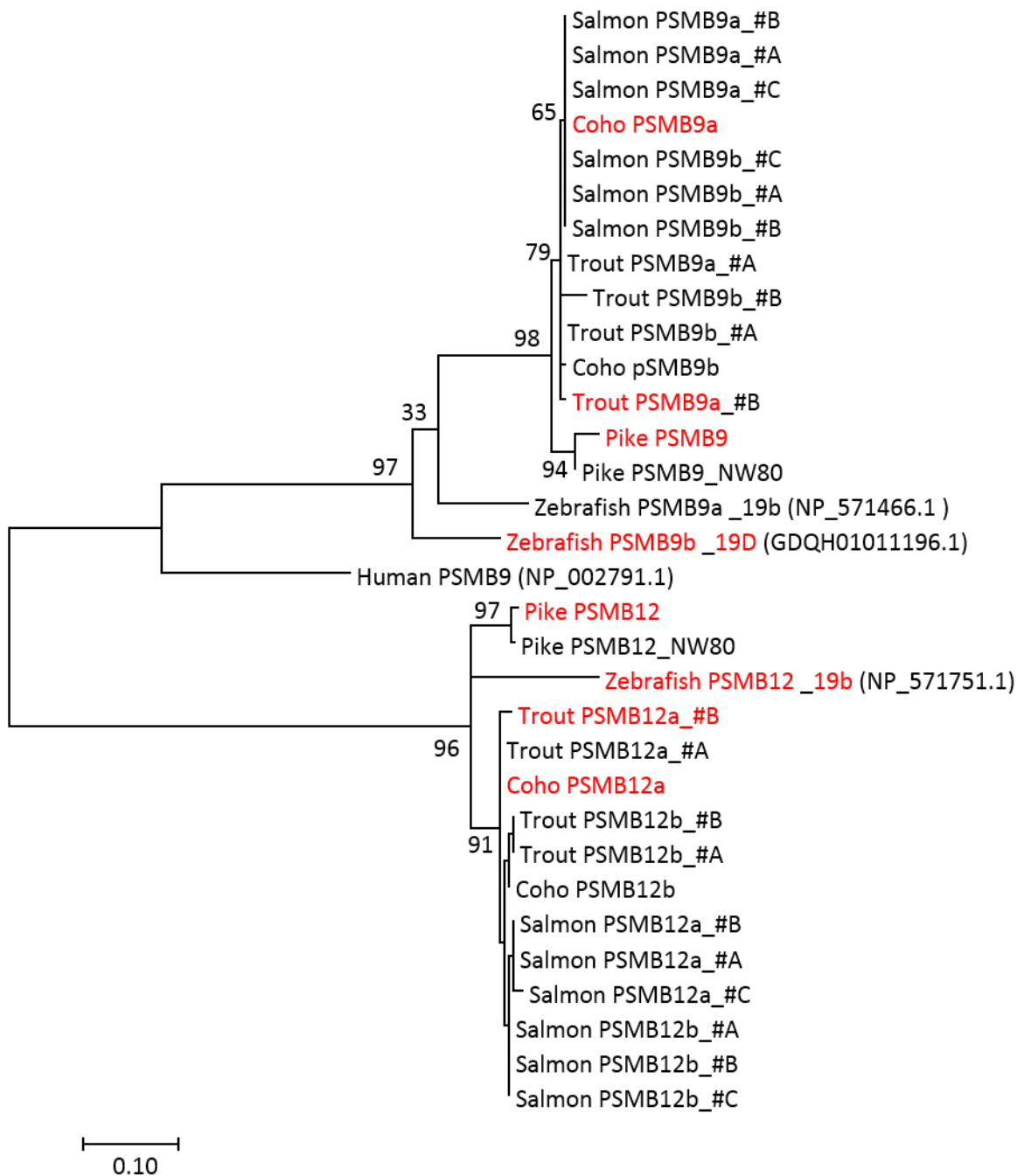


Figure S3b. Alignment of deduced PSMB9 amino acid sequence sequences

Alignment of deduced PSMB9 amino acid sequences from coho salmon (Onki), Atlantic salmon (Sasa), rainbow trout (Onmy), Northern pike (Eslu), zebrafish (Dare) and human sequences. Sequence references can be found in Additional file 2: Text S1 or in Figure S3a. The two Northern pike sequences are assumed allelic variants where EsluPSMB9_NW80 originates from scaffold NW_017859580.1 and EsluPSMB9 from the assembled region on Chr.10. For zebrafish haplotype PSMB9 sequences, see McConnell et al. (1). PSMB12 gene sequences originating from PSMB8F containing haplotypes are shown using red font. Amino acid residues are colour coded according to physiochemical properties.

		*	20	*	40	*																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
OnkiPSMB9a	:	M	L	E	E	S	S	-	E	P	G	W	L	S	E	-	E	V	K	T	G	T	T	I	I	A	I	E	F	D	G	G	V	V	L	G	S	D	S	R	V	S	A	G	E	T	V	V	N	-	R	V	M	N	K	L	S	L	L	:	55																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
SasaPSMB9a_#A	:	.	.	.	.	.	.	-	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.</

Additional file 5: Figure 3

	120	*	140	*	160	*			
OnkiPSMB9a	:	AHLIVAGWDKRGGGQVY-VTLNGLLSRQPF	AVGGSGSAYVYGFVDAEYRKAMSKEDCQ	:	170				
SasaPSMB9a_#A	:	-		:	170				
SasaPSMB9a_#B	:	-		:	170				
SasaPSMB9a_#C	:	-		:	170				
SasaPSMB9b_#A	:	-		:	170				
SasaPSMB9b_#B	:	-		:	170				
SasaPSMB9b_#C	:	-		:	170				
OnmyPSMB9a_#A	:	-	S	:	170				
OnmyPSMB9a_#B	:	-	S	:	170				
OnmyPSMB9a_#A	:	-	S	:	170				
OnmyPSMB9b_#B	:	-	S	:	170				
OnkiPSMB9b	:	K	S	:	170				
EsluPSMB9	:	R	S	Q.G	:	169			
EsluPSMB9_NW80	:	F	S	Q.G	:	170			
DarePSMB9a	:	K	-A.S.TK	I.....F.IN	K.N.T.RE	:	171		
DarePSMB9b	:	R.E	-A.S.T	F.I	AG.T.KE	:	169		
Human PSMB9	:	M	Q.E	-G.G.M.T	I.....TFI	Y.....A.KPG	P.E.R	:	172
	180	*	200	*	220	*			
OnkiPSMB9a	:	QFVVNTLSLAMSRDGS	SGGVAYLVTIDEKGAEK	CILGNELPTFYDQ	-----	:	217		
SasaPSMB9a_#A	:	-----	-----	-----	-----	:	217		
SasaPSMB9a_#B	:	-----	-----	-----	-----	:	217		
SasaPSMB9a_#C	:	-----	-----	-----	-----	:	217		
SasaPSMB9b_#A	:	-----	-----	-----	-----	:	217		
SasaPSMB9b_#B	:	-----	-----	-----	-----	:	217		
SasaPSMB9b_#C	:	-----	-----	-----	-----	:	217		
OnmyPSMB9a_#A	:	-----	-----	-----	-----	:	217		
OnmyPSMB9a_#B	:	S	-----	-----	-----	:	217		
OnmyPSMB9a_#A	:	-----	-----	-----	-----	:	217		
OnmyPSMB9b_#B	:	-----	-----TFVVP-----	-----	-----	:	206		
OnkiPSMB9b	:	-----	-----	-----	-----	:	217		
EsluPSMB9	:	I	-----	-----	-----	:	216		
EsluPSMB9_NW80	:	I	-----	-----	-----	:	217		
DarePSmb9a	:	E.....A.T.....G.....V.....KD.T.....V.....K.F.E	-----	-----	-----	:	218		
DarePSmb9b	:	E.....I.S.....G.....SESV.....Q.....PDTVEPSKAVKV	-----	-----	-----	:	227		
Human PSMB9	:	R.TTDAIA.....I.....TAA.VDHRV.....K.....E	-----	-----	-----	:	219		

Figure S3c. Alignment of deduced PSMB12 amino acid sequences

Alignment of deduced PSMB12 amino acid sequences from coho salmon (Onki), Atlantic salmon (Sasa), rainbow trout (Onmy), Northern pike (Eslu), zebrafish (Dare) against the human PSMB9 sequences. Sequence references can be found in Additional file 2: Text S1 or in Figure S3a. The two Northern pike sequences are assumed allelic variants where EsluPSMB12_NW80 originates from scaffold NW_017859580.1 and EsluPSMB12 from the assembled region on Chr.10. The zebrafish darePSMB12 sequence is from McConnell et al. (1). PSMB12 gene sequences originating from PSMB8F containing haplotypes are shown using red font. Amino acid residues are colour coded according to physiochemical properties.

		*	20	*	40	*	
OnkiPSMB12a	:	-MERHL-MDSQIKG--VSTGTTILAVTFNGGVIIGSDSRASIGGYVSSKTINKLI	:	52			
SasaPSMB12a_#A	:	-.....-.....-.....S.....	:	52			
SasaPSMB12a_#B	:	-.....-.....-.....S.....	:	52			
SasaPSMB12a_#C	:	-.....-.....-.....S.....	:	52			
OnmyPSMB12a_#A	:	-...KRF-.....	:	52			
OnmyPSMB12a_#B	:	-----	:	35			
SasaPSMB12b_#A	:	-.....F-.....S.....	:	52			
SasaPSMB12b_#B	:	-.....F-.....S.....	:	52			
SasaPSMB12b_#C	:	-----S.....	:	47			
OnmyPSMB12b_#A	:	-----N.....	:	47			
OnmyPSMB12b_#B	:	-----N.....	:	47			
OnkiPSMB12b	:	-----	:	47			
EsluPSMB12	:	-.....Y-...K-.....M.N.C.....	:	52			
EsluPSMB12_NW80	:	-.....Y-...K-.....M.S.C.....	:	52			
DarePSMB12	:	-...D...H-PY...VN-...K...M.ES.....	:	52			
Human PSMB9	:	MLRAGAPTGDLPAGE.H...M..E.D...VM...V.A.EAV.N-RVFD..S	:	55			
		60	*	80	*	100	*
OnkiPSMB12a	:	QVHDRIFCCIIAGSLADAQAVTKAAKFQISFHSIQMESPLVKAAASVLKELCYNNK	:	108			
SasaPSMB12a_#A	:		:	108			
SasaPSMB12a_#B	:		:	108			
SasaPSMB12a_#C	:		:	108			
OnmyPSMB12a_#A	:		:	108			
OnmyPSMB12a_#B	:		:	91			
OnmyPSMB12b_#A	:		:	103			
OnmyPSMB12b_#B	:		:	103			
SasaPSMB12b_#A	:		:	108			
SasaPSMB12b_#B	:		:	108			
SasaPSMB12b_#C	:		:	103			
OnkiPSMB12b	:		:	103			
EsluPSMB12	:	T.....L.....M.....	:	108			
EsluPSMB12_NW80	:	T.....L.....M.....	:	108			
DarePSMB12	:	M.....L.....IMR...S...	:	108			
Human PSMB9	:	PL.E..Y.ALS..A...ADM.AY.LEL.G.EL.E...L..N.VRNIS.KYR	:	111			
		120	*	140	*	160	
OnkiPSMB12a	:	EELQAGFITAGWDRKKGPQVYTVALGGMLLSQPFTIGSGSGSTYIYGYADAKYKPDM	:	164			
SasaPSMB12a_#A	:		:	164			
SasaPSMB12a_#B	:		:	164			
SasaPSMB12a_#C	:	VM.....	:	164			
OnmyPSMB12a_#A	:		:	164			
OnmyPSMB12a_#B	:		:	147			
OnmyPSMB12b_#A	:		:	159			
OnmyPSMB12b_#B	:		:	159			
SasaPSMB12b_#A	:		:	164			
SasaPSMB12b_#B	:		:	164			
SasaPSMB12b_#C	:		:	159			
OnkiPSMB12b	:		:	159			
EsluPSMB12	:	K.....L.....T.....	:	164			
EsluPSMB12_NW80	:	K.....L.....T.....	:	164			
DarePSMB12	:	...R.....I.V.S.....V...F...	:	164			
Human PSMB9	:	.D.S.HLMV...QRE.G...-GT...TR...A...F...V.A...G.	:	166			

	*	180	*	200	*	220	
OnkiPSMB12a	:	SKEECLQFATNALALAMGRD	NVSGGVAHLVVITEEGVEH	IIVPGDKLPKFHDE	:	217	
SasaPSMB12a_#A	:	K.....	V.....	:	217		
SasaPSMB12a_#B	:	K.....	V.....	:	217		
SasaPSMB12a_#C	:	K.....	V.....	:	217		
OnmyPSMB12a_#A	:			:	217		
OnmyPSMB12a_#B	:			:	200		
OnmyPSMB12b_#A	:	.R.....	V.....	:	212		
OnmyPSMB12b_#B	:	.R.....	V.....	:	212		
SasaPSMB12b_#A	:		V.....	:	217		
SasaPSMB12b_#B	:		V.....	:	217		
SasaPSMB12b_#C	:		V.....	:	212		
OnkiPSMB12b	:	.R.....	V.....	:	212		
EsluPSMB12	:	.R.....	G.....V.....	:	217		
EsluPSMB12_NW80	:	TR.....	G.....V.....	:	217		
DarePSMB12	:	TL...AT...S.....	V.....A...K...V...E.....	:	217		
Human PSMB9	:	.P...RR.T.D.I.....	S...GS.....IY...T...AA...D.R...L.NE...Y...	:	219		

References:

1. McConnell SC, *et al.* (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
2. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular biology and evolution* 25(7):1307-1320.
3. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.

Additional file 6: Figure S4. PSMB7, PSMB10, PSMB13 data

Table of Contents		Page
S4a	Phylogeny of deduced PSMB7, PSMB10 and PSMB13 amino acid sequences	2
S4b	Alignment of deduced PSMB7, PSMB10 and PSMB13 amino acid sequences	3

Figure S4a. Phylogeny of deduced PSMB7, PSMB10 and PSMB13 amino acid sequences

Phylogenetic analysis of PSMB7, PSMB10 and PSMB13 sequences from selected species. Haplotype sequences originating from Atlantic salmon (Salmon), Rainbow trout, (Trout), Coho, Northern pike (Pike), and zebrafish are shown using extensions a/b_#A-C or _19. The remaining sequences have accession numbers in parenthesis. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. with frequency model (1). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Salmonid and Northern pike haplotypes sequences are described in Materials and Methods while Zebrafish haplotype sequences are from McConnell et al.(3) and medaka haplotype sequences from Hd-rR (4), HN1 (5) and cab (6). Other sequences have Genbank references in parenthesis. PSMB10 and -13 sequences residing in PSMB8F containing haplotypes are shown using red font. The tree is unrooted. Sequences can be found in Additional file 2: Text S1.

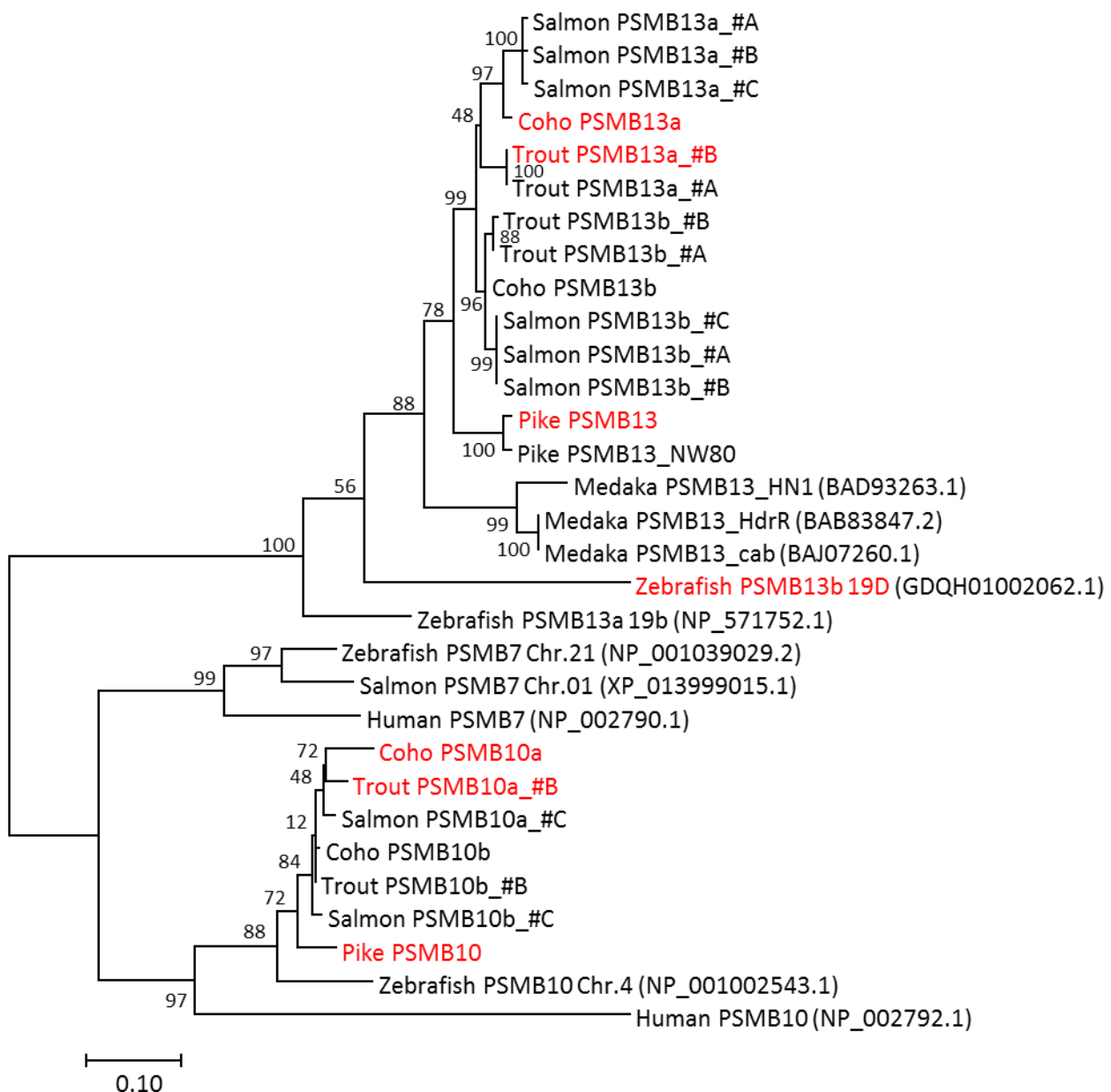


Figure S4b. Amino acid sequence alignment PSMB7_10_13 sequences

The SasaPSMB13a_#C sequence is incomplete but most likely does not represent a pseudogene. Sequence references can be found in Figure S4a or in Additional file 2: Text S1. Gene sequences originating from PSMB8F containing haplotypes are shown using red font. Onki is Coho salmon, Sasa is Atlantic salmon, Onmy is rainbow trout, Eslu is Northern pike, Orla is medaka and Dare is zebrafish.

		*	20	*	40	*	60	*	
OnkiPSMB13a	:	MALS-NVLEIPTSGFNFENVARNVALEGLLEGGQTKTPKPMKTGTTIAGLVCKEGVVLGAD	TRATSGEVVAD	:	71				
OnmyPSMB13a_#A	:	...-...V.T.A.....C.....Q.....A.....VL..N.....		:	71				
OnmyPSMB13a_#B	:	...-...V.T.A.....C.--...Q.....A.....VL..N.....		:	69				
SasaPSMB13a_#A	:	...-...K.....H...L.....		:	71				
SasaPSMB13a_#B	:	...-...T.....H...L.....		:	71				
SasaPSMB13a_#C	:	...-...H...L.....		:	71				
SasaPSMB13b_#A	:	...T-...V.T.A.....S.....H..A.....V...D.....		:	71				
SasaPSMB13b_#B	:	...T-...V.T.A.....S.....H..A.....V...D.....		:	71				
SasaPSMB13b_#C	:	...T-...V.T.A.....S.....H..A.....V...D.....		:	71				
OnmyPSMB13b_#A	:	...-...V.T.A.....S.....H..A.....I...D.....		:	71				
OnmyPSMB13b_#B	:	...-...V.T.A.....S.....H..A.....I...D.....		:	71				
OnkiPSMB13b	:	...-...V.T.A.....S.....H..A.....V...D.....		:	71				
EsluPSMB13	:	...P-...T.PT.....F.....R..A.....V...D.....S.....		:	71				
EsluPSMB13 NW80	:	...P-...T.P.....D..F.....R..A.....V...D.....S.....		:	71				
OrlaPSMB13 HdrR	:	...-...DS.AA....D.A...A.F...F...--Q...L.....V.F.D.....S.....		:	69				
OrlaPSMB13 HN1	:	...-...DS.AA....D.A...A.F...F...--Q...L.....V.F.D.....S.....		:	69				
OrlaPSMB13 cab	:	...-...DS.AA....D.A...A.F...F...--Q...L.....V.F.D.....S.....		:	69				
DarePSMB13a_19b	:	...TSH...PSLC....AT..IV..NGA.E.KI.P..AL...V.F.D.....D.....		:	72				
DarePSMB13b_19D	:	...FAT--...KT.EP..C...SL..E..DS.VK.---.K..AL...V.F.D.....NK...		:	67				
SasaPSMB10a_#C	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.....DDM....		:	72				
SasaPSMB10b_#C	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.....DDM....		:	72				
OnmyPSMB10b	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.....DDM....		:	72				
OnkiPSMB10b	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.....DDM....		:	72				
OnmyPSMB10a	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.....T.DDM....		:	72				
OnkiPSMB10a	:	...LHNSRPPQPQSA..S...TR..AV...N.SELGYSS..AR.....I.F.D..I.A.....DDM....		:	72				
EsluPSMB10	:	...LNTSRPPQPQSG..S...DSR..AV...N.SELGYSS..AR.....I.F.D..I.....DDM....		:	72				
DarePSMB10_4	:	...LNTST--KTL.G..S...TR..AV..AN.SELGYSA.NAR.....F.D..I.....DDM....		:	70				
Human PSMB10	:	...L---KPALE.RG..S...CQ..AS..RV.P.L--.V.HAR.....FQD..I.....NDS....		:	67				
DarePSMB7_21	:	...MATVS.CQYQPG..S...CR..AL..ADITKLGFSS.AAR.....C.I.Y.D.....E.MI....		:	71				
SasaPSMB7_LG01	:	...MATLS.CQPQLG..S...CK..AV..AEVTKLGCNI.AAR.....C.V.F.D.L.....E.MI....		:	71				
SasaPSMB7_LG11	:	...--TLS.CQTQLG..S...GK.LGLV.D-----L.....E.MI....		:	43				
Human PSMB7	:	...MAAVS.YAP.VG..S.D.CR..AV..ADFAKRGY.L..VR.....V.Y.D.I.....E.M....		:	71				

Additional file 6. Figure S4

	80	*	100	*	120	*	140					
OnkiPSMB13a	:	KMCAKIH	YISPN	YI--	CCGAGTAADTEKTTDLLSS	NTIFSMNSGRN	PRVVM	AVN	ILQDMLFR	YRGQIGASL	:	141
OnmyPSMB13a_#A	:	.	.	.	.	.	.	.	.	.	.	141
OnmyPSMB13a_#B	:	.	.	.	.	.	.	.	.	.	.	139
SasaPSMB13a_#A	:	.	.	.	.	.	.	.	.	.	.	141
SasaPSMB13a_#B	:	.	.	.	.	.	.	.	.	.	.	141
SasaPSMB13a_#C	:	.	.	.	.	S	--	.	.	.	.	133
SasaPSMB13b_#A	:	.	.	.	.	M	--	.	.	.	.	141
SasaPSMB13b_#B	:	.	.	.	.	M	--	.	.	.	.	141
SasaPSMB13b_#C	:	.	.	.	.	M	--	.	.	.	.	141
OnmyPSMB13b_#A	:	.	.	.	.	M	--	.	.	.	.	141
OnmyPSMB13b_#B	:	.	.	.	.	M	--	.	.	.	.	141
OnkiPSMB13b	:	.	.	.	.	M	--	.	.	.	.	141
EsluPSMB13	:	.	.	.	.	A	--	.	.	.	.	141
EsluPSMB13_NW80	:	.	.	.	.	A	--	.	.	.	.	141
OrlaPSMB13_HdrR	:	.	.	.	.	--	--	.	.	.	.	139
OrlaPSMB13_HN1	:	.	.	.	.	--	--	.	.	.	.	139
OrlaPSMB13_cab	:	.	.	.	.	--	--	.	.	.	.	139
DarePSMB13a_19b	:	.	.	.	.	A	--	.	.	.	.	142
DarePSMB13b_19D	:	.	E	.	.	LA	--	.	.	.	.	137
SasaPSMB10a_#C	:	.	N	M	.	A	--	.	.	.	.	142
SasaPSMB10b_#C	:	.	N	M	.	A	--	.	.	.	.	141
OnmyPSMB10a	:	.	N	M	.	A	--	S	.	.	.	142
OnmyPSMB10b	:	.	N	M	.	A	--	.	.	.	.	142
OnkiPSMB10a	:	.	N	M	.	A	--	CS	.	.	.	142
OnkiPSMB10b	:	.	N	M	.	A	--	.	.	.	.	142
EsluPSMB10	:	.	N	M	.	A	--	.	.	.	.	142
DarePSMB10_4	:	.	N	M	.	A	--	.	.	.	.	140
Human PSMB10	:	.	S	E	.	F	A	K	--	.	.	137
Dare PSMB7_21	:	.	N	S	.	.	--	.	.	.	.	141
SasaPSMB7_LG01	:	.	N	S	.	.	--	.	.	.	.	141
SasaPSMB7_LG11	:	.	N	S	.	.	--	CPSS	--	.	.	107
Human PSMB7	:	.	N	S	.	F	--	.	.	.	.	141

	*	160	*	180	*	200	*
OnkiPSMB13a	:	ILGGVDCSGNHLYTVGPPYGSIDNVPYLAMGSG-DLAALGILEDRFKPNMEMEEAKELVRDAIHSGIMSDLGS	:	212			
OnmyPSMB13a_#A	:	 T K - A	:	212			
OnmyPSMB13a_#B	:	 T K - A	:	210			
SasaPSMB13a_#A	:	 T Q -	:	212			
SasaPSMB13a_#B	:	 T Q -	:	212			
SasaPSMB13a_#C	:	 T Q - - - - -	:	185			
SasaPSMB13b_#A	:	 T - L	:	212			
SasaPSMB13b_#B	:	 T - L	:	212			
SasaPSMB13b_#C	:	 T - L	:	212			
OnmyPSMB13b_#A	:	 T - L Q	:	212			
OnmyPSMB13b_#B	:	 T - L . . . Q	:	212			
OnkiPSMB13b	:	 T - L	:	212			
EsLuPSMB13	:	 T . S D . K . T - N L	:	212			
EsLuPSMB13_NW80	:	 T . S D . K . T . S - L	:	212			
OrlaPSMB13_HdrR	:	 T VNK - HDL . L . K A	:	210			
OrlaPSMB13_HN1	:	 T VNK - G . QHDL . L . K . . K V N	:	210			
OrlaPSMB13_cab	:	 T VNK - HDL . L . K A	:	210			
DarePSMB13a_19b	:	 T . S M . K - N . . M V . DL . Q . A . S . . QA . . C	:	213			
DarePSMB13b_19D	:	L T . G DI DM . K . . F - N . P . M G . S . . DL . S . . H YA . V . N	:	208			
SasaPSMB10a_#C	:	. V VT . A . . . S . Y . H . . Y . KL . F . T . . . - AG . . IS . F . . . YR L K AA . FC	:	213			
SasaPSMB10b_#C	:	. V VT . A . . . S . Y . H . . Y . KL . F . T . . . - AG . . IS . F . . . YR L K AA . FC	:	212			
OnmyPSMB10a	:	. V VT . A . . . S . Y . H . . Y . KL . F . T . . GSGAG . . IS . F . . . YR . . . AQL . . . K AA . FC	:	214			
OnkiPSMB10a	:	TM . KAAI . - - - - IFEDR - - - - YR . N . EVRCS - - - - SF . L - - YLSLAQL . . . K AA . FC	:	197			
OnmyPSMB10b	:	. V VT . A . . . S . Y . H . . Y . KL . F . T . . . - AG . . IS . F . . . YR L K AA . FC	:	213			
OnkiPSMB10b	:	. V VT . A . . . S . Y . H . . Y . KL . F . T . . . - AG . . IS . F . . . YR L K TA . FC	:	213			
EsLuPSMB10	:	. V VT . A . . . SAY . H . . Y . KL . F . T . . . - AGP . IS . F R L . D . Q VA . FC	:	213			
DarePSMB10_4	:	. V VN . AQ . . S . Y . H . . Y . KL . F . T . . . - AAS . ISVF . . . Y L Q TA . FC	:	211			
Human PSMB10	:	. V LT . PQ . . G . H . H . . YSRL . FT . L . . . - QD . . AV Q . . . TL . A . QG . LVE . VTA . LG	:	208			
DarePsmB7_21	:	V T . P . . . SIY . H . . T . KL . VT - S . . MAVF . . . YR . D . . E . D . S AA . FN	:	212			
SasaPSMB7_LG01	:	V N . P . . . SIY . H . . T . KL . VT - S . . MAVF . . . Y . QD . . E . D . R AA . FN	:	212			
SasaPSMB7_LG11	:	- - - - -	:	-			
Human PSMB7	:	V VT . P . . . SIY . H . . T . KL . VT - S . . MAVF . . K . R . D . . E N . . SE . . AA . FN	:	212			

	220	*	240	*	260	*	280				
OnkiPSMB13a	: GNNIDICVITKQGV	DI	IRPYQSEYKDKRQKRYKYGPG	-TTSILTEKIVPLELEV	VQETVQQMDTA	---	: 277				
OnmyPSMB13a_#A	:		R. R.	-	R.	---	: 277				
OnmyPSMB13a_#B	:		R. R.	G.	R.	---	: 276				
SasaPSMB13a_#A	:		R. T.	-	K.	---	: 277				
SasaPSMB13a_#B	:		R. T.	-	K.	---	: 277				
SasaPSMB13a_#C	:			T.	K.	---	: 239				
SasaPSMB13b_#A	:		R. R.	-	R.	---	: 277				
SasaPSMB13b_#B	:		R. R.	-	R.	---	: 277				
SasaPSMB13b_#C	:		R. R.	-	R.	---	: 277				
OnmyPSMB13b_#A	:		R. R.	-	R.	---	: 277				
OnmyPSMB13b_#B	:		R. R.	-	M.	---	: 265				
OnkiPSMB13b	:		R. R.	-	R.	---	: 277				
EsluPSMB13	:	L.	F. M. H.	-	P. QM.	R.	: 277				
EsluPSMB13_NW80	:		F. R. H.	-	P. QM.	R.	: 277				
OrlaPSMB13_HdrR	:		F. ET.KPK.	R.	PV. K. V.	E. IQ.	: 275				
OrlaPSMB13_HN1	:	AE.	F. F. R.EIK.	RQ.	PV. K. T.	KM. T. K.E.V	: 275				
OrlaPSMB13_cab	:		F. ET.KPK.	R.	PV. K. V.	K. E. IQ.	: 275				
DarePSMB13a_19b	:	L. E.	HK. P. NY.	AK. KS.	P. KTVNK.	DL. H.	M.E. ASS.	: 281			
DarePSMB13b_19D	: . H. L.	RD.	 QHIG.	EGQ. K.	-	AV. LHHI.	MDLIE. S. K.E.S	: 273			
SasaPSMB10a_#C	: . S.V.L.	QA. Q. L.S.	DQPAQ. G.	KEGQ. K.	-	AV. KTVT.	P.D. D.SI.L.	AQ.	: 278		
SasaPSMB10b_#C	: . S.V.L.	QA. Q. L.S.	DQPAQ. G.	KEGQ. K.	-	AV. KTVT.	P.H. D.SI.L.	Q.	: 277		
OnmyPSMB10a	: . S.V.L.	HA. Q. L.G.	DQPAQ. G.	SRGQ. T.	K.	-	AV. KTVT.	P.DI. D.SI.L.	Q.	: 279	
OnkiPSMB10a	: . S.V.L.	QA. Q. L.G.	DQPA.	-	-	AV. KTVT.	P.DI. D.SI.L.	Q.	: 250		
OnmyPSMB10b	: . S.V.L.	QA. Q. L.G.	DQPA.	-	-	AV. KTVT.	P.D. D.SI.L.	Q.	: 266		
OnkiPSMB10b	: . S.V.L.	QA. Q. L.G.	DQPAQ. G.	EGQ. K.	-	AV. KTVT.	P.D. D.SI.L.	Q.	: 278		
EsluPSMB10	: . S.V.L.I.	QA. Q. L.A.	DQPAQ. G.	KEGQ. K.	-	AV. KTVT.	P.DI. D.S.	L. Q.	: 278		
DarePSMB10_4	: . S.V.L.	DKK. L.	T. DQPVH.	NQ. GGT.	R. K.	-	AV. SKTVT.	T.D. D.S.	HV. E.	: 276	
Human PSMB10	: . G.V.A.	 T.	AKLL. TLSSPTEPV.	SG. HFV.	-	AV. QTVK.	T. L.	E. A.	EVE.	: 273	
DarePSMB7_21	: . S. V.	GK. L.	HDIANK. GV.	TGS. R.	KH.	-	GV. SKAVT.	N. DM.	E. S.	T. S.	: 277
SasaPSMB7_LG01	: . S. L.	GR. L.	HDMANK. GV.	TGN. KQ.	-	GV. KSVIK.	D. E.	T.	: 276		
SasaPSMB7_LG11	:									: -	
Human PSMB7	: . S. L.	S. NKL.	FL. TVPNK.	GT. LG.	RCEK.	-	AV. T.	I. LE.	T. S.	: 277	

References:

1. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutations data matrixes from protein sequences. *Computer Applications in the Biosciences* 8:275-282.
2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. McConnell SC, et al. (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
4. Matsuo M, Asakawa S, Shimizu N, Kimura H, & Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). *Immunogenetics* 53:930-940.
5. Tsukamoto K, et al. (2005) Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. *Immunogenetics* 57:420-431.
6. Nonaka MI & Nonaka M (2010) Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics* 62(5):319-332.

Additional file 7: Text S3. Alignment of deduced TAP2 amino acid sequences

Salmonid and pike sequence references can be found in Additional file 2: Text S1. The assumed allelic Northern pike variants are from the genome (EsluTAP2) or from the unplaced genomic scaffold EsluNW80= NW_017859580.1) where some residues in the Pike TAP2 sequence are polished using the TSA sequence GATF01026336.1. Zebrafish sequences DareTAP2a_19B= NP_001006594.1 and DareTAP2d/e_19D are described in (1). Remaining sequence references are frog (*Xenopus laevis*) Xela TAP2_I=NP_001081860.1, Xela TAP2_II=AAI70327.1, human TAP2= AHW47941.1 and the chicken /Gaga) sequences are shown with accession numbers in the alignment. See rat (2), chicken (3) and frog (4) references for individual TAP polymorphisms. Polymorphic residues within species are shown using yellow shading. Variable residue positions in Northern pike TAP2 sequences coinciding with variable positions in chicken or rat sequences are shown using green shading. Cyan shading shows human TAP2 residues that crosslink with bound peptide (5) while predicted transmembrane regions (<http://proteininformatics.charite.de/rhythm/index.php?site=home>) are shown using grey shading in the coho (Onki) TAP2a sequence.

OnkiTAP2a	:	-MTKMMILRTCVFAMAVG-LCIDITTLA---FGASISKGTGRTSDAFGNVVRLLWVVAGIRLVL	:	59
SasaTAP2a_#A	:	-----AL-----P..F.....	:	59
SasaTAP2a_#B	:	-----AL-----P..F.....	:	59
SasaTAP2a_#C	:	-----M.A.....FCAPC..E..E..PI.FGT.....	:	59
OnmyTAP2a_#A	:	-----A..V..V..V..FCAPG..E.....PK.F.T.....S.....	:	58
OnmyTAP2a_#B	:	-----A..VT.....FCAPG..E.....PI.F.T.....S.....	:	58
SasaTAP2b_#A	:	-----R.....T.....FCATGL.....PISF.V..L.....E.....F	:	59
SasaTAP2b_#B	:	-----R.....T.....FCATGL.....PISF.V..L.....E.....F	:	59
SasaTAP2b_#C	:	-----R.....T.....FCATGL.....PISF.V..L.....E.....F	:	59
OnmyTAP2b_#A	:	-----FCATG.....P..PISF.V..SLL.....I.F	:	58
OnmyTAP2b_#B	:	-----FCATG.....P..PISF.V..SLL.....I.F	:	58
OnkiTAP2b	:	-----FCATG.....PISF.L..LL.....I.F	:	59
EsluTAP2NW80	:	-----IKA..L.F.LISD.II..FSAPGS.S---PI.C.....L.NV.I..V.....V	:	57
EsluTAP2	:	---M..I.IK..I.L.I.L.---I..IFSAPG.....I.QI.Y-T..Y..SV.I..V..F.F.V	:	58
OrlaTAP2_cab	:	MADR.IRKVVALIL.LC---L.LSLFY-----ASDRFSF..H.A...LS.AL.WSAVS	:	50
OrlaTAP2_HN1	:	MADR.IRKVVALIL.LC---L.LSLFY-----ASDRFSF..H.A...LS.AL.WSAVS	:	50
OrlaTAP2_HdrR	:	MADR.IRKVVALML.LC---L.LSLFY-----ASDRFSF..H.A...LS.AL.WSAVS	:	50
DareTAP2a_19B	:	-----RKVL...CMLC---F..LIVLILDLAETLL.SP.SSL.YN--LF.Q.AES.L.CT..Y	:	54
DareTAP2d_19D	:	-----ELKMQH.K.LI---V.LAMSCVLHYSTN---AFI.INSNVFLLAQ..L.S.VKWIG.R	:	53
DareTAP2e_19D	:	-----	:	-
XelaTAP2_I	:	-----MVMLQ---LPLIV.LLDVSLNYL.AHLVYKYCPDKLLAA..L.SL.K.PM.G	:	49
XelaTAP2_II	:	-----MASLS---FPVILFMADF.LNYFAAKLIYKYYPYEILGA..LLSL.K.PV.A	:	49
GagaBAG69399	:	-----MAMPPYILRLSC..LLADLALMLALAHFFPAL.HLGW.GS.LE..L..LV.G	:	52
GagaAEE25620	:	-----MAMPPYILRLSC..LLADLALMLALAHFFPAL.HLGW.GS.LE..L..LV.G	:	52
GagaBAG69412	:	-----MAMPPYILRLSC..LLADLALMLALAHFFPAL.HLGW.GS.LE..L..LV.G	:	52
GagaBAG69426	:	-----MAMPPYILRLSC..LLADLALMLALAHFFPAL.HLGW.GS.LE..L..LV.G	:	52
Rat TAP2_I	:	-----MALS--YPRPWAS.LLVDLALLGLLQRLSLGTLTPPGLPG..LEGTL..GV.W	:	50
Rat TAP2_II	:	-----MALS--HPRPWAS.LLVDLALLGLLQSLGTLTPPGLPG..LEGTL..GV.W	:	50
Human TAP2	:	-----MRLP--DLRPW.S.LLVDAALLWLQ.PLGTLTPQGLPG..LEGTL..GG.W	:	50

```

OnkiTAP2a      : GLSL--LTLGSIKPVFKRWLAVHCFLAPVYETGRMLMYGGSPERVYRSLGLGSPSLWPLCTVAA : 121
SasaTAP2a_#A   : ..T.--.....R.....S.....G.....L..A.. : 121
SasaTAP2a_#B   : ..T.--.....R.....S.....G.....L..A.. : 121
SasaTAP2a_#C   : ..T.--.....R.....S.....G.....L..A.. : 121
OnmyTAP2a_#A   : .....I.....T.SACG.--G.....L..A.. : 118
OnmyTAP2a_#B   : ....GSF.....I.....SACG.--G.....L..A.. : 120
SasaTAP2b_#A   : .....L.....K..H.S..SP.G.--G.....L..A.. : 119
SasaTAP2b_#B   : .....L.....K..H.S..SP.G.--G.....L..A.. : 119
SasaTAP2b_#C   : .....L.....K..H.S..SP.G.--G.....L..A.. : 119
OnmyTAP2b_#A   : .V.--.....L.....Q..H.S..STHG.--G.....L..A.. : 118
OnmyTAP2b_#B   : .V.--.....L.....Q..H.S..STHG.--G.....L..A.. : 118
OnkiTAP2b      : .V.--.....L.....T.....Q..H.S..STHG.--G.....L..A.. : 119
EsluTAP2NW80   : TV.VS--.....I.....T..L.....V.....I.H.S.L.SA.GP.--G.R.LV.... : 117
EsluTAP2       : SV.VT--..E.....I.....A..L..I.H.R.L.SSGP.--G.R.LV.... : 118
OrlaTAP2_cab   : FTVQ--VI..DL..LI.YITA.SL.QA.F.S.SRT.H.E--DSLCGRAA--DARC.LMF.G.. : 108
OrlaTAP2_HN1   : FTVQ--VI..DL..ALI.YITA.SL.QA.F.S.SRT.H.E--DSLCGRAA--DARC.LMF.G.. : 108
OrlaTAP2_HdrR  : FTVQ--VI..DL..LI.YITG.SL.QA.F.S.SRT.H.E--DSLCGRAA--DARC.LMF.G.. : 108
DareTAP2a_19B  : AG.T--..SKDAAD.LIR..IT...IGS.....AMFNRYRLD-----KFTV.LVG.... : 108
DareTAP2d_19D  : CFLRG---SWREAAVQ.CV..S.L.C.M..S.QLT.INTQ..NWSGC.S--C.GKTIISAA.T : 111
DareTAP2e_19D  : ----- : -
XelaTAP2_I     : TV..R..QDTPWLSSSRMYIFTMSLCQ.L.Q.L..VFISH...LPKSFEB-----NLYSHLIV : 106
XelaTAP2_II    : IV.VR..KCPAWVTISQLTFVLT.LSL.LHQ.L.FLFSSHV..MHSSFP-----SF.SHLPP : 106
GagaBAG69399   : .AGQ----.LAPRGPPGG--AVLLSLGPAIFL.V.GYVG-LPGAAPVLLAM-AT..WLV.THGT. : 109
GagaAEE25620   : .AGQ----.LAPRGPPGA--AVLLSLGPAIFL.L.GYVG-LPGAAPVLLAM-AT..WLV.THGT. : 109
GagaBAG69412   : .AGQ----.LAPRGPHGA--AVLLSLGPAIFL.L.GYVG-LPGAAPVLLAM-AT..WLV.THGT. : 109
GagaBAG69426   : .AGQ----.LAPRGPHGA--AVLLSLGPAIFL.L.GYVG-LPGAAPVLLAM-AT..WLV.THGT. : 109
Rat TAP2_I     : ..LK----V.GLLRLVGTF.PLL.LTT.LFFSL.ALVGSTMSTS.L.VAS-A.WGWLLAGY... : 109
Rat TAP2_II    : ..LK----V.GLLRLVGTF.PLL.LTN.LFFSL.ALVGSTMSTS.V.VAS-A.WGWLLADYG.V : 109
Human TAP2     : ..LK----.RGLLGFVGTL.LPL.LAT.LTVSL.ALVA.A.RAPPA.VAS-APW.WLLVGYG.. : 109
               60          *          80          *          100

OnkiTAP2a      : AAAALFWETTFPDSNGESNG---KQKKQKARVLFMRVLIFYRPDILLVGAFIFLSLAVLC--- : 179
SasaTAP2a_#A   : .....T.....Y.....T..... : 179
SasaTAP2a_#B   : .....T.....Y.....T..... : 179
SasaTAP2a_#C   : .....T.....Y.....T..... : 179
OnmyTAP2a_#A   : .....K.....E.T.....H.....T..... : 176
OnmyTAP2a_#B   : .....E.T.....H.....T.....A.... : 178
SasaTAP2b_#A   : .....K..KE.....E.T...A.....Y.....T.....A...S--- : 177
SasaTAP2b_#B   : .....K..KE.....E.T...A.....Y.....T.....A...S--- : 177
SasaTAP2b_#C   : .....K..KE.....E.T...A.....Y.....T.....A...S--- : 177
OnmyTAP2b_#A   : .....K...SE.....E.T.....Y.....T..... : 176
OnmyTAP2b_#B   : .....K...SE.....E.T.....Y.....T.....KYF : 179
OnkiTAP2b      : .....K...E.....E.T.....Y.....T..... : 177
EsluTAP2NW80   : .....KEK.E.....L...K..R..K...F.T..... : 178
EsluTAP2       : .T..F.....E.E.....L..VR.CKH.S.I.IC..... : 176
OrlaTAP2_cab   : LG.....I..ADDAET---EE...S.....RLFK..Y...LAGLL..... : 166
OrlaTAP2_HN1   : LG.....I..ADEAET---EE...S.....RLFK..Y...LAGLL..... : 166
OrlaTAP2_HdrR  : LG.....I..ADDAET---EE...S.....RLFK..C...LAGLL..C.... : 166
DareTAP2a_19B  : .L.C....I.L.T.E....ER.....I..IRL...YI..F..V..A.... : 166
DareTAP2d_19D  : VL.C....VS.S.T.RKAESSESVE..E.N.A...VRYSK..VA..S..V...I--- : 172
DareTAP2e_19D  : -----METADS-----RHLKTIF.L.SLS...L.PIC...L..VA..IG--- : 39
XelaTAP2_I     : FLSCML.DLL..YIFPKKTIKEEDNESK.E..N.I.L.KYSK..W.Q.S...SC.T..LIG--- : 167
XelaTAP2_II    : IF.CFL.DLVS.KLLPK.TEGEANN.-KRE.DN.I.LVKYSI..W.P.C...V..T..LIG--- : 166
GagaBAG69399   : VV.L.T.SLLV.TVATG-----A.EAE.W.PLR.L.ALAW.EWPF.GC...L..A..A.G--- : 163
GagaAEE25620   : VV.L.T.SLLV.TVATG-----A.EAE.W.PLR.L.ALAW.EWPF.GC...L..A..A.G--- : 163
GagaBAG69412   : VV.L.T.SLLV.TVATG-----A.EAE.W.PLR.L.ALAW.EWPF.GC...L..A..A.G--- : 163
GagaBAG69426   : VV.L.T.SLLV.TVATG-----A.EAE.W.PLR.L.ALAW.EWPF.GC...L..A..A.G--- : 163
Rat TAP2_I     : .LSLAV.AVLS.AGAQ.KE-----PGQENN.A.MI.L.RLSK..LPF.IV....AM..WG--- : 165
Rat TAP2_II    : .LSLAV.AVLS.AGAQ.KE-----PGQENN.A.MI.L.RLSK..LPF.IV....AM..W--- : 165
Human TAP2     : GLSWSL.AVLS.PGAQ.KE-----DQVNNK..MW.L.KLS...LP...A..F..V...G--- : 165
               *          120          *          140          *          160

```



```

OnkiTAP2a      : -EMFIPFYTGKVIDILGTQYKWNNFLTAIILMGLYSLGSSFSAGCRGGLFMCAINSFTCRMKVK : 242
SasaTAP2a_#A   : -.....E : 242
SasaTAP2a_#B   : -.....E : 242
SasaTAP2a_#C   : -.....E : 242
OnmyTAP2a_#A   : -.....S...E : 239
OnmyTAP2a_#B   : -.....S...E : 241
SasaTAP2b_#A   : -.T.....AS...D.....E : 240
SasaTAP2b_#B   : -.T.....AS...D.....E : 240
SasaTAP2b_#C   : -.T.....AS...D.....E : 240
OnmyTAP2b_#A   : -.....AS...D.....S...E : 239
OnmyTAP2b_#B   : I.....AS...D.....S...E : 243
OnkiTAP2b      : -.....AS...D.....S...E : 240
EsluTAP2NW80   : -Q.C.....A...K.D.R...F.FF.....C.....Q : 241
EsluTAP2       : -.....I.....K.M.D.R.V...F...C.....I...Q : 239
OrlaTAP2_cab   : -.....R.....SHFEQ.E..S.LLF.....V...V.....VL...S..... : 229
OrlaTAP2_HN1   : -.....R.....SHFEQ.E..S.LLF.....V...V.....VL...S..... : 229
OrlaTAP2_HdrR  : -.....R.....SHFEQ.E..S.LLF.....V...V.....VL...S..... : 229
DareTAP2a_19B  : -.....L...E.....SH.Q.D..RS...F...F.....V...Q : 229
DareTAP2d_19D  : -D.C.....L.....EH.QP.S.MS..F...L.....L.S.L.....T.LSRLNK.VRLM : 235
DareTAP2e_19D  : -D..M.AF...I..A.NSTFDHGT.TS..FF...T...G.....LT.SRL.K.VRER : 102
XelaTAP2_I     : -..I.V.Y...R.....S.K.NESV.ME..LF.AIF.IT...F..S.....F.SHS.L.Q.LRLL : 230
XelaTAP2_II    : -.....Y.M.R.....S.N.NERE..V.MFF.ATF.IT...IFS.....FSLAPL.R.LRLL : 229
GagaBAG69399   : -.TSV.YC..RAL.V.RQGDGLAA.TA.VG..C.A.AS..LF.....TFIRFR.VL.TRDQ : 226
GagaAEE25620   : -.TSV.YC..RAL.V.RQGDGLAA.TA.VG..C.A.AS..LF.....TFIRFR.VL.TRDQ : 226
GagaBAG69412   : -.TSV.YC..RAL.V.RQGDGLAA.TA.VG..C.A.AS..LF.....TFIGFR.VL.TRDQ : 226
GagaBAG69426   : -.TSV.YC..RAL.V.RQGDGLAA.TA.VG..C.A.AS..LF.....TFIRFR..IL.TRDQ : 226
Rat TAP2_I     : -.TL...H.S.R.....GDFDPDA.AS..FF.C.F.V...L.....S.LFTMSRINL.IREQ : 228
Rat TAP2_II    : -.TL...H.S.R.....GDFDPDA.AS..FF.C.F.V...L.....S.LF.ESRINL.IREQ : 228
Human TAP2     : -.TL...H.S.R.....GDFDPHA.AS..FF.C.F.F...L.....C.TYTMSRINL.IREQ : 228
               *          180          *          200          *          220

OnkiTAP2a      : LFGALVKQEIGFFETIKT--GDITSRLSTDTTLMGRAVALNVNVLLRTLTKTVGMLSIMMSLSW : 304
SasaTAP2a_#A   : .....M..--..... : 304
SasaTAP2a_#B   : .....--..... : 304
SasaTAP2a_#C   : .....--..... : 304
OnmyTAP2a_#A   : .....D.....--.....A..... : 301
OnmyTAP2a_#B   : .....D.....--.....A..... : 303
SasaTAP2b_#A   : .....S.....--.....K.A..L..... : 302
SasaTAP2b_#B   : .....S.....--.....K.A..L..... : 302
SasaTAP2b_#C   : .....S.....--.....K.A..L..... : 302
OnmyTAP2b_#A   : .....S.....--.....K.A.....I..... : 301
OnmyTAP2b_#B   : .....S.....--.....K.A.....I..... : 305
OnkiTAP2b      : .....D.S.....--.....K.A.....I..... : 302
EsluTAP2NW80   : ...T.....N.L--.....A.....T..... : 303
EsluTAP2       : .....V.....APK.....A..... : 301
OrlaTAP2_cab   : ..E..T.....--E...K.....A.T.C.A.....F.....I..... : 291
OrlaTAP2_HN1   : ..E..T.....--E...K.....A.T.C.A.....F.....I..... : 291
OrlaTAP2_HdrR  : ..E..T.....--E...K.....A.T.C.A.....F.....I..... : 291
DareTAP2a_19B  : ...S.IR.D.....--.....V..L...Y..V..... : 291
DareTAP2d_19D  : ..N.....DK..--..I...V..K..SQS..M...I...S...S...I..Y..I... : 297
DareTAP2e_19D  : ..SCF...D.A...KTTK--..L.TC..S.AS..S.SL.A...I...S..M.I.IYCF.IQ.C. : 164
XelaTAP2_I     : ..RS.I..D.A...LT..--.....Q..ARVS..I.A...II...V.CI.VY.F.F... : 292
XelaTAP2_II    : ..RSVI...A...DN.S--.....TH..ARVS.SI.A...IT..M.V.C.VY.F.F.I.. : 291
GagaBAG69399   : ..SS..YRDLA..QKTTA--AELA...T..V..ASNVL...I..M..N.GQVL.LCAF.LG..P : 288
GagaAEE25620   : ..SS..YRDLA..QKTTA--AELA...T..V..ASNVL...I..M..N.GQVL.LCAF.LG..P : 288
GagaBAG69412   : ..SS..YRDLA..QKTTA--AELS...T..V..ASNVL..I..M..N.GQVL.LCAF.LG..P : 288
GagaBAG69426   : ..SS..YRDLA..QKTTA--AELA...T..V..ASNVL...I..M..N.GQVL.LCAF.LG..P : 288
Rat TAP2_I     : ..SS.LR.DLA..QET..--ELN...S..S..S.WLPP.A.I...S.V.V..LYYF.LQV.P : 290
Rat TAP2_II    : ..SS.LR.DLA..QET..--ELN...S..S..SQWLS..A.I...S.V.V..LYYF.LQV.P : 290
Human TAP2     : ..SS.LR.DL...QET..--ELN...S.....SNWLP..A.....S.V.V..LYGF.L.I.P : 290
               *          240          *          260          *          280

```

```

OnkiTAP2a      : KLTLLMLMETPITGLLQSVHDNYYQRLSKEVQDSIARANEAGETVGGIRTVRSFKTEQHEAGR : 368
SasaTAP2a_#A   : .....M..... : 368
SasaTAP2a_#B   : .....M..... : 368
SasaTAP2a_#C   : .....M..... : 368
OnmyTAP2a_#A   : .....V.....L.....A..... : 365
OnmyTAP2a_#B   : .....V.....A..... : 367
SasaTAP2b_#A   : .....G.Y.....L.....M.....M.....A.....N.....RS..... : 366
SasaTAP2b_#B   : .....G.Y.....L.....M.....M.....A.....N.....RS..... : 366
SasaTAP2b_#C   : .....G.Y.....L.....M.....M.....A.....N.....RS..... : 366
OnmyTAP2b_#A   : .....V.....G.Y.....L.....T.....M.....A.....N.....RS.....SH : 365
OnmyTAP2b_#B   : .....V.....G.Y.....L.....T.....M.....A.....N.....RS.....SH : 369
OnkiTAP2b      : .....G.Y.....L.....T.....M.....A.....N.....RS.....H : 366
EsluTAP2NW80   : .....IY.....S.....A.....L.....T.....A.....S..... : 367
EsluTAP2       : .....IY.....D.....L.....T.....A.....I..... : 365
OrlaTAP2_cab   : .....V.....I.NIY.TK.....FQDM.N.M..T.DT.N.V.F..QV.....H..T...S. : 355
OrlaTAP2_HN1   : .....V.....I.NIY.TK.....S.FQDM.N.M..T.DT.N.V.F..V.....H..T...S. : 355
OrlaTAP2_HdrR  : .....V.....I.NIY.TK.....FQDM.N.M..T.DT.N.V.F..V.....H..TQ...S. : 355
DareTAP2a_19B  : .....L.....NIY.TH..K.....M.Q...D....A.S.....K.....LG..H. : 355
DareTAP2d_19D  : .....TFI..A..LIAIA..KIYNTH.EQ...D...V.....T...A..A..V.....CM..PS... : 361
DareTAP2e_19D  : P.A...SAF.S...ITAEKIYNK...G.VSS.KK..VKS.QV...S.YN..A...QGA.TV.QRL : 228
XelaTAP2_I     : Q....TFIS..A..F.WIT..K.YN...H.K.VQK.....SSS.L..R..MFESV...Y..AA.EE..K. : 356
XelaTAP2_II    : Q....TFISS..L.WII..KMYNH..H..D.V.K.....SS.L..K..IIESV...Y..AA.EE..K. : 355
GagaBAG69399   : R..M.A.L.V.LAVTARK.Y.TRH.M.QRA.L.AA.DTGA.VQ.SISS.EM..V.NG.EE..EH. : 352
GagaAEE25620   : R..M.A.L.V.LAVTARK.Y.TRH.M.QRA.L.AA.DTGA.VQ.SISS.EM..V.NG.EE..EH. : 352
GagaBAG69412   : R..M.A.L.V.LAVTARK.Y.TRH.M.QRA.L.AA.DTGA.VQ.SISS.EM..V.NG.EE..EH. : 352
GagaBAG69426   : R..M.A.L.V.LAVTARK.Y.TRH.M.QRA.L.AA.DTGA.VQ.SISS.EM..V.NG.EE..EH. : 352
Rat TAP2_I     : R..F.S.LDL.L.IAAEK.YNPRH.AVL..I..VV.K.GQVVR.A...LQ....GA.EQ.VS. : 354
Rat TAP2_II    : R..F.S.LDL.L.IAAEK.YNPRH.AVL..I..AV.K.GQVVR.A...LQ....GA.EQ.FR. : 354
Human TAP2     : R....S.LHM..F..IAAEK.YNTRH.EVLR..I..AV...GQVVR.A...LQ....GA.E..VC. : 354

```

300 * 320 * 340 *

```

OnkiTAP2a      : YNDRILMDTHNLKTRRDVTMAVYLLLRLTAVVMQVAMLYYGRFLFIQRGQMTTGNLVSFILYQSD : 432
SasaTAP2a_#A   : .....R.....S..... : 432
SasaTAP2a_#B   : .....R.....S..... : 432
SasaTAP2a_#C   : .....R.....S..... : 432
OnmyTAP2a_#A   : .....R.....L.....S..... : 429
OnmyTAP2a_#B   : .....R.....L.....S..... : 431
SasaTAP2b_#A   : .DH.....R.....K.....L.....S.....N : 430
SasaTAP2b_#B   : .DH.....R.....K.....L.....S.....N : 430
SasaTAP2b_#C   : .DH.....R.....K.....L.....S.....N : 430
OnmyTAP2b_#A   : .DH.....T.....R...V.....L.....S.....N : 429
OnmyTAP2b_#B   : .DH.....T.....R...V.....L.....S.....N : 433
OnkiTAP2b      : .DH.....R...V.....L.....S.....N : 430
EsluTAP2NW80   : .D.....I..S.....R.....I.L.....F.....Q..... : 431
EsluTAP2       : .DN.....I.....R.....I.L..T.....F.....V..Q..... : 429
OrlaTAP2_cab   : .DK...E..S..K....R.I..A...GLA...FI.....K..... : 419
OrlaTAP2_HN1   : .DK...E..S..K....R.I..A...GLA...FI.....K..... : 419
OrlaTAP2_HdrR  : .DK...E..S..K....R.I..A...GLA...FI.....K..... : 419
DareTAP2a_19B  : .DG...E.....R..I...I..M..ELG..K.....Y...S.....Q. : 419
DareTAP2d_19D  : .DE...T.....R.....V...KSLG..L...C...QL..KS..A..S.....G. : 425
DareTAP2e_19D  : .DTS..EE...YF..IH..S.R.T...FM..FQLC.R.LI.W..HQM.TS...P.....ME : 292
XelaTAP2_I     : .EKS..RE..Y..Q..C...F.R.L...VIK..INLST...L...C..QQL..HS..NISI..KM.....ME : 420
XelaTAP2_II    : .EKS..RE..V..Q...F.R.L...VT...VNLGP..IT....QIL.RS.FISS.KM.....ME : 419
GagaBAG69399   : .SQV..DR..LR..RDQ...ER..IF...IQ.VLQLAV.ALV..C.HQQLRE.TL.A.S..A....TK : 416
GagaAEE25620   : .SQV..DR..LR..RDQ...ER..IF...IQ.VLQLAV.ALV..C.HQQLRE.TL.A.S..A....TK : 416
GagaBAG69412   : .SQV..DR..LR..RDQ...ER..IF...IQ.VLQLAV.ALV..C.HQQLRE.TL.A.S..A....TN : 416
GagaBAG69426   : .SQV..DR..LR..RDQ...ER..IF...I..VLQLAV.ALV..C.HQQLRE.TL.A.S..A....TK : 416
Rat TAP2_I     : .KEA..ERCRQ.WW...LEKEL..VI..VM.LG...LI.NC.VQQ.LA.EV.R.G.L..L...EE : 418
Rat TAP2_II    : .KEA..ERCRQ.WW...LEKSL..VIQ.VM.LG...LI.NV.VQQ.LA.EV.R.G.L..L...EE : 418
Human TAP2     : .KEA..EQCRQ.YW...LER..L...V...VLHLGV..ML..SC.LQQM.D.EL.Q.S.L..MI...ES : 418

```

360 * 380 * 400 *

```

OnkiTAP2a      : LADNIRTLIYIFGDM LNSVGAAGKVFEYLDREPQVSTKGT LQPETLTGHVQFHNLSFSYPTRQ- : 495
SasaTAP2a_#A   : .....H.N.....- : 495
SasaTAP2a_#B   : .....H.N.....- : 495
SasaTAP2a_#C   : .....H.N.....- : 495
OnmyTAP2a_#A   : .....- : 492
OnmyTAP2a_#B   : .....- : 494
SasaTAP2b_#A   : .GA.....H.....- : 493
SasaTAP2b_#B   : .GA.....H.....- : 493
SasaTAP2b_#C   : .GA.....H.....- : 493
OnmyTAP2b_#A   : .GT.....H.....- : 492
OnmyTAP2b_#B   : .GT.....H.....- : 496
OnkiTAP2b      : .GT.....H.....- : 493
EsluTAP2NW80   : .....IF.....HK- : 494
EsluTAP2       : .....VY.....IC- : 492
OrlaTAP2_cab   : .G.....T.....V.....T..I..D.K.K.DEMK...TYCH...A..SNP- : 482
OrlaTAP2_HN1   : .G.....T.....V.....K..I..D.K.K.DEMK...TYRH...A..SNP- : 482
OrlaTAP2_HdrR  : .GV.....T.....V.....T..I..D.K.K.DEMK...TYRH...A..SNP- : 482
DareTAP2a_19B  : .G.....Q..KSE..ID.N.M.KD.K...K.QK.T...R.P- : 482
DareTAP2d_19D  : .G.SY.....V.MYS.....A.....K.L.D.D.S.H.K.A.Q.H.K.T.F..S.P- : 488
DareTAP2e_19D  : IGGHVQ..VELHI.LIE.LES.D...FM.HK.S.LS-.D.A.DQ.K...M.K.I...SSCP- : 354
XelaTAP2_I     : SG.Y...VHMLSEVTH.A..GH...H...QV.E..NT.K.C.DN.R.QFE.K.VT...S.P- : 483
XelaTAP2_II    : SNSY...VHMLSEITH.A...E...Q...K.K..AS.C.C.DK.Q.RFE.R.VT...S.P- : 482
GagaBAG69399   : AGSCVQA.A.SY..L.SNAV..C...D...W.RP.GAG..YV.TR.R..IT..RV..A...P- : 479
GagaAEE25620   : AGSCVQA.A.SY..L.SNAV..C...D...W.RP.GAG..YV.TR.R..IT..RV..A...P- : 479
GagaBAG69412   : AGSCVQA.A.SY..L.SNAA..C...D...W.RA.GAG..YM.TR.R..IT..RV..A...P- : 479
GagaBAG69426   : AGSCVQA.A.SY..L.SNAA..C...D...NW.RA.GAA..YM.TR.R...T..RV..A...P- : 479
Rat TAP2_I     : VGHHV.N.V.MY...SN...E...S...R.NLPKP...A.PRVE.R.E.QDV...S.P- : 481
Rat TAP2_II    : VGHHVQN.V.MY...SN...E...S...R.NLPNP...A.PR.E.R.E.QDV...S.P- : 481
Human TAP2     : VGSYVQ..V..Y...SN...E...S.M..Q.NLPSP...A.T..Q.V.K.QDV..A..N.P- : 481
               420          *          440          *          460          *          480

OnkiTAP2a      : ERKVLQGFSLERLPGQLTALVGPSGGGKSTCVSLLERFYQPQQGEILLDGLPLQSYQH HYLHRK : 559
SasaTAP2a_#A   : ..... : 559
SasaTAP2a_#B   : ..... : 559
SasaTAP2a_#C   : ..... : 559
OnmyTAP2a_#A   : ..... : 556
OnmyTAP2a_#B   : .....K. : 558
SasaTAP2b_#A   : G.....Q..H..... : 557
SasaTAP2b_#B   : G.....Q..H..... : 557
SasaTAP2b_#C   : G.....Q..H..... : 557
OnmyTAP2b_#A   : .....Q..H..... : 556
OnmyTAP2b_#B   : .....Q..H..... : 560
OnkiTAP2b      : .....Q..H..... : 557
EsluTAP2NW80   : ...I..D...K..K.....EQ...N..... : 558
EsluTAP2       : .H.I..D...K..K.....Q..... : 556
OrlaTAP2_cab   : S.T...EL...K..M.....E.....Q...E.D.....N..R..D...F.K. : 546
OrlaTAP2_HN1   : SKT...EL...K..M.....E.....Q...E.D.....N..R..D...F.K. : 546
OrlaTAP2_HdrR  : S.T...EL...K..M.....E.....Q...E..N.....N..R..D...F.K. : 546
DareTAP2a_19B  : DHN..KD...K...I...M.....T...K...D...K...S. : 546
DareTAP2d_19D  : DQAA..KD...K...M...M.....Q...Q...K...K...Q. : 552
DareTAP2e_19D  : DK...ENV.F.MK..TI...V.....A...KHL.E.S.D.F...R..KE.DPK.F.Q. : 418
XelaTAP2_I     : DAPS...NV.FK.P..SV.....T.....E..I.A...I..GE.K.Q...S. : 547
XelaTAP2_II    : .FLT.KNV.FA.P..SV.....T.....E..E.D...KS.KE.K.E...S. : 546
GagaBAG69399   : ..L...DVTF...SEV..A.LN.S...A.....E.GA..V...V..RD.E.R...Q : 543
GagaAEE25620   : ..L...DVTF...SEV..A.LN.S...A.....E.GA..V...V..RD.E.R...Q : 543
GagaBAG69412   : ..L...DVTF...EV..A.LN.S...A.....E.GA..V...V..RD.E.R...Q : 543
GagaBAG69426   : ..L...DVTF...EV..A.LN.S...A.....E.GA..V...V..RD.E.R...Q : 543
Rat TAP2_I     : .KP...LTFT.H..KV.....N.S...VAA..QNL...TG.KV...E..VQ.D...Q : 545
Rat TAP2_II    : .KP...LTFT.H..KV.....N.S...VAA..QNL...TG.QL...E..VQ.D...Q : 545
Human TAP2     : D.P...K.LTFT...EV.....N.S...VAA..QNL...TG.QV...EK.ISQ.E.C...SQ : 545
               *          500          *          520          *          540

```

```

OnkiTAP2a      : IAMVGQEPVLFSGSVKDNIAYGLADCSLERVQEAARRANAHSFISQLEKGYD TDVGERGGQLSG : 623
SasaTAP2a_#A   : .....I..... : 623
SasaTAP2a_#B   : .....I..... : 623
SasaTAP2a_#C   : .....I..... : 623
OnmyTAP2a_#A   : V.....I.....H..... : 620
OnmyTAP2a_#B   : V.....I.....H..... : 622
SasaTAP2b_#A   : V.....I.....M..... : 621
SasaTAP2b_#B   : V.....I.....M..... : 621
SasaTAP2b_#C   : V.....I.....M..... : 621
OnmyTAP2b_#A   : V.....I.....M..... : 620
OnmyTAP2b_#B   : V.....I.....M..... : 624
OnkiTAP2b      : V.....I.....M..... : 621
EsluTAP2NW80   : V.....I.....x.....KE.....x.....QG.....R.....Y.....N..... : 622
EsluTAP2       : V.....I.....K.....G.....K.....Y.....N..... : 620
OrlaTAP2_cab   : VTA.S.....L.....IR.....T.....SN.TM.EIE.....CK.....D.MK.DQ.....E.....G.....AK : 610
OrlaTAP2_HN1   : VTA.S.....L.....IR.....T.....SN.M.EIK.....CK.....D.MK.DQ.....E.....G.....AK : 610
OrlaTAP2_HdrR  : VTA.S.....L.....IR.....T.....SN.M.EIK.....CK.....D.MK.DQ.....E.....G.....AK : 610
DareTAP2a_19B  : V.....T.R.....QG.M.....K.....SK.....K.....NL..... : 610
DareTAP2d_19D  : V.....D.....R.....T.YDQKK.ED.KE.....D.CR.....T.....CL.A..... : 616
DareTAP2e_19D  : V.V.S.....E.AR.IQK.T.DSY.EDM.T.VKQ.IN.VLS.D.N.E.....H.V..... : 482
XelaTAP2_I     : V.L.A.....A.TIRE.G.K.I.EKQLK.....K.K.DT.....EM.L.....IDS.A.A..... : 611
XelaTAP2_II    : ..L.A.....A.RG.S.QNIPE.QLKA.....K.ED.EGMD.A.....DS.A.A..... : 610
GagaBAG69399   : V.L.....IR.....ME.QE.EIIA..A.G.LG..A.Q.FG.....A..... : 607
GagaAEE25620   : V.L.....IR.....ME.EE.EIIA..A.G.LG..A.Q.FG.....A..... : 607
GagaBAG69412   : V.L.....IR.....ME.EE.EIIA..A.G.LG..A.Q.FG.....A..... : 607
GagaBAG69426   : V.L.....IR.....ME.EE.EIIA..A.G.LG..A.Q.FG.....A..... : 607
Rat TAP2_I     : VVL.....R.EDAQ.MA..QA.C.DD..GEMTN.IN.EI..K.S..AV..... : 609
Rat TAP2_II    : VVL.....R.EDAQ.MA..QA.C.DD..GEMTN.IN.EI..S..AV..... : 609
Human TAP2     : VVS.....RN.....QS.EDDK.MA..QA.H.DD..QEM.H.IY.....K.S..AA : 609
                *                560                *                580                *                600

OnkiTAP2a      : GEKQRIAIARALIREPQVLILDEVTSALDTESEHMQEALA-----SCPSQTLLVIAHRLKTI : 681
SasaTAP2a_#A   : ..... : 681
SasaTAP2a_#B   : ..... : 681
SasaTAP2a_#C   : ..... : 681
OnmyTAP2a_#A   : ..... : 678
OnmyTAP2a_#B   : ..... : 680
SasaTAP2b_#A   : ..... : 679
SasaTAP2b_#B   : ..... : 679
SasaTAP2b_#C   : ..... : 679
OnmyTAP2b_#A   : .....T..... : 678
OnmyTAP2b_#B   : .....RHPSTHKLECF.L..... : 688
OnkiTAP2b      : ..... : 679
EsluTAP2NW80   : .....I.....V..... : 680
EsluTAP2       : .....V..... : 678
OrlaTAP2_cab   : S.R...G.....K.....I.S..I.N.NK.LQ.....NR.N..V...S... : 668
OrlaTAP2_HN1   : S.R...G.....K.....I.S..I.N.NK.LQ.....N.N..V...S... : 668
OrlaTAP2_HdrR  : S.R...G.....K.....I.S..I.N.NK.LQ.....NR.N..V...S... : 668
DareTAP2a_19B  : .....S.....Q.....Q.S-----C.T..... : 668
DareTAP2d_19D  : .Q.....Q.I.....S.S.....K.D.....RR.N..... : 674
DareTAP2e_19D  : .....Q.....S.....N.K.I.N.....CFAT..R.IV..... : 540
XelaTAP2_I     : .Q...V.L...A.K.KL....AS.C..A.T..EI.QS.Q-----TL.DVS..I....R... : 669
XelaTAP2_II    : .Q...V.L...A.K.KL....AS.C..A.T..EI.KS.Q-----IEDLS..I....R.V... : 668
GagaBAG69399   : .Q.....V.R.T.....A.....GD.DA.L.QWVR-----NGGDR.V.L.T.QPRML : 665
GagaAEE25620   : .Q.....V.R.T.....A.....GD.DA.L.QWVR-----NGGDR.V.L.T.QPRML : 665
GagaBAG69412   : .Q.....V.R.T.....A.....GD.DA.L.QWVR-----NGGDR.V.L.T.QPRML : 665
GagaBAG69426   : .Q.....V.H.T.....A.....GD.DA.L.QWVR-----NGGDR.V.L.T.QPRML : 665
Rat TAP2_I     : .Q...L.....V.N.R.....A.....A.C.QAL.TWR-----QEDR.M.....H.V... : 666
Rat TAP2_II    : .Q...L.....V.N.R.....A.....A.C.QAL.TWR-----QEDR.M.....H.V... : 666
Human TAP2     : .Q...L.....V.D.R.....A.....VQC.QAL.DWN-----RGDR.V.....Q.V... : 666
                *                620                *                640                *                660

```

OnkiTAP2a	: ERADQIILIDQGTVQEQGTHQELMDRKGSYKLERLFTEDDVPH-----	: 726
SasaTAP2a_#A	:A.-----	: 726
SasaTAP2a_#B	:A.-----	: 726
SasaTAP2a_#C	:A.-----	: 726
OnmyTAP2a_#A	:R.....MS.-----	: 723
OnmyTAP2a_#B	:R.....MS.-----	: 725
SasaTAP2b_#A	:R.....R.S.L.....G.....R.....TS.-----	: 724
SasaTAP2b_#B	:R.....R.S.L.....G.....R.....TS.-----	: 724
SasaTAP2b_#C	:R.....R.S.L.....G.....R.....TS.-----	: 724
OnmyTAP2b_#A	:L.....G.....R.....TS.-----	: 723
OnmyTAP2b_#B	:L.....G.....R.....TS.-----	: 733
OnkiTAP2b	:L.....G.....R.....TS.-----	: 724
EsluTAP2NW80	:L.....KV..EM.....R.....V..D..K-----	: 725
EsluTAP2	:L.....KV..EM.....R..I.....K-----	: 723
OrlaTAP2_cab	:VV..IG..K.K.K.....L..V.....E.....RTN-----	: 711
OrlaTAP2_HN1	:VV..IG..K.K.K.....L..V.....E.....RTN-----	: 711
OrlaTAP2_HdrR	:VV..IG..K.K.K.....L..V.....E.....RTN-----	: 711
DareTAP2a_19B	:V..S..ELV..K.....E.....EK.....R.....SD..KTTKQEKEKSDTVKTQ	: 725
DareTAP2d_19D	:V..K..E..L..T.....E.....N.....R.....S..EKEADEKEIKQG-----	: 726
DareTAP2e_19D	: .K.....V..K..S..S..K.....V.....KK..I.....R..SR..N..SLET-----	: 585
XelaTAP2_I	: QK...N..LVLEG..YLV.....E..VEK.--RECINGSAMGTYCKMEIMMWROVI---	: 720
XelaTAP2_II	: QK...LVLE..G..NLV.....EK..VQKQ..V..HR..LNGTSLQNSGKVN-----	: 714
GagaBAG69399	: .K...RVVLEH...A.M..PA..RT..G.P.SR.LQHKKKK-----	: 705
GagaAEE25620	: .K...RVVLEH...A.M..PA..RT..G.P.SR.LQHKKKK-----	: 705
GagaBAG69412	: .K...RVVLEH...A.M..PA..RT..G.P.SR.LQHKKKK-----	: 705
GagaBAG69426	: .K...RVVLEH...A.M..PA..RT..G.P.SR.LQHKKKK-----	: 705
Rat TAP2_I	: QN...VLVLK...QLV..H---DQ..R..EQDV..AH..VQQRLEA-----	: 703
Rat TAP2_II	: QN...VLVLK...QLV..H---DQ..R..EQDV..AH..VQQRLEA-----	: 703
Human TAP2	: Q...H...LVLE..KL..KL---AQ.-----	: 686

*

680

References:

1. McConnell SC, *et al.* (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
2. Rudolph MG, *et al.* (2002) Crystal structures of two rat MHC class Ia (RT1-A) molecules that are associated differentially with peptide transporter alleles TAP-A and TAP-B. *Journal of molecular biology* 324(5):975-990.
3. Walker BA, *et al.* (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proceedings of the National Academy of Sciences of the United States of America* 108(20):8396-8401.
4. Ohta Y, *et al.* (2003) Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in *Xenopus*: further evidence for co-evolution among MHC class I region genes. *European journal of immunology* 33(11):3017-3027.
5. Nijenhuis M & Hammerling GJ (1996) Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. *Journal of immunology* 157(12):5467-5477.

Additional file 8: Figure S5. Tapasin data

Table of Contents		Page
S5a	Phylogeny of deduced TAPBP amino acid sequences	1
S5b	Alignment of deduced TAPBP amino acid sequences	2

Figure S5a. Phylogeny of deduced TAPBP amino acid sequences

Deduced TAPBP amino acid sequences from salmonid and pike haplotypes are described in Additional file 2: Text S1 and the human reference is shown in the figure. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (1). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Sequences linked to the PSMB8F gene variant are shown with red font. Salmon TAPBPb_50 and _59 refers to the genomic location of these sequences on Atlantic salmon chromosome 14 (see main text Figure 1 and Additional file 1: Figure 1).

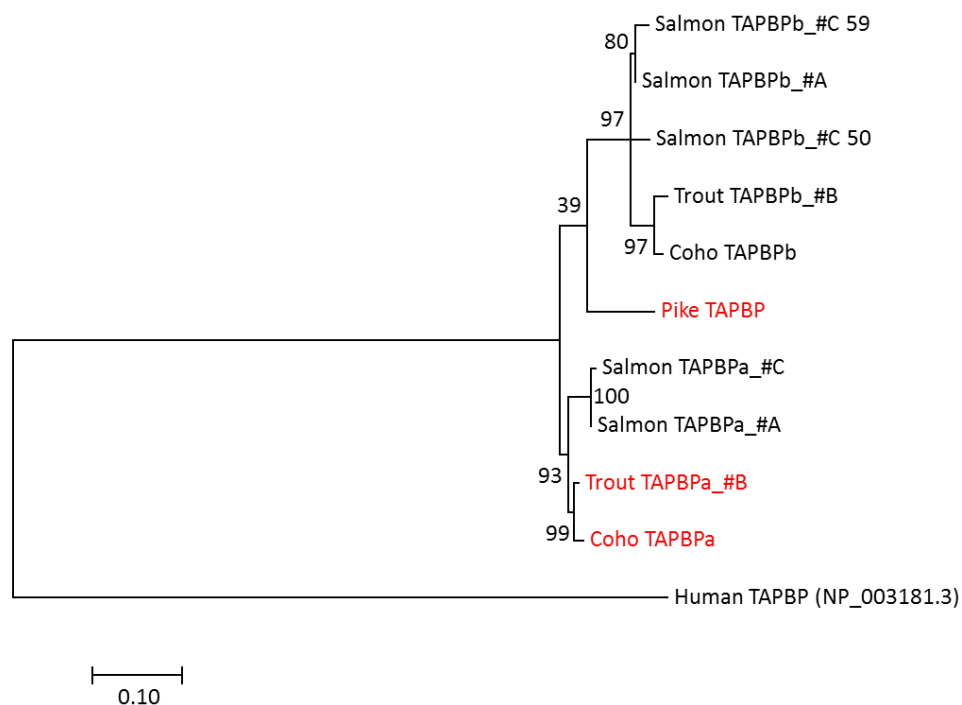


Figure S5b. Amino acid alignment of TAPBP sequences

Salmonid and Northern pike haplotype sequence references can be found in Additional file 2: Text S1. Other sequence references are: Human (Hosa) TAPBP AAG33061.1; Chicken (Gaga) TAPBP1 NP_001029988.2, TAPBP2 CAL47156.1, TAPBP3 CAL47157.1, TAPBP4 CAL47152.1, TAPBP5 CAL47155.1, TAPBP6 CAL47188; Zebrafish (Dare) Haplotype A DareTAPBP_AB strain AL672151:89.363-94.015 CAD58767.1, Haplotype B: DareTAPBP_Tue1 19: 7,119,769-7,124,430 ENSDARG00000079402, DareTAPBP_Tue2 19: 7,147,063-7,151,766 ENSDARG00000045011, DareTAPBP_CG2 GDQH01003123.1; Medaka (Orla) TAPBP_HdrR TAPBP BA000027:361.815-364.809 BAB83851.2, OrlaTAPBP_HN1 TAPBP AB183488:358.533-361.920 BAD93267.2, OrlaTAPBP_cab TAPBP AB450971:194.680-198.000 BAJ07263.1. Reference for the rainbow trout TAPBP sequence is (3), for zebrafish TAPBP sequences (4), medaka TAPBP sequences (5-7) and chicken TAPBP polymorphism (8). Numbering is according to human TAPBP. Residues known to interact with MHCI are indicated in cyan [TN3-TN6] while yellow shading shows human cysteine interacting with Erp57, lysine interacting with TAP and ER retention signal (9). Glycosylation sites are underlined. Sequences linked to the PSMB8F gene variant are shown with red font.

SasaTAPBPpa #C	:	M-ANISTIIYKLSFLAFTYFIHVY---GTSCP---VLECWFVQEKPGRRGGGFPAAMIQEKSLI	:	55
SasaTAPBPpa #A	:	.-----Q-----A.	:	55
OnmyTAPBPpa #B	:	.-----Q-----A.	:	55
OnkiTAPBPpa	:	.-----Q-----A.	:	55
SasaTAPBPpb #C	:	-T.....IQ-....S.M....A.....H...S.....	:	52
SasaTAPBPpb #A	:	-T.....IQ-....S.M....A.....H...S.....	:	52
OnmyTAPBPpb	:	-----MD-----H...S.....	:	35
OnkiTAPBPpb	:	-T.....IQ-....S.VM....A.....H...S.....	:	52
EsluTAPBP	:	-X.....III.I.TN.....R.....S.....	:	55
DareTAPBP_AB	:	-SD...VF.I.VI...L.YGSC--SG..----H...LSTP.S....I	:	55
DareTAPBP_tue1	:	-SGF.....I.IIV.LCFG.C--A.....H...LSTP.S....M	:	54
DareTAPBP_tue2	:	-SE.....I.VF.LCFG.C--SA.....H...LSTP.S....M	:	55
DareTAPBP CG2	:	-SD...VF.I.VI...L.HGF--SA.....H...LSTP.S....M	:	55
OrlaTAPBP_HdrR	:	-TDF.....LV.LWC...ACSSSSS---L.....V....LAG.TT...T..	:	58
OrlaTAPBP_cab	:	-TGF.....YLV.LWC.NACSSSSS---L.....V....LAG.TT...T..	:	58
OrlaTAPBP HN1	:	-TDF.....LV.LWC..AC---SS---L.....V....LAG.TT...T..	:	55
GagaTAPBP1	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
GagaTAPBP2	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
GagaTAPBP3	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
GagaTAPBP4	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
GagaTAPBP5	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
GagaTAPBP6	:	A.GLRL----L.GLCWSQFRVEDAA.P.PPPAPVR.ALL-.GV..R..L.GGGNARPA..	:	57
HosaTAPBP	:	KSL.L----L.....ALGLA.AVSAGPA.I.....EDAS.K-----LAKRPGA.	:	45
		* 20 * 40		
		TN4		
SasaTAPBPpa #C	:	YINTDPSEETKSQQGPSADINHDRVYYVTDPAAILCSSSLHPPEGSVHKPQCEINFPMPQPS	:	118
SasaTAPBPpa #A	:	Q.....S.....I.....	:	118
OnmyTAPBPpa #B	:	Q.....S.....I.....	:	118
OnkiTAPBPpa	:	Q.....R.....S.....I.....T.....	:	118
SasaTAPBPpb #C	:	D.....R.K..L.T.--.....T.....M.....Q.....	:	113
SasaTAPBPpb #A	:	D.....R.K..L.T.--.....T.....M.....Q.....	:	113
OnmyTAPBPpb	:	D.....R.K..PT.--.....T.....Q.S.....	:	96
OnkiTAPBPpb	:	D.....R.K..PT.--.....T.....Q.S.....	:	113
EsluTAPBP	:	WKRPE...R..T..X..L..I.....TF.....Q.....	:	118
DareTAPBP_AB	:	F.R.EAY...I.TELH.P...SSS.....GTF..AA.N..K..N..K.....HA.	:	118
DareTAPBP_tue1	:	F.R.EAF...I.TELH.P...SSS.....GTF..AA.N..K..N..K.....HA.	:	117
DareTAPBP_tue2	:	F.R.EAY...I.TELH.P...SSS.....GTF..AA.N..K..N..K.....HA.	:	118
DareTAPBP CG2	:	F.R.EAY...AETH.P...SSS.....GTF..AA.N..K..N..K.....HA.	:	118
OrlaTAPBP_HdrR	:	H.RK.ERDDGAA.SRA.-S...P..I.VI...ET..HR.FN..R..IK....V...L....	:	120
OrlaTAPBP_cab	:	H.RK.EHDDGAA.SRA.-S...P..I.VI...ET..HR.FN..R..IK....V...L....	:	120
OrlaTAPBP HN1	:	H.RK.ERDDGAA.SRA.-S...P..I.VI...ET..HR.FN..R..IK....V...Q....	:	117
GagaTAPBP1	:	RFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
GagaTAPBP2	:	RFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
GagaTAPBP3	:	RFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
GagaTAPBP4	:	RFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
GagaTAPBP5	:	RFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
GagaTAPBP6	:	LFGG-----AETP.EPGPEPEVTFN.S.WGT.TPLGV-.R---TP.S..L..TN..TG	:	109
HosaTAPBP	:	LLRQG-----GEPP.RP.LDPELYLS.H...GA.QAAFRRY.R.-APA.H..MSRV.L.A	:	101
		* 60 * 80 * 100		
		TN3		

SasaTAPBP _a #C	: TVQWVVPLTDSAHSPIYLQADWYSAALQGLDQGLGLSSVMRAPATATKEP-----TVVLSVS	: 174
SasaTAPBP _a #A	:	: 174
OnmyTAPBP _a #B	:A.....A.....	: 174
OnkiTAPBP _a	:A.....A.....	: 174
SasaTAPBP _b #C	:T.....T.....F.....R.....N.....M.....-----K.L.....	: 169
SasaTAPBP _b #A	:T.....T.....F.....R.....N.....-----K.L.....	: 169
OnmyTAPBP _b	:T.....T.....F.....R.....N.L.....-----K.....	: 152
OnkiTAPBP _b	:T.....T.....T.T.....F.....R.....N.L.....-----K.....	: 169
EsluTAPBP	: S...T.S.....S.....N.....S.....-----A.L.....	: 174
DareTAPBP_AB	: M.R.ASA.....Q.V.....F.V.A...D..T.NI....S.S.....-----K.I.....	: 174
DareTAPBP_tue1	: M.R.ASA.....Q.V.....F.V.A...E..T.NI....L.S.D.....-----K.I.....	: 173
DareTAPBP_tue2	: M.R.ASA.....Q.V.....F.V.A...E..T.NI....S.S.....-----K.I.....	: 174
DareTAPBP_CG2	: M.R.ASA.....Q.V.....F.V.A...E..T.NI....S.S.....-----E.I.....	: 174
OrlaTAPBP_HdrR	: SLA..RS..EN.F..V.....F..SFH.V.E.R.VATIT...SG...H-----S.I...LT	: 176
OrlaTAPBP_cab	: SLA..RS..EN.F..V.....F..SFH.V.D.R.VATIT...SG...H-----S.I...L	: 176
OrlaTAPBP_HN1	: SLA..RS..EN.F..V.....F..SFH.V.DER.VATIT...SG...H-----S.I...L	: 173
GagaTAPBP1	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.EGT----ITAA.A.A.L	: 166
GagaTAPBP2	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.EGT----ITAA.A.A.L	: 166
GagaTAPBP3	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.EGT----ITAA.A.A.L	: 166
GagaTAPBP4	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.EGT----ITAA.A.A.L	: 166
GagaTAPBP5	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.EGT----ITAA.A.A.L	: 166
GagaTAPBP6	: SDP.SR...HPD.R...PTAGGQ.WV..VG--TP.Y.VTALLQGGMG.ERT----ITAA.A.A.L	: 166
HosaTAPBP	: SAK.ASG...PAQN C ..RA.DGA.LMVSIS--SPV.S...LL.PQPEPQQEPVLITMA....T.L	: 162
C115 (C95 mature protein) human ERp57 bond		
SasaTAPBP _a #C	: SRTPLVRSRLGEPVVLDCGFW-MEATSPLSGS-----GFAVEWRYQFRGDGRLLVLAYDGKTD	: 230
SasaTAPBP _a #A	:	: 230
OnmyTAPBP _a #B	:-A.TS.....	: 230
OnkiTAPBP _a	:-A..S.....	: 230
SasaTAPBP _b #C	: ...FMI.....-VD.S.....	: 225
SasaTAPBP _b #A	: ...FMI.....-VD.S.....	: 225
OnmyTAPBP _b	: ...FMI.....-VD.S.....	: 208
OnkiTAPBP _b	: ...FMI.....-VD.S.....	: 225
EsluTAPBP	: ...V.....R.....-A..S.....-T.....N.....	: 230
DareTAPBP_AB	: .K..V..C.....L.....-IDPS..H.....-..SI.....E.....N.....	: 230
DareTAPBP_tue1	: .K..V..C.....L.....-VDPS..H.....-..SI.....E.....N.....	: 229
DareTAPBP_tue2	: .K..V..C.....L.....-IDPS..H.....-..SI.....E.....N.....	: 230
DareTAPBP_CG2	: .K..V..C.....L.....-IDPS..H.....-..SI.....E.....N.....	: 230
OrlaTAPBP_HdrR	: .T.MTAQA...Q.FT...S.-ADPL...ET-----	: 232
OrlaTAPBP_cab	: .T.MTAQA...Q..T...S.-ADPL...ET-----	: 232
OrlaTAPBP_HN1	: .T.MTAQA...Q..T...S.-ADPL...ET-----	: 229
GagaTAPBP1	: TH..AL.A.V.S.IH.H.A.A-APPS-----P.VL...H.N..A..VL...SS.A	: 216
GagaTAPBP2	: TH..TL.A.V.S.IH.H.A.A-APPS-----P.VL...H.N..A..VL...SS.A	: 216
GagaTAPBP3	: TH..TL.A.V.S.IH.H.A.A-APPS-----P.VL...H.N..A..VL...SS.A	: 216
GagaTAPBP4	: TH..TL.A.V.S.IH.H.A.A-APPS-----S.VL...H.N..A..VL...SS.A	: 216
GagaTAPBP5	: TH..TL.A.V.S.IH.H.A.A-APPS-----S.VL...H.N..A..VL...SS.A	: 216
GagaTAPBP6	: TH..TL.A.V.S.IH.H.A.A-APPS-----P.VL...H.N..A..VL...SS.A	: 216
HosaTAPBP	: TH..AP.V...QDAL..LS.AY.PP..EAAS.LAPGPPP.GL T ..R.H L ..K..H.L..ATPGLN	: 225
TN6 TN7		
SasaTAPBP _a #C	: RFAETKEKRAGLDFTALHET-----GNASLILQEAQVRHTGTNICTVYLPYLLAQVAVELE	: 286
SasaTAPBP _a #A	:	: 286
OnmyTAPBP _a #B	:G.....E.....Y.....	: 286
OnkiTAPBP _a	:G.....E.....Y.....A.....	: 286
SasaTAPBP _b #C	:IQ.EG.....E.....S..Y.....	: 281
SasaTAPBP _b #A	:Q.EG.....E.....S..Y.....	: 281
OnmyTAPBP _b	:Q.EGV.....D.....S..Y.....	: 264
OnkiTAPBP _b	:Q.EGS.....D.....S..Y..M.....	: 281
EsluTAPBP	: .L...Q...G.E.....R-----S..Y.....I.X.....	: 286
DareTAPBP_AB	:S.SG.EM.I.G.YQ-----E.S...S..Y.....H.....D..	: 286
DareTAPBP_tue1	:S.SG.EM.I.G.YQ-----V...E.S...S..Y.....H.....D..	: 285
DareTAPBP_tue2	:S.SG.EM.I.G.YQ-----E.S...S..Y.....H.....D..	: 286
DareTAPBP_CG2	:S.SG.EM.I.G.YQ-----E.S...S..Y.....H.....D..	: 286
OrlaTAPBP_HdrR	: .L.D.L.EG.T...E...R-----TK.Q.S..Y..S.F.....M...	: 288
OrlaTAPBP_cab	: .L.D.L.EG.T...E...R-----TK.Q.S..Y..S.F.....M...	: 288
OrlaTAPBP_HN1	: .L.D.L.DG.T...E...R-----TK.Q.S..Y..S.F.....M...	: 285
GagaTAPBP1	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
GagaTAPBP2	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
GagaTAPBP3	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
GagaTAPBP4	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
GagaTAPBP5	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
GagaTAPBP6	: -.PRATPG.E.LLGTRDGD---GVTAVT.R.AQPSPGDE..Y..S.F..HGHT.TVLQ.H	: 273
HosaTAPBP	: GQMPAAQEG.-VA.A.WDDDEPWGPWT..GTFW.PRV.PFQ E ..Y L A.IH...Q G ..TL.A	: 287
TN5 TN6		
Glycosylation site N253 (N233 mature protein)		


```

SasaTAPBPa_#C : IVEPPSLSIFPSPLPLSVPGQV--VKVQCEASGFFPL-SLDFHWELT-GPDGKVRP----LGQ : 341
SasaTAPBPa_#A : .....VM.....-.....-..... : 343
OnmyTAPBPa_#B : .....E.Y.....-..... : 341
OnkiTAPBPa : .....E.Y.....-..... : 341
SasaTAPBPb_#C : .....M.....-T.....Y.....E...V...G.R.S.---- : 336
SasaTAPBPb_#A : .....M.....-T.....Y.....E...V...G.R.S.---- : 336
OnmyTAPBPb : ..... : 266
OnkiTAPBPb : .....M.....-T.....Y.....E.R.M.....-..... : 336
EsluTAPBP : .....X.....-T.....S.....EL...M-ET..N..... : 341
DareTAPBP_AB : .....L.....L.....-LS.....A.H-T..LG..F..A...SLS---- : 341
DareTAPBP_tue1 : .....L.....L.....-LS.....A.H-T..LG..F..A...SLS---- : 340
DareTAPBP_tue2 : .....L.....L.....-LS.....A.H-T..LG..FI..A...ALS---- : 341
DareTAPBP_CG2 : .....L.....L.....-LS.....A.H-T..LG..F..A...SLS---- : 341
OrlaTAPBP_HdrR : V.....H.....A.....-LNI.....V...-PV.LS..FR..A...S.S----.S : 343
OrlaTAPBP_cab : V.....H.....A.....-LNI.....V...-PV.LS..FR..A...S.S----.S : 343
OrlaTAPBP_HN1 : V.....H.....A.....-LNI.....V...-PV.LS..FR..A...S.S----.S : 340
GagaTAPBP1 : VF...KVTLS.KN.-VVA..MS--AELR.HV...Y...-DVTVT.QRRA.SS.TSQSPRDTVMD : 332
GagaTAPBP2 : VF...KVTLS.KN.-VVA..TS--AELR.HV...Y...-DVTVT.QRRT.SS.TS.SPRDTVMD : 332
GagaTAPBP3 : VF...KVTLS.KN.-VVA..TS--AELR.HV...Y...-DVTVT.QRRT.SS.TS.SPRDTVMD : 332
GagaTAPBP4 : VF...KVTLS.KN.-VVA..MS--AELR.HV...Y...-DVTVT.QRRA.GS.TS.SPRDTVMD : 332
GagaTAPBP5 : VF...KVTLS.KN.-VVA..TS--AELR.HV...Y...-DVTVT.QRRA.GS.TSQSPRDTVMD : 332
GagaTAPBP6 : VF...KVTLS.KN.-VVA..MS--AELR.HV...Y...-DVTVT.QRRA.GS.TS.SPRDTVMD : 332
HosaTAPBP : VYK..KV.LM.AT.ARAA..EA-PELLL.LV.H.Y.SGG.EVE...RG..G.RSQKA---E.. : 346
          *          300          *          320          *          340

```

```

C362
SasaTAPBPa_#C : GSVTGHRRQGFNDTYSQTSRLLEDSAKLDLGRGGEVTCVAVHP---GGTR--RASVTLNIVIGIN : 399
SasaTAPBPa_#A : ..... : 401
OnmyTAPBPa_#B : .....T.....-.....-.....V. : 399
OnkiTAPBPa : .....T.....-.....-.....V. : 399
SasaTAPBPb_#C : .....S.....V.....R---.....IT.VS : 394
SasaTAPBPb_#A : .....S.....V.....LR---.....IT.VS : 394
OnmyTAPBPb : ..... : 270
OnkiTAPBPb : .....S.F.....-.....H---.....IA.VS : 392
EsluTAPBP : ...S.....G.....T..Q.....-.....-.....R.....VS : 399
DareTAPBP_AB : .....RAS.G.F..S.....T..R.A...IS...K.D---.....AA.....VA : 399
DareTAPBP_tue1 : .....RAS.G.F..S.....T..R.A...IS...K.D---.....AA.....VA : 398
DareTAPBP_tue2 : ...V...RAS.G.F..S.....T..R.A...IS...K.D---.....AA.....VG : 399
DareTAPBP_CG2 : .....RAS.G.F..S.....T..R.A...IS...K.D---.....AA.....VA : 399
OrlaTAPBP_HdrR : ..MS...AW.G...ST..Q..TS.....IS.....R---.....S...FS : 401
OrlaTAPBP_cab : ..MS...AW.G...ST..Q..TS.....IS.....R---.....S...FS : 401
OrlaTAPBP_HN1 : ..MS...AW.G...ST..Q..TS.....IS.....R---.....S...FS : 398
GagaTAPBP1 : SWTS...AA.G...R.AAAR.IP.RPQ-HH.DVYS..VT.T--ALAKPM.V..R.LLA.TE : 391
GagaTAPBP2 : SWTS...AA.G...R.AAAR.IP.RPQ-HH.DVYS..VT.T--ALAKPM.V..R.LLA.TE : 391
GagaTAPBP3 : SWTS...AA.G...R.AAAR.IP.RPQ-HH.DVYS..VT.I--ALAKPM.V..R.LLA.TE : 391
GagaTAPBP4 : SWTS...AA.G...R.AAAR.IP.RPQ-HH.DVYS..VT.T--ALAKPM.V..R.LLA.TE : 391
GagaTAPBP5 : SWTS...AA.G...R.AAAR.IP.RPQ-HH.DIYS..VT.T--ALAKPM.V..R.LLA.TE : 391
GagaTAPBP6 : SWTS...AA.G...R.AAAR.IP.HPQ-HH.DVYS..VT.T--ALAKPM.V..R.LLA.TE : 391
HosaTAPBP : RWLSAL.HHS.GSV.LSGH.QPPPVTTTE-QH.ARYA.RIH..SLPASG--S.E...E.A.LS : 406
          *          360          *          380          *          400

```

```

SasaTAPBPa_#C : GPSIEDSMAMVAVALGLYGLIKIVSWTF--SSGSDDTNSQEK-----KVK : 442
SasaTAPBPa_#A : ..... : 444
OnmyTAPBPa_#B : .....A..... : 442
OnkiTAPBPa : ..... : 442
SasaTAPBPb_#C : A....Y.....I.....V.....VAYT... : 437
SasaTAPBPb_#A : A....Y.....I.....V.....VAYT... : 437
OnmyTAPBPb : ..... : -
OnkiTAPBPb : A....Y.....I.....QV.....VA.T...FTHIGVIIILSQLQ... : 449
EsluTAPBP : A.....C.....VI.....I.....E.AEKKA..... : 440
DareTAPBP_AB : A.....L...M..FL...S..D.G.SELND.....E- : 442
DareTAPBP_tue1 : A.....L...M..FL...S..D.G.SELND.....E. : 442
DareTAPBP_tue2 : A.....L...M..FF...S...AELND.....E. : 439
DareTAPBP_CG2 : A.....L...M..FL...S..D.G.SELND.....E. : 443
OrlaTAPBP_HdrR : S.....G...V.....FA...T.....EAAKPN.....D. : 444
OrlaTAPBP_cab : S.....G...V.....FA...T.....EAAKPN.....D. : 444
OrlaTAPBP_HN1 : S.....G...V.....FA...T.....EAAKPN.....D. : 441
GagaTAPBP1 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
GagaTAPBP2 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
GagaTAPBP3 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
GagaTAPBP4 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
GagaTAPBP5 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
GagaTAPBP6 : ..HL..ITGLFL..FV.C...R---LY-PKAARPKETK.....SQ : 430
HosaTAPBP : ...L...VGLFLS.FL.L..F..ALG.AA-VYL.TCKD.KK.....AE : 448
          *          420          *          440

```

TAP association ER retention signal

References:

1. Whelan S & Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution* 18(5):691-699.
2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. Landis ED, *et al.* (2006) Identification and regulatory analysis of rainbow trout tapasin and tapasin-related genes. *Immunogenetics* 58(1):56-69.
4. McConnell SC, *et al.* (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(34):E5014-5023.
5. Matsuo M, Asakawa S, Shimizu N, Kimura H, & Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). *Immunogenetics* 53:930-940.
6. Nonaka MI & Nonaka M (2010) Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics* 62(5):319-332.
7. Tsukamoto K, *et al.* (2005) Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. *Immunogenetics* 57:420-431.
8. van Hateren A, *et al.* (2013) A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. *The Journal of biological chemistry* 288(45):32797-32808.
9. Dong G, Wearsch PA, Peaper DR, Cresswell P, & Reinisch KM (2009) Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* 30(1):21-32.

Additional file 9. Figure S6. Calnexin and calreticulin data

Table of Contents		Page
S6a	Phylogeny of deduced CANX, CALR, CALRL and CLGN amino acid sequences	4
S6b	Alignment of deduced CANX and CLGN amino acid sequences	1
S6c	Alignment of deduced CALR and CALRL amino acid sequences	5

Figure S6a. Phylogeny of deduced CANX, CALR, CALRL and CLGN amino acid sequences

The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel 2008 model (1). The tree with the highest log likelihood (-10176.0329) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Sequence references are shown in parenthesis and Atlantic salmon sequences are shown using red font. Atlantic salmon sequences are also presented in Additional file 2: Text S1. CALR is calreticulin, CALRL is calreticulin-like, CANX is calnexin and CLGN is calmegin.

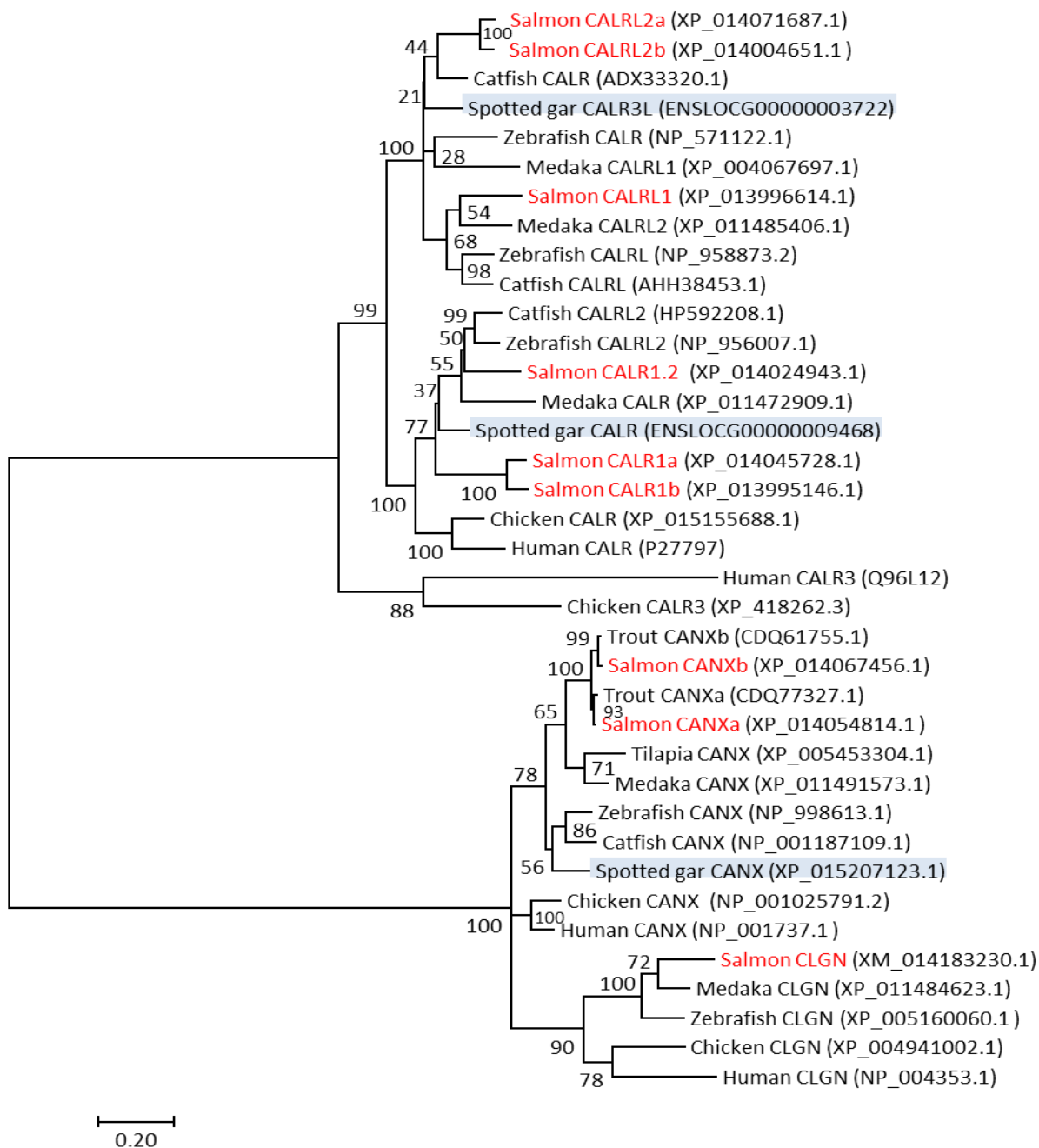


Figure S6b. Alignment of deduced CANX and CLGN amino acid sequences

Sequence references are shown in Fig. S6a. See Fuller et al. (3) for the catfish CANX gene sequence. Residues are color coded according to physiochemical properties and numbering is consecutive. Hosa is human, Gaga is chicken, Sasa is Atlantic salmon, Onmy is rainbow trout, Orla is medaka, Dare is zebrafish, Icpu is catfish and Leoc is Spotted gar. Numbering on top is consecutive. Amino acid residues are colour coded according to physiochemical properties.

		*	20	*	40	*	60											
HosaCANX	:	--	MEGKWL-LC-MLLVLG-TAIVE	----	AHDGHD	-----	DD--VIDIEDDLDDVIE	----	: 40									
GagaCANX	:	--	.M...-YVT..A...-LAAQ	-----	E..VG	-----	.GD.....GV	----	: 43									
SasaCANXa	:	--	.LNV	--	VV..AV	.LCSTLLLTTVAAHQEEEE	-PIMELAE	---	MGV..E.E.LGL--GE : 54									
OnmyCANXa	:	--	.LNV	--	VV..AV	.LCSTLLLTTVAAHQEEEE	-PIMELAE	---	MGVD.E.E.LGL--GE : 54									
SasaCANXb	:	--	.L.VR	--	YVL..AV	.LWSTLLLTTVTAHQ.EE	.EPIVEMGG	---	M.V...MEELDH--GE : 55									
OnmyCANXb	:	--	.L.VR	--	YVL..AV	.LWSTLLLTTVTAHQEEEE	.EPIMEMGG	---	M.V...EMEELDH--GE : 55									
OrlaCANX	:	--	.HQRVT	--	FLLVAAGLLFLSLA	-PVSWAEGGLEE	----	GLD...-L.V..E..LGLAGTDD : 53										
DareCANX	:	--	.L.MR	--	VA..S	.SLCLLLM-GPVR	AQE	EEEEA	-----	DME--M.V..AI..MQ----- : 46								
IcpuCANX	:	--	.L.VR	--	VL..A	.SVCVT	LQ--VRAQE	EDE	-----	E...-VHV...GNLG--DE : 45								
LeocCANX	:	QT	.L..R	--	ML..A	.W	.STLV	-----	QA	.SE	-----	T...-V...VAG	-----	: 43				
SasaCLGN	:	--	.KLW	.GWA	.VL	.S	.LALS	VSQ	---	AEAQD	-----	L...-LEM	.EAMV	EDMVEDRE	: 46			
HosaCLGN	:	--	.HFQAFW	.LG	.S	.FISI	-----	NAEFM	-----	...-VET	...-FEE	: 33						
		*	80	*	100	*	120											
HosaCANX	:	-----	EVE-DSK	PD	TTAPPSSPKVTYKAPVPTGEVYFADSFDRG	TL	SGWILSKAKKDD	: 92										
GagaCANX	:	-----	.I...N	.ESST	.PA	R	VE	.K...D	: 95									
SasaCANXa	:	--	ELLDG	---	...PEDA	DT	PPG	.PA	E	M..HF	.E	DS	V....E	: 112				
OnmyCANXa	:	--	ELLDG	---	...PEDA	DT	PPG	.PA	E	M..HF	.E	DS	V....E	: 112				
SasaCANXb	:	--	ELLDG	---	...ADM	PPG	.V	V	E...HF	.E	M	DS	V....E	: 110				
OnmyCANXb	:	--	ELLDG	---	...ADM	PPG	.PV	E	HF	.E	M	DS	V....E	: 110				
OrlaCANX	:	EGEGLEE	---	.G...EA	.PAPKT	.PV	E	V..HF	.E	D	V....A	: 112						
DareCANX	:	----	EE	----	DI	----	EEEQKAP	A	T....E	M..H	.EA	.K...Q	.V..Q...G	: 100				
IcpuCANX	:	--	ELLDG	DG	--	.DLEDEEK	P	PTAA	T....E	K..HF	.E	.K...D	V..Q...G	: 105				
LeocCANX	:	----	---	---	EEE	.KQAS	.A	E	.A	.DF	.E	D	E	: 93				
SasaCLGN	:	DSAEIEGEA	EM	.G	----	AEGEK	G	.DANV	.FQ...T	...I	SET	.D	.S	.DR	.QV	.TM	.E	: 109
HosaCLGN	:	N	-----	---	SEEI	.VNESEL	.SEIK	.T	.Q	.I	ET	.S	.R	.A	.V		: 83	
		*	140	*	160	*	180	*										
HosaCANX	:	TDDEI	AKYDGKWEVEEMKESK	LP	GD	KGLVLM	SR	AKHHAISAKLNK	PFL	FD	TKPLIVQYEVNFQ	N	: 156					
GagaCANX	:		QD	.DT	VT	S	.S..V	I		: 159								
SasaCANXa	:	I	.ED	D	.D	K	Q	.LR	.I	I	Q	: 176						
OnmyCANXa	:	I	.ED	D	.D	K	Q	.LR	.I	Q	: 176							
SasaCANXb	:	I	.ED	D	.DG	K	Q	.LR	.I	Q	: 174							
OnmyCANXb	:	I	.ED	D	.DG	K	Q	.LR	.I	Q	: 174							
OrlaCANX	:	A	.E	A	.D	K	R	.LR	.T	S	: 176							
DareCANX	:	I	.ED	QD	K	.K	L	.LR	.T	T	: 164							
IcpuCANX	:	I	.ED	DT	K	L	.LR	.T	T	: 169								
LeocCANX	:	I	.E	Q	VD	K	L	.V	T	: 157								
SasaCLGN	:	A	S	.QL	.N	.V	M	K	ASL	.DR	RDDA	.V	D	: 173				
HosaCLGN	:	M	.E	.SI	R	.I	.L	.NQV	.R	K	V	.A	.I	.AD	D	: 147		
		200	*	220	*	240	*											
HosaCANX	:	GIECGGAYVKLLSK	TP	ELNLDQ	FHDK	TPYT	IMFG	PD	KCGEDYKLHFIFRHK	NP	KTGIYEEKHAK	: 220						
GagaCANX	:		QD	.DT	VT	S	.S..V	I		: 223								
SasaCANXa	:	.D	Q	.D	E	.V	E	E		: 240								
OnmyCANXa	:	.D	Q	.D	E	.V	E	E		: 240								
SasaCANXb	:	.D	Q	.D	E	.V	E	E		: 238								
OnmyCANXb	:	.D	Q	.D	E	.V	E	E		: 238								
OrlaCANX	:	.D	TE	V	.R	E	E			: 240								
DareCANX	:	.D	Q	.D	.D	.EE	.V	E	EF		: 228							
IcpuCANX	:	.D	A	QSA	E	.V	E	E		: 233								
LeocCANX	:	.D	Q	.D	V	E	E			: 221								
SasaCLGN	:	.D	I	DEDD	.D	.E	.N	.R	.S	RS	.LNKDL		: 237					
HosaCLGN	:	.D	I	AD	.DD	.I	.EN	.Y	.S	.I	H	VF		: 211				

```

      260          *          280          *          300          *          320
HosaCANX : RPDADLKTYFTDKKTHLYTLILNPDNSFEILVDQSVVNSGNLLNDMTTPVNPSPREIEDPEDRKRP : 284
GagaCANX : .....M.....S.....P.....N.Q.. : 287
SasaCANXa : K...R..Y.....VV.....V...T.....T.....I..AA....D.H.. : 304
OnmyCANXa : K..S..R..Y.....VV.....V...T.....T.....I..AA....D.H.. : 304
SasaCANXb : K...R..Y.....VV.....V...A.....T.....I..AA....D.H.. : 302
OnmyCANXb : K..S..R..Y.....VV.....V...A.....T.....I..AA....D.H.. : 302
OrlaCANX : K...RS.....FS.VV.....V...T.....K.....PA....H.. : 304
DareCANX : K..S..RS.Y.....V...T...I..T.....S...V.....PA....D.H.. : 292
IcpuCANX : KA.S..R..Y.....V...I..T.....A...PA....D.H.. : 297
LeocCANX : K...R..Y.....V...I..T.....VS.A...PA....D... : 285
SasaCLGN : .A.V...KFY.....V.....Y.....S.SR.S..H.VV....TK..D..T.S.. : 301
HosaCLGN : P..V...KF...R.....VM...DT..V...T...K.S..E.VV..IK.PK.....N.K.. : 275

      *          340          *          360          *          380
HosaCANX : EDWDERPKIPDPEAVKPDDEWDEDAPAKIPDEEATKPEGWLDDEPEYVPDPDAEKPEDWDEDMDG : 348
GagaCANX : .....D.....A..N.V.....A..... : 351
SasaCANXa : .....Q..D...E.....KQ...D.V..D...Q..TS...V..... : 368
OnmyCANXa : D.....Q..D...E.....KQ...D.V..D...Q..TS...V..... : 368
SasaCANXb : .....Q..D...E.....KQ...D.V..D...S..TS...V..... : 366
OnmyCANXb : .....Q..D...E.....KQ...D.V..D...S..TS...V..G..... : 366
OrlaCANX : .....Q..D...E.....Q...DN.V..D.....IG...V..D..... : 368
DareCANX : .....Q..D...E.....A..D.V..D.....IS...V..... : 356
IcpuCANX : D.....D.....A..D.V..D.....IS...V..... : 361
LeocCANX : .....D...E.....S.....I..... : 349
SasaCLGN : .....A.....N...E...Q..PD.V.....F.A..A.N..D...E... : 365
HosaCLGN : .E...A.....S...E...SE..Q..E.SSVV..A.....KFI..N...D..N..T.. : 339

      *          400          *          420          *          440
HosaCANX : EWEAPQIANPRCESAPGCGVWQRPVIDNPNYKKGWKPPMIDNPSYQGIWKPRKIPNPDFFEDLE : 412
GagaCANX : .....K.....T...M.....VN..... : 415
SasaCANXa : .....P.AL..T...A...M.....A.....N..V.....A.....H : 432
OnmyCANXa : .....P.AL..T...A...M.....A.....N..V.....A.....H : 432
SasaCANXb : .....VP.AL..T...A...M...S.....A.....N..V.....A..A...H : 430
OnmyCANXb : .....VP.AL..T...A...M...S.....A.....N..V.....A..A...H : 430
OrlaCANX : .....P..A.....E.K..M...H...A.....N.....A..... : 432
DareCANX : .....P..V..T...A..E..M.....S.....N..V.....H : 420
IcpuCANX : .....P.SA.....K.E..M.....V...N..V.....Y...H : 425
LeocCANX : .....P..K..T.....A.....N..V.....H : 413
SasaCLGN : .....VP..V..T...Q..KP.T.N..Q.....A.LV...H...V.S...Q...Y...A : 429
HosaCLGN : .....L..A..RI--..E.KP.M...K...V.R..LV...N...S.....Y...DH : 401

      *          460          *          480          *          500          *
HosaCANX : PFRMTPFSAIGLELWSMTSDIFFDNFIICADRRIVDDWANDGWGLKKAADGAAEPGVVGQMIEA : 476
GagaCANX : ..K.....V.....TE.AVA.....MA : 479
SasaCANXa : .....N.V.....S.....F.TNE.HTAER.....E.....L.N..MT : 496
OnmyCANXa : .....N.V.....S.....F.TNE.HTAER.....E.....L.N..MT : 496
SasaCANXb : .....N.V.....S.....F.TNE.HTA.R.....E..L...L.N..MT : 494
OnmyCANXb : .....N.V.....S.....F.TNE.HTA.R.....E.....L.N..MT : 494
OrlaCANX : .....A...V.....F.TN..NTA.R..A.....E.....LAA..LT : 496
DareCANX : .....V.....S.....F.TS..NVAER.....E.....L.N..T : 484
IcpuCANX : .....AV..V.....S.....F.TS..NVA.K..EE.....E.....L.N..MS : 489
LeocCANX : .....V.....S.....M.SS.AVA.Q.....E.....FS..LTM : 477
SasaCLGN : .....K..S.R.V.....E.Y.....TSYKEVA.R..G.S.....LVAS.N...L.S.LTM : 493
HosaCLGN : ..LL.S...L.....Y.....SEKEVA.H..A...RW.IMIAN.NK...LK.LMA : 465

      520          *          540          *          560          *
HosaCANX : AEERPWLWVVIILTVALPVFLVILFCCSGKKQT---SGMEYKKTDAPQPDVKE--EE--EEKEE : 533
GagaCANX : .....I..F.....V.....P---AA.....VS--D.--R.E.- : 535
SasaCANXa : ..D.....V...V.LI.IVV.F.T...AAAP-AAAD...E.....--KA.- : 554
OnmyCANXa : ..D.....V...V.LI.I.V.F.T...AVAP-AAAD...E.....--E.KA.- : 555
SasaCANXb : ..D.....V...V.LV.I.V...T...T---AAAD...E.....--VV.KA.- : 551
OnmyCANXb : ..D.....V...V.LV.I.V...T...TAA--AAAD...E.....--VV.KA.- : 553
OrlaCANX : .....I..V...G..LV.LFV.....S.A--PAD...E.....--E.- : 553
DareCANX : .....I..V...LV.I.V...T...SSASTPAAK...E.....--A--E.- : 543
IcpuCANX : .....V.....V.IFV...T...KPAT-AA...E.....--E.- : 547
LeocCANX : .....I.....VV.....KP---AA...E.....D--Q--E.- : 533
SasaCLGN : .....G..IG.TV...WPK.---SDDDYV...V.P.K..E.--DDD- : 548
HosaCLGN : ..GH...LI.LV.AGV.IA..ITS..WP-R.VKKKHKDT...ICI.QT.GVL.Q--...- : 525

```

```

      580          *          600          *          620          *          640
HosaCANX : EKDKG-----DEEEE-----GEEKLE-EKQKSDAEEDGGTVSQ----- : 565
GagaCANX : -...-----K...E-----EAN.....-...N...-I.SA-----EE-E : 570
SasaCANXa : -E.Q-----VK..KS-----QPAAA.....DSPAEKEE-----EEEEEE : 589
OnmyCANXa : -E.Q-----VK..KS-----QPEAAG.....DSPAEKE----- : 583
SasaCANXb : -A.Q-----VK..KS-----QPAA-..-N....DSPAEKEE-----VNEEEE : 585
OnmyCANXb : -A.Q-----VK..KS-----QPAAE...-N....DSPA--EE-----VNEEEE : 586
OrlaCANX : -----E...K-----SPAA-..-N-----DE-D : 569
DareCANX : -----AKD-----PEA--...EE.DST.DGDG-----DDTA : 569
IcpuCANX : -----E...ARG-----ES--...A....SPAEEA-----PE-E : 575
LeocCANX : -.ED.EEEEKE.G...E-----KED...P.-...G..GA.EH-----EE-E : 575
SasaCLGN : -EEEE-----E...KGDQAAEAKATESLTPA.-A.-.EK...ET.AAGGDSVEDEEDEEEEE : 604
HosaCLGN : -EKAALKEKPMDL...KK---QNDGEMLEK...ES.P.E-..EE.IE-----I : 566

      *          660          *          680          *
HosaCANX : ----EEED-----RKPK---AEDEEI-----LNRSPRNR-KPRRE : 592
GagaCANX : EEEE.....-..HA---S..E.T-----V.....-...KD : 601
SasaCANXa : EEEE...EEEEEAADVNEEEEAATDQ---VRDKTT-----TR----- : 622
OnmyCANXa : -EEE...EEEEEAADVNEETEATDQ---KQ...DV-----R.....T..G.KD : 626
SasaCANXb : DEEE..DEVTEE-----VREQ----- : 601
OnmyCANXb : -EEE...EVTEE-----NFLKSRES---QKPLDV-----CPM----- : 614
OrlaCANX : DEEE.QNE-----AADE---QKLEDE-----R..R...-T.KD : 599
DareCANX : ENGDKDDN-----TSNE---KS..D-----R.....-S.KD : 600
IcpuCANX : EEADKQ...-----KSDE---KQ...DV-----R....SKARA.KD : 607
LeocCANX : EEGERDADKTA-----KVAEEPLKNEKKEDV-----RQ.....-...KD : 615
SasaCLGN : DEGE...ETKSN-----EAASDDQ---K..ADEGGHSGVDGHKQ.V.K.-RV.KD : 650
HosaCLGN : IEGQ...SNQSNK-----SGSEDE---MK.ADESTGS-GDGPVK.V.K.-RV.KD : 610

```

Figure S6c. Alignment of deduced CALR and CALRL amino acid sequences

Sequence references are shown in Figure S6a. Numbering below the alignment relates to the mature human CALR sequence. For catfish CALR sequences, see (4). Residues are color coded according to physiochemical properties. Cyan shaded residues are involved in glycan binding (5) and acidic region and ER-retention signals are shaded yellow (6).

```

SasaCALR1a : ----MLVSVLL--MIALASAKPSVYFREQFED-DAWNTRWVESSHRSYDGKFVL : 47
SasaCALR1b : ----.RG.M.FSAL.....E..L..K.....G...T..... : 50
SasaCALR1.2 : --MTTMLIL.MTVLV.SIFGES.....E...G...KS.....K.....H : 52
SasaCALRL1 : --MQV.G.FAIIIS--IF.VHST...Q...L.G...K...LD.K.KA...EWK. : 50
SasaCALRL2a : --MRVA.AI.AVFASVAVTIDAT...K...Q.G...KS..LV.E.K...EWK. : 52
SasaCALRL2b : --MRVA.AIFSVFASVAVTIDAT...K...Q.G...KS..LV.K.KT...EWK. : 52
OrlaCALRL1 : --MKFCSAL.ALFA--GVAVEAT.....E.T.G.G.KS..L..K.K...QWK. : 50
OrlaCALR3 : --MNT.SLFVGLIA--.YCVHAKI....E.L.G.E.RS...N.K.K...EWK. : 50
OrlaCALRL2 : ----.GG-----LLDIDR.S----- : 11
DareCALR : --MRITAA.CFISA-LAFI.HAD...K...L.G.G.KS.....K.K...QWK. : 51
DareCALRL : MQISI.QLISAALL--ALA.NAE..LN...L.G...RS...N.E.K...Q.K. : 52
DareCALRL2 : --MTA.SLLFMAVSV..IT.ESN.....G.S.RS.....K.KT..... : 52
IcpuCALR : --MKISLALFAVFA-VGITVDAT.F.K...L.GEGFKS..L..K.K...QWK. : 51
IcpuCALRL : -MQSEFK.FTIFSAALLAF.VDAT...K...L.G...TS...N.E.K...Q.K. : 53
IcpuCALRL2 : --MTA.CLLFISLSF..IT.ES.....G...KS..L..T.K..... : 52
HosaCALR3 : ---MARAL.Q.WAICM.RV.LAT...Q.E.L.GEH.RN..LQ.TND.RF.H.R. : 51
HosaCALR : ---ML.SVP..LGLLG..V.E.A...K...L.G.G.TS..I..K.K..F..... : 51
      -17          *          1          *          20          *

```

SasaCALR1a : TAGK**FY**GD**AE**DK**GL**QTSQDAH**Y**SSSSAR**FE**PF**SN**Q**GT**L**VI**Q**FT**V**K**HEQN**IDC** : 101
 SasaCALR1b : .P.....P.....R..... : 104
 SasaCALR1.2 :V**RS**.....R...A**S**...SI...KDQ..... : 106
 SasaCALRL1 : ...N.....R...AA.....E.....K... : 104
 SasaCALRL2a :A.....R...AL**S**.....E...S...V.....K... : 106
 SasaCALRL2b :A.....R...AM**S**.....E...P...V.....K... : 106
 OrlaCALRL1 : SS.....A...V.....R...AL...P...E.....K... : 104
 OrlaCALR3 : ...N.....R...AA.....E...S.....K... : 104
 OrlaCALRL2 : --.....G.....A...DD...K**QP**.....S... : 63
 DareCALR : .S.....P**L**.....R...AL**S**...DS.....K... : 105
 DareCALRL : ...N.....R...AT**V**.....E.....K... : 106
 DareCALRL2 : S.....AM...GD**N**KD**QP**.....S..... : 106
 IcpuCALR : S.....L.....R...AL**S**.....E...P.....K... : 105
 IcpuCALRL : ...D.....R...AT.....E...S...V.....K... : 107
 IcpuCALRL2 : S.....AI**S**...D...KD**QP**.....G... : 106
 HosaCALR3 : SS.....HK.....T**NGR**...AI...K...K.....Y.....KM... : 105
 HosaCALR : SS.....E.....R...AL**S**.....K**Q**...V..... : 105

40 * 60 * 80

SasaCALR1a : GGG**YIK**L**F**PADLDQAD**M**HGDS**N**Y**NI**M**F**GP**DI**C**GP**AT**KKI**H**VI**I**NY**K**GK**N**HL**IRK : 155
 SasaCALR1b :K.....G...V.....S... : 158
 SasaCALR1.2 :N**EE**.....T.....G...V...F.....N... : 160
 SasaCALRL1 : ...V**V**...ST.....Q**Y**.....YS...V...F.....K... : 158
 SasaCALRL2a : ...V**I**.....A.....Q**Y**.....YS...V...F.....K... : 160
 SasaCALRL2b : ...V**I**...N...A.....Q**Y**.....YS...V...F.....K... : 160
 OrlaCALRL1 : ...V**V**...S...S...K**YV**.....YS...V...F...Q...K... : 158
 OrlaCALR3 : ...V**V**...S...T...S**YV**.....YT...V...F.....K... : 158
 OrlaCALRL2 :SE**N**E...V.....G...V...F.....N... : 117
 DareCALR : ...V**V**...EM**TE**...E**Q**Y.....YS...V...F...Q...K... : 159
 DareCALRL : ...V**V**...N...N...Q**Y**.....YS...V...F...Q...K... : 160
 DareCALRL2 :S...N**E**...T.....G...V...F.....N... : 160
 IcpuCALR : ...V**I**.....E**Q**YL.....YS...V...F...Q...K... : 159
 IcpuCALRL : ...V**V**...N...L...Q**Y**.....YS...V...F.....K... : 161
 IcpuCALRL2 :S...K**EE**N...T.....G...V...F.....N... : 160
 HosaCALR3 : ...V...I...K**NLN**K**Q**Y.....FD**I**...V...L**HF**N**Y**EN**K**... : 159
 HosaCALR : ...V...NS...T...E.....G...V...F...V...N... : 159

* 100 * 120 * 140

SasaCALR1a : DI**RCK**DDE**Y**THLYTLILNP**D**NT**Y**EV**K**ID**N**KK**V**ESGSLEEDWDILPP**KKV**KD**PEA** : 209
 SasaCALR1b : .V..... : 212
 SasaCALR1.2 :V.....F..L..I..D... : 214
 SasaCALRL1 : EV**K**.....L.....NQ.....E.....T.....F..A..TI..... : 212
 SasaCALRL2a : E**K**.....L.....Q.....N**E**.....T..D.....T..... : 214
 SasaCALRL2b : E**K**.....L.....Q.....N**E**.....T..D.....A..TI..... : 214
 OrlaCALRL1 : .**K**.....L.....Q...R...E.....D...M...I..D... : 212
 OrlaCALR3 : E**K**.....L.....Q...Q...E.....F.....F...I..D... : 212
 OrlaCALRL2 : ...F.....V**V**.....N.....F...I..D... : 171
 DareCALR : .**K**.....L.....R...Q.....E.....F...I..... : 213
 DareCALRL : .**K**.....L.....Q.....E.....F...I..... : 214
 DareCALRL2 :S...V.....F..S..I..... : 214
 IcpuCALR :F**N**...R...Q.....N**E**.....F...I..... : 213
 IcpuCALRL : .**K**.....L.....Q.....E.....F...I..... : 215
 IcpuCALRL2 :S...V**V**.....D...F...I..D... : 214
 HosaCALR3 : L...V**GF**.....R...LS**D**...G**Q**SI...I**Y**...NL**T**SL...ET**S**AE : 213
 HosaCALR : ...F...V**R**.....S**Q**.....D...F...I..D... : 213

* 160 * 180 *

SasaCALR1a	:	VKPDHWDERERMEDPDDKKPEDWDRPENIADPDAKQPEDWDDMDGEWEPPMVS	:	263
SasaCALR1b	:	...D...V...K...N...	:	266
SasaCALR1.2	:	E...D...K.NID...E...VA...P...K.D...T	:	268
SasaCALRL1	:	K...D...D.PK.D.AE.A...VA...P...K.D...ED...VIT	:	266
SasaCALRL2a	:	K...ED...D.AKID...T.T...EK...P...I.D...VD...IP	:	268
SasaCALRL2b	:	K...ED...D.PKID...T.T...G.EK...P...K.D...VD...VIP	:	268
OrlaCALRL1	:	K...SD...D.ATID...S.T...Q...T.P...K.D...ED...IT	:	266
OrlaCALR3	:	K...ED...D.PKID...E.T...EK...Y.P...K...QED...I	:	266
OrlaCALRL2	:	K...ED...D.KIP...K...P...K...IA	:	225
DareCALR	:	K...D...D.AKID...E.T...K...P...K.D...ED...A.IP	:	267
DareCALRL	:	K...ED...D.AKID.ET.T...K...P...K...ED...A.IP	:	268
DareCALRL2	:	K...ED...KID...E.Q...E...K...P...K.D...E...T	:	268
IcpuCALR	:	K...ED...D.AKID...E.T...K...P...K.D...ED...IP	:	267
IcpuCALRL	:	K...ED...D.AKID.ET.T...K...P...K...ED...IP	:	269
IcpuCALRL2	:	K...ED...KIG...K...T.P...K.D...E...T	:	268
HosaCALR3	:	S.--D.EQTK-----N.AQ...EK--HFL.ASTSKQS...NGDL...D.PA...LQ	:	257
HosaCALR	:	S...ED...AKID...T.S...K...H.P...K...E...VIQ	:	267
		200 *		
			220 *	
			240 *	
SasaCALR1a	:	NPDYKGEWKPRITIDNPDYKKGWLHPEIDNPDISADSEIYRFDSIGVIGLDLWQV	:	317
SasaCALR1b	:	...K...V...E...	:	320
SasaCALR1.2	:	...KQ...A...V...E.T.P...QYA...	:	322
SasaCALRL1	:	...E...KQ...N...A.I...E.A...T.K.D.S.L...	:	320
SasaCALRL2a	:	...E.Q...AKQ...E...A.V...E.T.AS...K.N...L...	:	322
SasaCALRL2b	:	...E.Q...KQ...T.V...E.T.TS...K.N...L...	:	322
OrlaCALRL1	:	...EF...KQMS...N...A.V...V...E.AP.ANM.L.N...L...	:	320
OrlaCALR3	:	...E...KQ...A.V...E.P.AN...K...L...	:	320
OrlaCALRL2	:	...D.NA.E...A...I...E.T.P...KY...	:	279
DareCALR	:	...E...KQ...S...T.V...E.A.DA...K...L...	:	321
DareCALRL	:	...E...KQ...N...A.V...EHVP.DQ...K.N...L...	:	322
DareCALRL2	:	...Q...A...V...E.TP...KY...	:	322
IcpuCALR	:	...E...KQ...N...T.E...TP.TS...KY.N...L...	:	321
IcpuCALRL	:	...E...KQ...N...I.E...E.P...K.N...	:	323
IcpuCALRL2	:	...KQ...A...V...E.TP...KYS...	:	322
HosaCALR3	:	K.P.QDGL...EG.H---.DV...RKMK.T.LTQYDLSE.EN...A...E...	:	307
HosaCALR	:	...E...Q...T.I...E.P.PS...AY.NF...L...	:	321
		260 *		
			280 *	
			300	
SasaCALR1a	:	KSGTIFDNLITDDATLAEVGNETWGQTKDPEKKMKESQEEKERKKLEAEEMA	:	371
SasaCALR1b	:	...V...Q...G	:	374
SasaCALR1.2	:	...N.PK...A...DRL...E...R...VN	:	376
SasaCALRL1	:	...VS...VKE...KF.A...V...E...--...DDK.RK.ED.KN	:	372
SasaCALRL2a	:	...G...IKE...F...A...DA...E...AR.E...KS	:	376
SasaCALRL2b	:	...G...VKE...F...T...E...DA...E...AR.E...KS	:	376
OrlaCALRL1	:	...IKE...AK...EV...G...RDD...DI...LR.E...KS	:	374
OrlaCALR3	:	...V...VKE...D...K...A...E...Q...--...QD.L.RKDE...KN	:	372
OrlaCALRL2	:	...N.PN...D...K...A.R...D.E...R.E.DKK	:	333
DareCALR	:	...VEE...KF.TD.VRQR-TREE...DQ...E...R.E...KS	:	374
DareCALRL	:	...F.S...VKE...F...A...G...--E.QE.K.RK.E...KN	:	374
DareCALRL2	:	...S...N.PK...T...A...A...ED...HDE...KS	:	376
IcpuCALR	:	...VKE...DF.K...A...G...DK...E...QR.E...KN	:	375
IcpuCALRL	:	...A...KE...F.KN...V...G...DE--D.K.RK.ED.KN	:	375
IcpuCALRL2	:	...N.PK...S...A...A...D...E...R.E.DKK	:	376
HosaCALR3	:	R...EEY.DNF.KA...E...G...RE.DAI.AKE.M...ARE---	:	358
HosaCALR	:	...N.EAY...F...V...AA...Q...DK.D.EQ.L.E.E.DKK	:	375
		* 320 *		
			340 *	

	Acidic region	ER retention signal	
SasaCALR1a	: RKEETKDEPGEEEEEEEELEHEEEEEDEEEGETGAQEEEEESDSIKDEL	: 421	
SasaCALR1b	: ...Q...E.....K.E.L.DR.E....-...E.....T.....	: 423	
SasaCALR1.2	: K...D--DED.DK..RD..EDYDD..E..-----...TDSKL....	: 416	
SasaCALRL1	: KEQA.EA.EEG....G.D.GE.E.TP.-----EGT...EAPG....	: 413	
SasaCALRL2a	: K-DTAD..GD.D.--D.P..EDD-DSPT.-----E..G.DPKKD....	: 415	
SasaCALRL2b	: K.DTAD..GD.D...D.SK.E.E-DSPT.-----EG...IPMKD....	: 418	
OrlaCALRL1	: K.DDEDEMDE.DD...KN..E.EDMDAG-----TDDD.DIKQ....	: 415	
OrlaCALR3	: KDQD.DGDDDEDID.D-...SK.EA..A-----LS.TD.DAQA....	: 412	
OrlaCALRL2	: ..D.DEE.DK.....D.E.EDG..P.L.E-----..D.SP.AV....	: 378	
DareCALR	: K.D-DNE.DE.D.D.D.P..DD.T..PP.-----EE..G.DDALP....	: 417	
DareCALRL	: KEQN.EAADE...D.G....E.D.T..PQ-----ED.GD.DVLQ....	: 417	
DareCALRL2	: ...A.-.EE...K...D..E.E....E..DE-----...TDSKL....	: 419	
IcpuCALR	: K..SEG..GDDD.P.D.Q..D.E..TPE.PNL--LDD..D..DKPS....	: 423	
IcpuCALRL	: KEQD.TTDDDDSDNDDDDNDE..GKE.TTE.PDV-.G..DDAQP....	: 424	
IcpuCALRL2	: .S..A...AD..DKD..D.DE.D.....-----DG...GTDSKL....	: 421	
HosaCALR3	: -----EE.LLSGKINRHEHYFNQFHR-----RN..	: 384	
HosaCALR	: ...EEA.DK.DD.DKD.D.EDE.DK.ED.-----..DVPGQA....	: 417	
	360 *	380 *	

References:

1. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular biology and evolution* 25(7):1307-1320.
2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. Fuller JR, *et al.* (2004) Characterization of the molecular chaperone calnexin in the channel catfish, *Ictalurus punctatus*, and its association with MHC class II molecules. *Developmental and comparative immunology* 28(6):603-617.
4. Liu H, *et al.* (2011) Molecular responses of calreticulin genes to iron overload and bacterial challenge in channel catfish (*Ictalurus punctatus*). *Developmental and comparative immunology* 35(3):267-272.
5. Kozlov G, *et al.* (2010) Structural basis of carbohydrate recognition by calreticulin. *The Journal of biological chemistry* 285(49):38612-38620.
6. Raghavan M, Wijeyesakere SJ, Peters LR, & Del Cid N (2013) Calreticulin in the immune system: ins and outs. *Trends in immunology* 34(1):13-21.

Additional file 10. Figure S7. Beta2-microglobulin (b2m) data

Table of Contents		Page
S7a	Phylogeny of deduced b2m sequences from selected species.	1
S7b	Alignment of deduced b2m amino acid sequences	2

Figure S7a. Phylogeny of deduced b2m sequences from selected species.

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (1). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Bootstrap values from 100 trials are shown next to the branches. Sequence references are shown in parenthesis. Salmon b2m1 and b2m11 represent the two sequence groups found in the 13 Atlantic salmon gene sequences (Additional file 2: Text S1). Unique duplications in the Ostariophysi and neoteleost lineages are colour shaded while Atlantic salmon sequences are shown in red font.

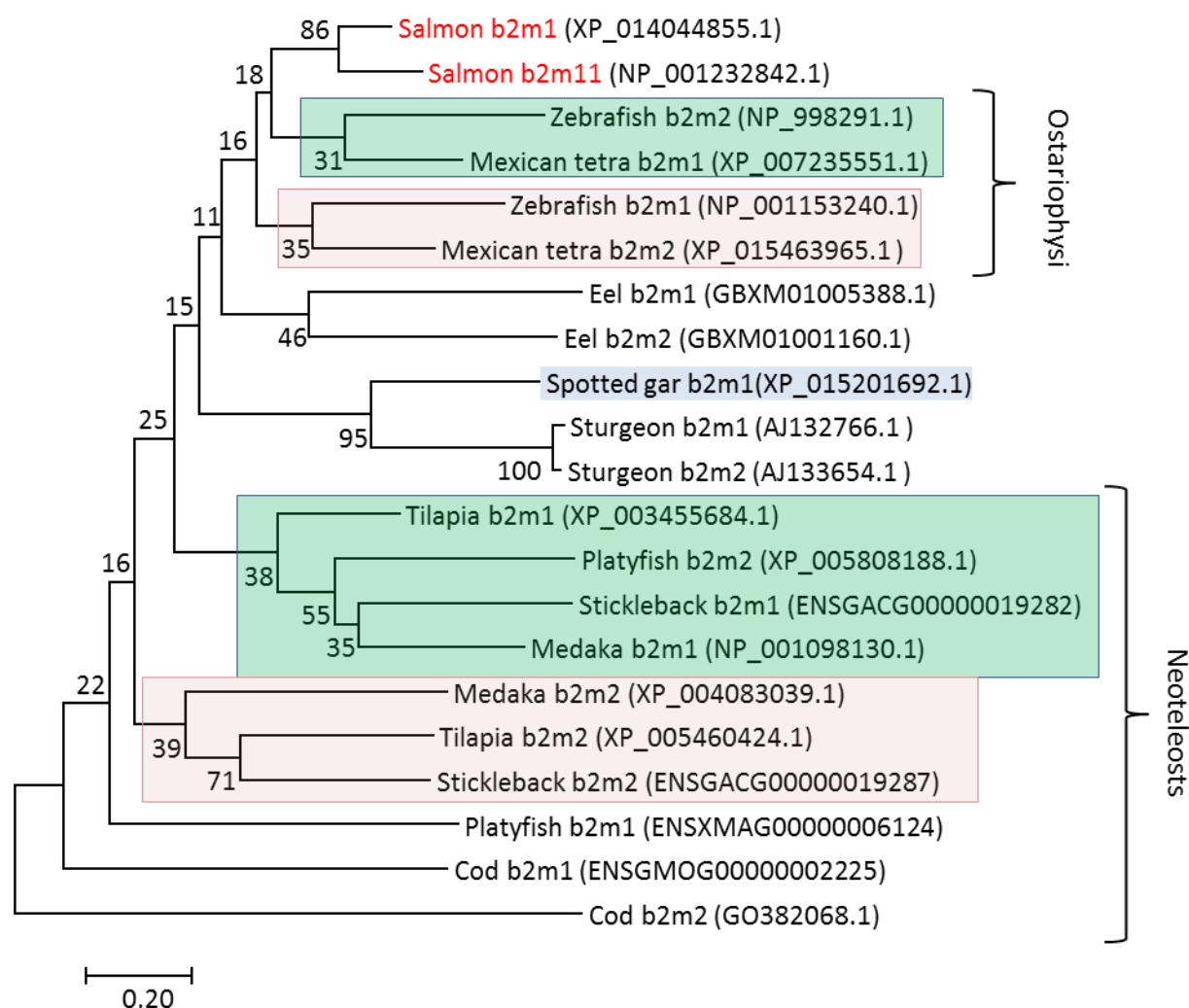


Figure S7b. Alignment of deduced b2m amino acid sequences

Sequence references are shown in Figure S7a. Abbreviations are as follows: Sasa is Atlantic salmon, Orni is Nile tilapia (*Oreochromis niloticus*), Gaac is stickleback (*Gasterosteus aculeatus*), Xima is platyfish (*Xiphophorus maculatus*), Orla is medaka (*Oryzias latipes*), Gamo is cod (*Gadus morhua*), Dare is zebrafish (*Danio rerio*), Asme is Mexican tetra (*Astyanax mexicanus*), Anan is eel (*Anguilla anguilla*), Acba is sturgeon (*Acipenser baeri*), Leoc is spotted gar (*Lepisosteus oculatus*), Gaga is chicken (*Gallus gallus*), Hosa is human (*Homo sapiens*). Trout b2m sequences were described in Shum et al. (3) while Kondo et al. (4) describe two distinct b2m variant sequences in other teleosts. Sequences are color coded according to physiochemical potential.

```

          *          20          *          40          *          60
Sasab2m1 : --MKSI---LSIVVLVLIYSA--VESKESPPKVQVYSRNPNGNFGDKNTLICHVSGFHPPDISIQ : 57
Sasab2m11: --..TV---..VIAFCVFLGF--INA.....EH.KD..... : 57
Ornib2m1 : --..AV---..CLA.IAAF.C.--..A.Y....T.....EY.KE.V.....N.....T.T : 57
Ornib2m2 : --..TF---VFLA.VL.LGLTDLASAA.N.S.....T.S..QY.TA.....E.T.E : 59
Gaacb2m1 : --..MLF---FTLAA.LA.SF.--QDPNHT.....H...E.DQE.....T.. : 57
Gaacb2m2 : --..IL---VCAF.VG.LSL.S.SQD.T.S...E...KL.EY.KV.....TD...E.K.E : 58
Ximab2m1 : --..SQT---..AV...LPM.QSN-TFQIQFS...I.....LY.N..V.....E.K.D : 58
Ximab2m2 : --..LL---..CLAA..A.FC.--..SC.I....LI.....E..KQ.S.....N.....T.D : 57
Orlab2m1 : --..EL---FF.AA.AAFCC.--AF..T....T.....D..QY.KE.....N.....T.R : 57
Orlab2m2 : --..G.VVTVVL.C.DLGVSI-K.RN.QA.....L..TL.KE.....T.....E..K : 61
Gamob2m1 : --..IT---FCA.AVA.LLCT--.....V.....SK.AEL.EA.....L.QA....E.T.E : 57
Gamob2m2 : --..GVFGVKMNPMLISVFCGLVLSMAIT.....T.E.AEP.TG.S...YLNN.Q..EVKVD : 62
Dareb2m1 : --..RAL---ITFAL.C.L.IT--..QG.V.T...H...HF.EY.KP.....Y..S.....E : 57
Dareb2m2 : --..WFK---..AV.A..FLNAS--CLA..T.....EYDKD.....Y.....T.E : 57
AsmeB2m1 : --..LM---..FC..FAFLGVS--SFA.....E..KD.V.....T.D : 57
Asmeb2m2 : --..LL---..F..ACFL.IS--A.A.....YS..E..KP.H.....E.....N.E : 57
Ananb2m1 : --..N...---TSFC.AVLLVG--..N.IH.A.....KL..Q.DQP.V.L.Y..N.....K.T : 57
Ananb2m2 : --..LL---..CLA..AIALIS--..NGVH.A.....ML.VG.K.....K.. : 57
AcbaB2m1 : MFLR.TF--VAAL.AC.A.IHL-GDAT.AM...I...V.EL.KP.....N.KL. : 61
AcbaB2m2 : MFLR.TF--VAAL.AC.A.IHL-GDAT.AM...I...V.EL.KP.....Q..N.KL. : 61
Leocb2m1 : --..VLY--..TAL.AF.A.LDS-AAAAAAE...I.....EYDKS.....D.....K.E : 59
Gagab2m : --..GKA---AAV.LVT.VALLG-LAQADLT.....F.ASA.T..V.N.FAA....K...T : 58
Hosab2m : --..SRS---VALA..A.LSLSG-L.AIQRT..I.....H.AEN.KS.F.N.Y.....S..EVD : 58

          *          80          *          100          *          120
Sasab2m1 : LLKNGVEIPDAKQTDLAFEQGWQFHLTKSVGFPTDSGEEYTCRVRHLK--NL-KTYTWEADM : 116
Sasab2m11: .....G.....V.N.D.H.....H.A...KE..N.A.K.T.GQ--DT..FG..SN. : 116
Ornib2m1 : .....G.....V.N.D.H.....H.A...KE..N.A.K.T.GQ--DT..FG..SN. : 116
Ornib2m2 : ....DKV...GSQ.....EN.HY.M..H.P...TKEDVFS...T.TG--KK-R..I.... : 118
Gaacb2m1 : .M.D...L.N..L.....K...R.V...H.T...RR..K...K.T.GT--TV-RD.A..SN. : 116
Gaacb2m2 : ..R.RQA..G.M.....AS.HY...H.P...QP.DTFS.N.T.VG--IS-.I...P.. : 117
Ximab2m1 : .....Q..N.F.....EN.HY...HAN...VE..R.M...T.VG--TT-TQHE...V : 117
Ximab2m2 : ..CD.TTL.NPV.....K.D.H.....AT...E..KTCI...N.AG--TS-ND.N..PN. : 116
Orlab2m1 : .MEDDQ.L.E.....K.N.H.....AP...KQ.AR...K.T.SG--LA-.D...SN. : 116
Orlab2m2 : .....N.L.GGG.....ES.HY...HAP...EK..S...V.T.TT--GT-.....P.V : 120
Gamob2m1 : ...D.AVL.N.N...S.DKT.H...S.FADI..SKE.S.A.Q.T.IG--KR-TVHA..PN. : 116
Gamob2m2 : ..E..AV..G.V.S..M..SQ..Y...R.P.I.RE.AR.A..N.MG--RI-TDHG.DMND : 121
Dareb2m1 : .....QVMS.T.....K.....A...EK.D...S...M.--ET-.KFS..PN. : 116
Dareb2m2 : .....QP.EKTT.....F.....FAA.K.RP..V...Q...T--ET-.....P.. : 116
AsmeB2m1 : ..E..TP..N.....T...K..VM.S.K...MQ--TT-..N..S.E : 116
Asmeb2m2 : ..R...V.TE.....K.....P...QKN...A...T.MQ--KAPQK.S.DP.. : 117
Ananb2m1 : ...DN..M.G.T.....S.....P...QA..K.A...Q.QG--IT-.P.S..P.. : 116
Ananb2m2 : .ME..R...N.E.....T.....R..D...VE.K..S.V.K.QN-DAA-.S.Q..P.. : 117
AcbaB2m1 : ....D...EG.Q.S...H...H.....A...KE..H.S.K.E.STLREA-TRF..TP.. : 122
AcbaB2m2 : ....D...EG.Q.S...H...H.....A...KE..H.S.K.E.STLREA-TRF..TP.. : 122
Leocb2m1 : ....KQQMEGV..S.....H.....QT.P...KK.DI.S...S.STLSVP-.EF..P.. : 120
Gagab2m : .M.D...PMEG.QYS.MS.NDD.T.QRLVHAD...S..ST.A.K.E.ETLKEP-QV.K.DPEF : 119
Hosab2m : .....ER.EKVEHS..S.SKD.S.Y.LYYTE...TEKD..A...N.VTLSQP-.IVK.DR.I : 119

```

References:

1. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutations data matrixes from protein sequences. *Computer Applications in the Biosciences* 8:275-282.

2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. Shum BP, *et al.* (1996) Unexpected beta2-microglobulin sequence diversity in individual rainbow trout. *Proceedings of the National Academy of Sciences of the United States of America* 93(7):2779-2784.
4. Kondo H, *et al.* (2010) Identification of two distinct types of beta-2 microglobulin in marine fish, *Pagrus major* and *Seriola quinqueradiata*. *Veterinary immunology and immunopathology* 134(3-4):284-288.

Additional file 11: Figure S8. ERp57 and ERp57L data

Table of Contents		Page
S8a	Phylogeny of deduced ERp57 and ERp57L amino acid sequences	1
S8b	Alignment of deduced ERp57 and ERp57L amino acid sequences	2

Figure S8a. Phylogeny of deduced ERp57 and ERp57L amino acid sequences

The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel 2008 model (1). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7 (2). Bootstrap values resulting from 100 trials are shown next to the branches and accession numbers are shown in parenthesis. Salmon sequences are shown with red font. The tree was rooted using PDIA4 sequences. The Spotted gar ERp57L sequence was predicted using FGENESH (3) and can be found in Additional file 2: Text S1.

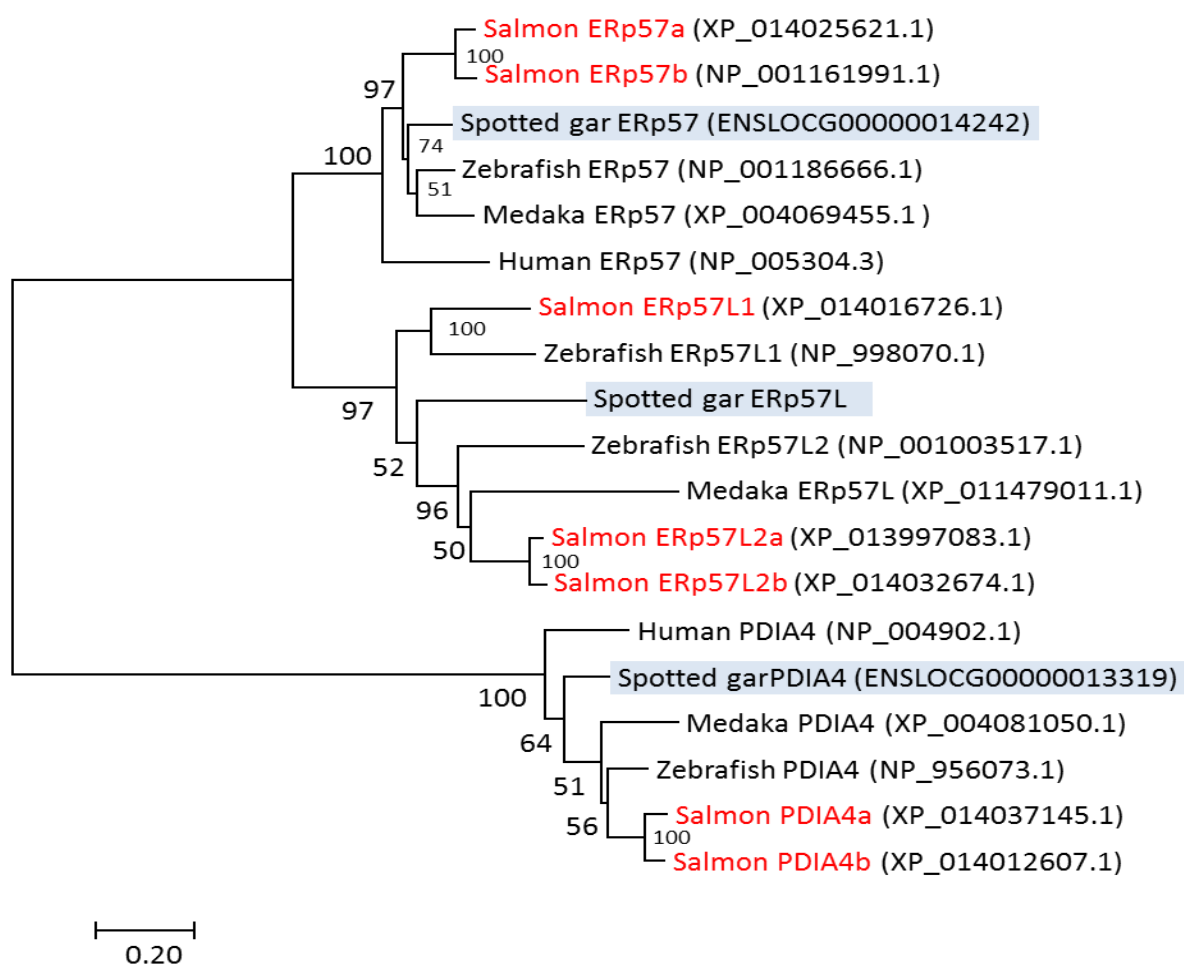


Figure S8b. Alignment of deduced ERp57 and ERp57L amino acid sequences

Sequence references are shown in Figure S8a. Residues are color coded according to physiochemical properties. Residues interacting with Calnexin and Calreticulin are shaded cyan while those interacting with Tapasin are shaded green (4). Numbering according to the mature human ERp57 sequence.

		-24	*	1	*	20	
HosaERp57	:	-----MRLRRLALFPGVALLLAAARLAAASDVLELTDDNFESRISDTGSAGLMLVE	:	51			
SasaERp57a	:	-----MLK.FF.VV-----L.GA.L...I.F...D.D.K.G.H.---MI...	:	40			
SasaERp57b	:	-----MLK.FF.IV-----L.GA.R...I.FS...D.D.K.G.H.---MI...	:	40			
DareERp57	:	-----MMLA.LFLVA-----F.AA.R.....Y...D.D...G.HD---I...	:	41			
LeocERp57	:	MNGWQTVPCSTSLV.Y.GVPR TSPCFL.VAVS.....F...D....A.HD---V...	:	56			
SasaERp57L1	:	-----MGT.GPFRM.LLF..AQQNVFV.-.....G.SD.HYTVAEYE---TV...	:	48			
DareERP57L1	:	-----MR.PCFIICAVLTVW..EG.....G.SD.DRSAGMHD---TL...	:	43			
SasaERp57L2a	:	-----MASFLS.IPPFLLSV.IFSGAAV.RG.....G.AD.DYLAAEHE---T...K	:	49			
SasaERp57L2b	:	-----MASFLS.IPAFTLSVVI FCGAVV.RG.....G.AD.DYLAAEHE---T...K	:	49			
DareERp57L2	:	-----DEMI.RGLLCILVCSLSSSAREH....K...AD.DYLAPEHE---TL...K	:	48			
LeocERp57L	:	-----MARLSV.L.CPAALL..AALG.VAV.GG.....G.AD.DGAVAQHE---TV...	:	51			
C57 forms disulfide bond with C95 in tapasin							
	*	40	*	60	*	80	
HosaERp57	:	FFP C57 CKRLAPEYEA AATRLKGIVPLAKVDCTANTNTCNKYGVSGYPTLKIFRDGE	:	110			
SasaERp57a	:	K.....V.....G.....VHN.V.Q.....	:	99			
SasaERp57b	:	F.V.....A.....VQN.V.Q.....K...	:	99			
DareERp57	:	SKV.G.....	:	100			
LeocERp57	:	S.....S.V...F.....G...	:	115			
SasaERp57L1	:	QQ.....T...K...T.S.....V.SE.GRF.N.....N...	:	107			
DareERp57L1	:	Q.....K...TLA.....V.SE.ERF.N.....N...	:	102			
SasaERp57L2a	:	.Y.....K...DF.T.....T.....PD.GRF.T.....N...	:	108			
SasaERp57L2b	:	.Y.....K...DF.T.S.....T.....SPD.GRF.T.....N...	:	108			
DareERp57L2	:	.Y.....K...F.S.S.....T.T.....EI.KH...N.....N.Q	:	107			
LeocERp57L	:	QK.....E.....A.....HSE.SRF.N.....N.Q	:	110			
	*	100	*	120	*	140	
HosaERp57	:	EAGAYDGP C95 TADGIVSHLKKQAGPASVPLRTEEEFKKFISDKDASIVGFFDDSFSEAH S	:	169			
SasaERp57a	:	D..P.....E.K..AD.T.YVG.R...V...A.DG.P.KA	:	158			
SasaERp57b	:	D.....S.IE.K..AD.T.YVG.R...V...A.GG.P.QA	:	158			
DareERp57	:	DS.G.....E.KN.AD.E.Y.G.R...V...A.GG.A.QG	:	159			
LeocERp57	:	.S.....EIK.V...E...G.R...VI...AEGG.TTQA	:	174			
SasaERp57L1	:	DFA.....S.....YM.....S.....HNGRDLDA.VNNF...V...SGVD.SQMA	:	166			
DareERp57L1	:	.S.....YM.....S..A.LK.ADLDG.VDNYE..V...SGED.AQLA	:	161			
SasaERp57L2a	:	D.SS....S.....HFM.....N..T..R.ADLEA.VNHF...V...SGPD.GQLA	:	167			
SasaERp57L2b	:	DSSS....S.....HYM.....N..T..S.ADVEA.VNHF...V...SGPDTAQLA	:	167			
DareERp57L2	:	.SSS....S.....DYM.....D..L.HS.LDLE...NH F...V..L.SGTD.SQLA	:	166			
LeocERp57L	:	.SS.....YM.....S..E.....DLDS.V.HF.G.V...LGPE.ADLA	:	169			
	*	160	*	180	*	200	
HosaERp57	:	EFLKAASNLRDNRYFAHTNVESLSVNEYDDNGEGIILFRPSHLTN C95 FEDKTVAYTEQKMT	:	228			
SasaERp57a	:	...S..A..ESF.....S.E.LQKHSVE.....R.N...EGS.KFS.DTF.	:	217			
SasaERp57b	:	...S..A..ESF.....SGE.LQKNGVE.....AR.S...ESVIKFS.D.F.	:	217			
DareERp57	:	...A..ES.....N.D.LKKHGID.....SPQ.S...SS.LF..D.F.	:	218			
LeocERp57	:	...S..A..ES.....S.E.LQKHGID...V...PR.S...ESS.K.S.D.Y.	:	233			
SasaERp57L1	:	...S.AM..SH.....TDL.GLKHGVESDTVV...PR.NS...SL.KSD.-AVS	:	224			
DareERp57L1	:	...S.A...S.....STDVGAGLK.GVD.CVL...PR.SS...NV.K...-HLS	:	219			
SasaERp57L2a	:	...VM.EHF.....IDMT.GLKHGVD T.RVL...PR.SS...ESVLHF.-TI.	:	225			
SasaERp57L2b	:	...GAM..HF.....INMT.GLKHGVD T.SVL...PR.SG...ESVLR F.-TI.	:	225			
DareERp57L2	:	...G..LM.ESF.....TDLQ.GQK.GVTH.S.L...APR.SS...ESV.PH.G-SLS	:	224			
LeocERp57L	:	...R...SM..S.....SA...RQRHGVO.AVL...APQ.AS...ESV.RHRG-AVR	:	227			

		*	220	*	240	*	260		
HosaERp57	:	SG	KIKKFIQENIFGICPHMTEDNKDLIQGKDLLIAYY	DVDYEKN	AKGS	NYWRNR	VMMVA	:	287
SasaERp57a	:	NA	...Q...D...M...D...QMK...V...P...K...					:	276
SasaERp57b	:	NAM	...D...M...D...QMKD...V...P...K...					:	276
DareERp57	:	A	...D...A...QLK...V...P...K...					:	277
LeocERp57	:		...D...L...QLK...V...FE...P...T...K...					:	292
SasaERp57L1	:	TASLRQ	...RD.V...L...L.AE.RENMR.R...V...LR.I...T...K...					:	283
DareERp57L1	:	VSSLH	...VKD...L.L...ME.R.TVRES...T.FFN...LR.P...T...I.K...					:	278
SasaERp57L2a	:	THTLRR	...RD...M...L.NE.R.KLK.Q...T...L.LQ.P...K.G					:	284
SasaERp57L2b	:	THTLRR	...RD...M...L.NE.R.KLK.Q...T...L.LR.P...K.G					:	284
DareERp57L2	:	VTGLRR	...RD...L...K...EVLRRK...T...L.LH.P...LK...					:	283
LeocERp57L	:	PDALRR	...R...M...L.HE.REQLMRQ...T.F.EL.LR.P...					:	286
		*	280	*	300	*	320		
HosaERp57	:	KKFLDAGHKL	NEFAVASRKTFSEHSLD-FGLE-STAGEIPVVAIR	TAKGEKFVMQEEFS				:	344
SasaERp57a	:	G	...Q.N...KNS.QDIAE-M.D-ASS.L...G...D.Y.T...					:	333
SasaERp57b	:	S	...Q.KT...KNS...DI.E-M.D-ASS.L...G...D.Y.A...					:	333
DareERp57	:	G	...Q.K.S...NKNR.DV.E-L.D-GSS.L.L.G...D.Y.K...					:	334
LeocERp57	:	S	...Q.K.S...NKNS...I.E-D-SS.L...G...D.Y...					:	349
SasaERp57L1	:	TQ.QSR	--SY...N.AE.QE.EEE...GP.DG.L.LIT.NRE.H.YS...T.					:	340
DareERp57L1	:	TQ.Q.R	--T...D.QE.QD.EEE...VSS.EG.DV.L.T...RA.Q.YS...T.					:	335
SasaERp57L2a	:	SQ.ASQ	--S...N.RD.VD.EEE...GA.DG.DL.F.T...RQ.F.YT.R...T.					:	341
SasaERp57L2b	:	YQ.ASQ	--S...N.RD.VD.EEE...GA.DG.DL.F.T...RQ.F.YT.R...T.					:	341
DareERp57L2	:	T.SSQ	--ML.S.N.ND.ME.EEE...SA.DGN.L.F.T...RT.D.YS.R...T.					:	340
LeocERp57L	:	SR.AGR	--S...D.RD.PD.EE-...GAEG-L.A.V.VL.Q.Y.R...T.					:	341
		*	340	*	360	*			
HosaERp57	:	DGKALERFLQDYFDGNL	KRYLSEPIPESNDGPVKVVVAENFDEIVN	NENKDVLI	EYFYA			:	403
SasaERp57a	:		...K...N...T...A...E.D...					:	392
SasaERp57b	:		...K...S.N...T...A...E...					:	392
DareERp57	:		...V.N...L...S.DDS...					:	393
LeocERp57	:		...K...V...L...E.DDS...					:	408
SasaERp57L1	:	S	...E...A.K.QV.AAS.N...D.E...PS...V...					:	399
DareERp57L1	:	S	...K.E.AKR.V...L.DT.A.DPE...V...					:	394
SasaERp57L2a	:	S	...V...A.R.I...K.K...S.E.DPE...					:	400
SasaERp57L2b	:	S	...E...A.R.I...K.K...S.E.DPE...					:	400
DareERp57L2	:	S	...S.E.A.R.V...V.AI.N.V...DT.E.DPE...					:	399
LeocERp57L	:		...E...A.R...V...T.DAS...					:	400
		380	*	400	*	420	*		
HosaERp57	:	P	CGHCNLEPKYKELGKELSKDPNIVIAKMDA	IANDVPSPYE	RG	PTIYFSPANKKL		:	462
SasaERp57a	:		...S...W...S...Q...F.A.GQ.M					:	451
SasaERp57b	:		...S...W...S...Q...F.A.GQ.M					:	451
DareERp57	:		...S...E...S...GR.Q					:	452
LeocERp57	:		...S...D...G...Q					:	467
SasaERp57L1	:		...S...T...Q.A.TH...PT.D.Q...F.V.GQ.D					:	458
DareERp57L1	:		...GN...PN.D.Q...V.SGQ.D					:	453
SasaERp57L2a	:		...S...A.Q.YS...QGFD.Q...AQ.S.D					:	459
SasaERp57L2b	:		...S...A.Q.YS...QGFD.Q...AR.D.D					:	459
DareERp57L2	:		...K...TA.M.YS...AG.D.Q...AA.GR.S					:	458
LeocERp57L	:		...R.T.QQ.G...EG.DIQ...A.VGQ.A					:	459
		440	*	460	*	480			
HosaERp57	:	NPKKYEGGRELSDFISYLQREATNPPV	IQQEEKPKKKKKAQEDL					:	505
SasaERp57a	:	S	...V...KK...L.A.ETS...-KNE					:	493
SasaERp57b	:	S	...A.V...K...L.A.ETS.NI-IE					:	493
DareERp57	:		...V...K...TV.V.D.K.S...-KSE					:	494
LeocERp57	:	S	...V...KK.S.L...EK.S...-KSE.SCLLERMKEQCLVDVN					:	525
SasaERp57L1	:	Q.R	...VN.LN.KE...H.L.LGTA					:	493
DareERp57L1	:	Q.RR	...VN.T.KK...LILDDS					:	488
SasaERp57L2a	:	Q.R	...AH.VK.K.K...SHV.VSGV					:	494
SasaERp57L2b	:	Q.R	...A.VK.K.K...SHI.VSGV					:	494
DareERp57L2	:	E.R	...A.VK.VNF.K...K.LILNGV					:	493
LeocERp57L	:	A.RR	...HAFHFHCISRW.KTRQVC.L					:	485

K/RXEL

References:

1. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular biology and evolution* 25(7):1307-1320.
2. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33(7):1870-1874.
3. Solovyev VV, Kosarev P, Seledsov I, & Vorobyev D (2006) Automatic annotation of eucaryotic genes, pseudogenes and promoters. *Genome Biology* 7(Supl.1):10.11-10.12.
4. Dong G, Wearsch PA, Peaper DR, Cresswell P, & Reinisch KM (2009) Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* 30(1):21-32.

Additional file 12. Text S4. Alignment of deduced TAPBP, TAPBPR and TAPBPL amino acid sequences

Alignment of deduced Tapasin (TAPBP), Tapasin-related (TAPBPR) and Tapasin-like (TAPBPL) amino acid sequences from selected species. Numbering above and below the alignment relates to the human TAPBP and TAPBPR sequences. Sequence references are as follows: Atlantic salmon: TAPBP_a#C NP_001117077.1, TAPBPR NP_001133983.1, TAPBPL1a XP_014069540.1, TAPBPL1b XP_014017660.1, TAPBPL2 XP_014062182.1; Northern pike TAPBP XP_010899738.2; Zebrafish TAPBP GDQH01003123.1, TAPBPR XP_001919985.2, TAPBPL AAI71514.1; Medaka TAPBPR XP_011483883.1, TAPBPL XP_004075780.1; Spotted gar TAPBP GFIM01016833, TAPBPR XP_015193320.1, TAPBPL GFIM01040944.1; Frog (*Xenopus laevis*) TAPBPL XP_018100952.1; Turtle (*Chrysemys picta bellii*) TAPBPL XP_005298961.1; Chicken (*Gallus gallus*) TAPBP1 NP_001029988.2, TAPBPR NP_001026543.1, TAPBPL merged transcripts BU342879.1, BU369515.1 and BX257449.3; Kiwi (*Apteryx australis mantelli*) TAPBPL XP_013817376.1; Opossum (*Monodelphis domestica*) TAPBPL XP_007485846.1; Human (*Homo sapiens*) TAPBPR NP_060479.3, TAPBP NP_003181.3. The six conserved cysteines are shown above the alignment (C1-C6). Highlighted residues are as follows: TAPBP and TAPBPR TC2, TC3 and TN3-TN6 are regions that interfere with MHCI binding (1-3), sites interaction with ERp57 (1), N-linked glycosylation site interacting with CALR (4), transmembrane site associating with TAP (5), ER retention motif (6) and TAPBPR cysteine C94 interacting with UGT1 (7).

		C1	TN4
		1 * 20 *	
Human TAPBPR	: -----MGTQ E GWCLLLC L ALSGAAETK P H P AEGQWR A V D VVLDCFLA F DGAHRG-----ALASSEDRAR :	59	
Chicken TAPBPR	: -----MLAVLAGTVEGLTPV P ELRRVDVVLGCSYVW E GGLSR-AFG-----GSEHP :	45	
Salmon TAPBPR	: -----MLEILLFYGLITCVSGQGADVVLSCLSLVEGGSMG-GM-----GGGALFSRTP :	48	
Zebrafish TAPBPR	: MQIVSDIQQNY S ISCLVSSCSVMLTLLVLSLSLYGLDGADVVLSCSLIEGGMG-GMGMMGMGAMFKRT P :	70	
Gar TAPBPR	: -----MR L AFGVLVLAS H LMSSVSGSVADVVLAACSLVEEGGGMG-GR L A-GAGALFTRDE :	53	
Medaka TAPBPR	: -----MI E IIFFGLFITCVNA D GAADVVLSCALVEEGSGLS-GM---EGGALFSRTP :	48	
Salmon TAPBP1la	: -----MSLNVNILLCFLFCGEVPGIQSFEQVPWLPCQLVDESVTFN-DE G --HAETEQH R D :	54	
Salmon TAPBP1lb	: -----MSLNVNIVICLFLCAEVPGIQSLEQVPWLPCM V DERVKFN-DE G --HAETQYQH R N :	54	
Zebrafish TAPBP1	: ---MIVLR T RIHNMFWTIGGFILALVCLGVNGSQSLRD V QWLPCQFVDEKVQIN-E E G--HTETKYIH R E :	64	
Medaka TAPBP1	: -----MGVFQKVVIIFYFVGIVVSQVQIPWLSQCFTEDEHVSNN-SE G --HIETQLIP R E :	51	
Salmon TAPBP12	: -----MKTFTVLLCLIP-----YAGVLGFLQVQWLR C RLKD E YVWTN-DE G --HIETNNSY R D :	50	
Gar TAPBP1	: -----MITTA V LLCVLIC S GLRCVR S ESLA E AGLLACE L VEER V SSG-L S G--APETQFVN R K :	55	
Frog TAPBP1	: -----MD S RLYLII L NLLIVCN D IQSFED L TSRK S RKLSCMYVDMRN K TL-----LGYN D IMK Q K :	55	
Chicken TAPBP1	: -----MAQ R TTLELP P CGHGSPQDGGAQCSNHK D PK Q H-CF C IPRLRRILQLP C ME M KKNIPLSS K T :	61	
Kiwi TAPBP1	: -----MN Y RTYLVA G CF L CIGIAKNLNGIATAGSSFRKS R QLACV F ETS K I I PL-----TK E LE I IR L N :	59	
Turtle TAPBP1	: -----MNF R IYTAVYC F ICTGILKHNGVATAGSTFRQSRRLSC L FET E T I PL-----LSEFTYIR Q N :	59	
Opossum TAPBP1	: -----MLGRGMLALRLLSPGLLWILGWHMVESGMTLGKAGQLPC L LEVALDPSE-----TLNGERKW Q H :	59	
Salmon TAPBP1a_27	: -----MANISTYIKLSFLAFTYFIHVYGTS C PVLECWFVQEK P GRG-G----GFPAAMIQ E K :	52	
Pike TAPBP1	: -----MAXISTYIKLSIIAFIYFTNVYGTSCP V LECWFVQEK P GRG-G----GFPAAMSQ E K :	52	
Zebrafish TAPBP1	: -----MSDISTVF K ISVIAFT L FHGFYGSACP V LECWFVQEK P GHG-G----GLSTPM S Q E K :	52	
Gar TAPBP1	: -----MSDTATIYRLTLLALWG L AAVCGS R CP V ECWYVQEK A GQ G -G----GFPAAMTQ E K :	52	
Chicken TAPBP1	: -----MAAGLRLLL A GLCWSQ F RV E DAASPPPPAPV R CAL L GVGRRG-----GLPGGN A RP :	54	
Human TAPBP1	: -----M K SL S LL L AVA L GLATAVSAGPAV I ECWFVQEK A SG K G-----LAK R P :	42	

	40	*	60	*	80	*	100			
Human TAPBPR	: ASLVLKQVPVLDDGSLEDFD		FGGGTLAGDQD	--PPIIFEASVDLVQIPQAEALLHADC	--SGKEVTC		ISRY	: 127		
Chicken TAPBPR	: ATLVLRGLSVTDDGTLGDVT		DYIIPQADHSSSPPIIFEASEQLVSI	PYAEALLHVDC	--SGEEVSC		ELSPY	: 114		
Salmon TAPBPR	: AMLVLRDLAVTFLDSPDLT		LPFNPPAVPDP	--NIILEAKVESP	ETPEADLLLHADC	--NEQEVTC		ISRY	: 115	
Zebrafish TAPBPR	: ATLVRDLFVENEGLSP		ELVTPYNPPETNPE	--DLIFELMLPSIQIPNADLLLHADC	--NEQEVTC		ISRY	: 137		
Gar TAPBPR	: ATLVLRDLFVTADES		LDVTFPAPP-APDPE	--NLIFEATVTSLEI	PNADSLHADC	--NEQEVTC		ISRY	: 119	
Medaka TAPBPR	: ATLVLRDVAVGPEEP		LEMLSPFVPPSPVPDPE	--LFLFEASVTSPDIPNTDRLLHADC	--NELEVLC		ISRY	: 115		
Salmon TAPBPL1a	: AGLQFGHPGDS	-----	ALNPN	--TITFLVTG	-----SKVDMRKYIEG	--VVEHQLQCE		IRRY	: 102	
Salmon TAPBPL1b	: AGLQFGHPGDS	-----	ALNPN	--AITFLVTG	-----SKVDMRKYIEG	--VVEHQLQCE		IRRY	: 103	
Zebrafish TAPBPL	: SVLQFGKDGDS	-----	ALHP	--AITFLVTG	-----SKVDMRRYLEG	--SEDTLQCE		IRRY	: 110	
Medaka TAPBPL	: AMVQFGLKGDA	-----	PVNPH	--AITFLITV	-----SKLDLQRSIEG	--VEVEELCEV		RRY	: 99	
Salmon TAPBPL2	: AVLQFENGSGNS	-----	ALLSE	--SITFLVAA	-----SKVDMRKFVEG	--PVDQLQCE		DIHRY	: 97	
Gar TAPBPL	: AALLLARPGREAD	-----	TAQRLQEAQRAG	--VVTFLATG	-----SAVDVTRHVP	--GTEQLQCE		IRRY	: 111	
Frog TAPBPL	: AFLVLSSRKQL	-----	SKEYFDSE	--GITFILKD	-----SPVNTMKFLKE	--HTLDLVCE		VNPY	: 105	
Chicken TAPBPL	: ETVLLNIRLLL		LPVAGAGGAQRLQTESLPAGN	--LPSFIVQE	-----PSVNILQHTDE	--DINMLDCRISPY		: 123		
Kiwi TAPBPL	: VRLLLSGTETSEE	-----	QRSHSENHPQDN	--IPSFIVKE	-----SSLDILQHANE	--DINTLD		CRISAY	: 115	
Turtle TAPBPL	: AWWLSSGAERP	-----	HQENPPAAD	--TLTFIVQES	-----SSDILQHVDE	--DTDKLECKINRY		: 111		
Opossum TAPBPL	: IQLVLSGSES	-----	QSLSHQGN	--SATFLLQD	-----RAGVLLPHMDE	--DVKLECKVSR		Y	: 108	
Salmon TAPBP _a _#C	: SLLYINTDPESEE	-----	TKSQQGPSADIN	--HDRVYYVTDPA	--AILCSSSLHPPEG	--SVHKPQCE		INPF	: 113	
Pike TAPBP	: SLLYINTDPEWKR	-----	PESQQR	PSTDIN	--XDRLYYITDPA	--ATFCSSSLHPPEG	--SVQKPQCE		INPF	: 113
Zebrafish TAPBP	: SLMFIRTEAYSEE	-----	TKAELHPPADIS	--SSRVYYVTDPA	--GTFCSAALNPPKG	--SVNKPKE		INPF	: 113	
Gar TAPBP	: SLLYVRTEPDP	-----	VGLQQNPPSID	--PKRIYYVTDAG	--ASLCSPSLRAAEG	--AVEKPKCE		VPSPF	: 111	
Chicken TAPBP	: ALLRFVSGDAET	-----	PPEPGPE	--PEVTFNVSD	-----PWGTLTPLGV	--PRTPPSC		ELNPT	: 104	
Human TAPBP	: GALLLRQGPGE	-----	PPRPDL	--PELYLSVHDPA	--GALQAAFRRYPRG	--APAPHC		EMSRF	: 96	

TN3

		*	120	*	140	*	
Human TAPBPR	:	FLQMTETTV-----	KTAAWFMANMQVSGGSPISLVMKTFRVTKNEA-----	LWHPTLNLPL	:	179	
Chicken TAPBPR	:	SFQQ-----	EGNGLCSASWFLATIRL--SSGISIVLLLRGPSCSSQKE--	GHDVTLHPKLRIPM	:	169	
Salmon TAPBPR	:	FPRN-----	AKEGSTEPAYFIGSLQIEGGGLSLTLILQTLPLDQS---	DRPALMQSKLELPL	:	169	
Zebrafish TAPBPR	:	VPR-----	GADMDSLPAHFFIGSIQLEGGGSLTLVLQTLHSETTQS--	DAQPLMQSKLNLPL	:	192	
Gar TAPBPR	:	FPRGPE-----	GTEPSPEDAFFIGSLQLEGGGVSLTLVLRTPQIPAEQEEDGGRPLRQSKLDLPL	:	179		
Medaka TAPBPR	:	SFK-----	GLQDSSDQTFMVISITVEELDFSISLILETLPLVEKHQS--	TLMHNKLGLELPL	:	167	
Salmon TAPBPL1a	:	STESPQMRW----	PGLGAQEH--DIWFTCTIRHTDGVFIITSFLRHTPATPTPG-----	QAHYLNWAAI	:	160	
Salmon TAPBPL1b	:	STEGIQVRW----	PGLGAQDH--DIWFTCTLRHTDGLFVITSFLRHTPATPTPG-----	QADYRNWAAI	:	161	
Zebrafish TAPBPL	:	STGGIVMRW----	PVAGAQBY--DVWFTCTLRHTQGFLVITTFILRHTPATPAQG-----	QVDFLQWMSV	:	168	
Medaka TAPBPL	:	STEGIHVRW----	PALGAEY--NRWYTCILRHRKDKFSVTAFLRHPSDQPPPG-----	QDYRSWPPPI	:	157	
Salmon TAPBPL2	:	SKGMSRVRW----	PSLGGTGH--DIWFTCTLRHTAGLFNITSFLRVTPATTMSA----	HQPDFLSWLTII	:	156	
Gar TAPBPL	:	STGGIQVPW----	KGTAGPD--SAWFTCTLRHSANEFIVTAFLLYPSAPAPQG-----	PI	:	159	
Frog TAPBPL	:	FTDNTQIIW----	PGIHVIENMEEAWFVSRFKHPNDKQFTLFFTHISQKSLEE-----	QDGSQQLQAF	:	164	
Chicken TAPBPL	:	FTANTQILW----	PGRKVKAHGLDTWFTCTIEHAAGKYTATAFLVQGHEDRE-----	HQTPGQLHQ	:	180	
Kiwi TAPBPL	:	FTANTQIIW----	PGREIRASSLDSWFICTIKHTAEKYMTTAFILVQTHKENE-----	NHSQGQLSP	:	172	
Turtle TAPBPL	:	FTDNTQILW----	PGLQTLQHLKLDWSWFTGTIKHLADKFTVTFVLVQSSASKEAE-----	NRVPEDVSQ	:	170	
Opossum TAPBPL	:	FTANTQILW----	PGLPPRQAQLPTWHLVITISHTDAQFKTASTFCFQSAESAQGT-----		:	158	
Salmon TAPBPα_#C	:	MPQPSTVQWVVPLTDSAHSPFIYLQADWYSAAALQGLDGQLGLSSVMRAPATATKE-----			:	166	
Pike TAPBP	:	MPQPSSVQWTVSLTDSAHSPFIYLQADWYSASLQGLNGQLSLSSVMRAPATATKE-----			:	166	
Zebrafish TAPBP	:	MPHASMVRWASALTDSAQSPVYLQADWFSVAAQGLDEQLTLNIMRAPASASKE-----			:	166	
Gar TAPBP	:	LPQPALVWRAAPLTAAGSSPAHLQADWLSAALRGLQGQLVLSSLLRAPASADQ-----			:	164	
Chicken TAPBP	:	NPQTGSDPWSRPLHPDARSPTAGGQWVAAVVG--TPQYGVTTALLQG-----			:	151	
Human TAPBP	:	VELPASAKWASGLTPAQNCPRALDGAWLMVSISS--SPVLSLSLLLRPQEPPEQEP-----			:	149	

ERp57 interaction C95 disulfide link to ERp57

TN6	TN7
-----	-----

N-linked glycosylation site

		IgC domain (PKV)											
		280	*				C5	*		320			
							300						
Human TAPBPR	:	IIQLNIQASPKVRLSLANEALL	-----	PTLIC	DIAGYYPL-DVVVTW	TREEL	-----	GGSPAQV	:	347			
Chicken TAPBPR	:	NIQLLVSEPPRVRVFPTE	--AS-LKRDET	--ITLTC	NIAGYYPL-DISVSW	TQKTP	-----	EDEVETISP	:	343			
Salmon TAPBPR	:	VIQLHVTQPPRVSLSEEK	--LV-FRDELP	--QKLSCH	CKNYYPL-DVQMEW	ISVSS	-----	TDSEPSVLS	:	353			
Zebrafish TAPBPR	:	IVQLHITQPPRITLSEEK	--LS-FPDRT	P--QRLSCH	CNRYYP	L-DVQVEW	FSQAP	-----	SSEESISLS	:	376		
Gar TAPBPR	:	VIQLHITQTPSVTLTPNQ	--LV-FQDDT	P--QRVIC	HCDHYYPL-DVQVEW	FSLS	-----	SASEPVPIS	:	359			
Medaka TAPBPR	:	VIQLHIIQPPHISLSEEK	--LM-LK--TP	--QTLKCF	SRNYYPL-DAQMEW	LFVSP	-----	TATEPELFFK	:	349			
Salmon TAPBPL1a	:	DITLQIMESPRVSLNVDS	-TIS-LVDRGE	--QKVVC	EAEGYYPL-DVEMEW	FR	-----	GGGLLEPEKL	:	332			
Salmon TAPBPL1b	:	EIALHIMEPPRVSLNVDS	-VIS-LVDTER	--QKMOV	CAVGYYP	L-DVEMEW	FR	-----	GGGLLEPEKL	:	335		
Zebrafish TAPBPL	:	DIPLSLSEQPRVSLNVGS	-TLS-MVIGQD	--QKLIC	EAESYYPL-DVDIEW	YREPH	-----	GGSPPLFL	:	340			
Medaka TAPBPL	:	ELNLHVEEPPRVSLNVGP	-TLV-LEEGKE	--QKVVC	TAENYYPL-DVEILW	HEQN	PAASGQ	RVGAPLPQEL	:	335			
Salmon TAPBPL2	:	DIALHII	EPPNVSLTTKE	-----	EQI--PRVVC	EA	NGFYPL-DVDIDL	FKETS	-----	SGQRP	PEKL	: 322	
Gar TAPBPL	:	DVQLQIMESPRVSLSVAT	-ALA-LQEEEE	--QKVVC	EITGYYPL-DVQVDW	LHEQG	-----	GGGRLPQVI	:	335			
Frog TAPBPL	:	VIHLEVVEPPIVNLNVKS	--LL-LTEGQE	--QKLVC	EA	SKYYPL-DVNIHW	LEGR	-----	DQTLPLPTVL	:	339		
Chicken TAPBPL	:	AIQLQITEKPTVTNVNS	--LS-LVEGEQ	--QKLVC	DVRNYYPA-DIHVQW	LE	REPQ	-----	SAGQLPDTV	:	344		
Kiwi TAPBPL	:	AIQLQIEEKPTVIVNVNS	--LS-LVEGEQ	--YKLIC	D	IRNYYPL-DAQAQW	LRELK	-----	GSRKVPD	VV	: 347		
Turtle TAPBPL	:	NIQLEIVEKPTVTNLADV	--LS-LVEGDE	--HKLVC	D	ISHFYPH-DADAQW	LQEP	-----	EQGMLPHVV	:	345		
Opossum TAPBPL	:	SIEVAVVEAPKVR	FSPTILT-LEEGEE	--EKLVC	DVRHYFPL-EAHVFW	L	RE	RL	-----	EGRMM	PEGV	: 333	
Salmon TAPBPα#C	:	AVELEIVEPPSLSIFPSP	--LPLSVPGQV--VKVQC	EA	S	GF	FPL-SLDFH	WELT	-----	GPDGKVRPL	:	339	
Pike TAPBP	:	AVELEIVEPPSLSIFXP	--LPLSVPGQV--VTVQC	EA	S	GF	FPL-SLELH	WELM	-----	ETDGNVRPL	:	339	
Zebrafish TAPBP	:	AVDLEIVEPPSLSILPSP	--LPLLVP	GQV--LSVQC	EA	S	GF	FAPH-TLDL	GWEFT	-----	GADGKLSL	: 339	
Gar TAPBP	:	AMELEVVEPPALLSPSP	--LW-ALPGHS--APVQC	EA	S	GF	FPL-ALELQ	WEFLPA	-----	GGAGEALAL	:	337	
Chicken TAPBP	:	VLQLHVFEPPKVTLS	PKN--LV-VAPGMS--AELRCH	V	SGFYPL-DVTVTW	QRRAG	-SSG--TSQS	PRD	TV	: 330			
Human TAPBP	:	TLELAVYKPPKVS	LMPAT--LARAAP	GEAP-PELLCLV	S	HFYPSGGL	EV	EWELRGG	-----	PGRSQKAE	:	344	
		*		280	*		300	*		320			
		TC2-TC3				C6							
		*	340	*		360	*		380				
Human TAPBPR	:	SGAS	FSSLS	QSVAGTYS	SISSSLTA	EPG----	SAG-ATYTCQVTHISLE	-EPLG	-ASTQVVP	ERRT	-----	: 406	
Chicken TAPBPR	:	SNTR	FSSSH	QSDGTYS	SINSYLSVNLATA	--QAP-ATYTC	HVSHVALE-APIS	-ISTHL	KAPEHTE	-----	: 404		
Salmon TAPBPR	:	DQVSL	SSSH	QHSDRTMS	ISSHLTLHPSTF	--PPG-TTVTC	RVTHPALD-TPLS	-LSLT	VETPEPD	-----	: 413		
Zebrafish TAPBPR	:	KDSSL	SSSH	QHSDGTFS	VSSHLTLLPDKH	--PPG-TVVTC	RVSHPSLE-TPTN	-ATLT	VEEPEDT	-----	: 437		
Gar TAPBPR	:	NGVYF	SSSH	QHSDGTYS	LAYSIVSPSED	--LTG-ATFSC	IVSHHLS-EPIT	-ASITV	SAPEES	-----	: 419		
Medaka TAPBPR	:	DQGS	LTSH	QHSDGTFS	LSSHLVPPAV	--PPG-TKIIC	RVSHVGLD-APLD	-VSVLV	ETPKID	-----	: 408		
Salmon TAPBPL1a	:	DTVLY	SSSH	HHQDGTYS	LSAFFLLHASLH	--DSG-SKYFC	RVSHSSQR-MPIR	-KSFTL	NVTEYDSWN	--A	: 396		
Salmon TAPBPL1b	:	DTVLY	SSSH	HHQDGTYS	LSAFFLLHASLH	--DSG-SKYFC	RVSHSSLR-MPIR	-KSFTL	LIVTEVDSWN	----	: 398		
Zebrafish TAPBPL	:	KNVLY	SSSH	HHQDGTYS	LSAFFYLQPGLE	--DSG-YKYTC	RVSHKSLT-TPVR	-KSFY	LVVTEPDS	-----	: 401		
Medaka TAPBPL	:	KNILF	SSSH	HHNQDGTYS	LSAFFYLKASLT	--LSG-KQF	SCSVSHPSLR-MPIK	-KSFIL	QVEEPSLM	----	: 398		
Salmon TAPBPL2	:	GNVLH	SSSH	HHQDGTYS	VAAVRLRPSQ	--DSGCE	TYSCRVSHVSLQGLYIH	-KSII	----	-GPGC	: 381		
Gar TAPBPL	:	KNILF	TS	HHADGTYS	ISAFFILRPSLA	--DHD-ARYT	CRVSHVSLR-TPIR	-KSFTL	TVTERLRIGSIL	: 401			
Frog TAPBPL	:	KNVLY	SSSH	HHNQDGTYS	LS	SSFFIFTASQL--DNG-AVLT	CRVEHMSLK-HPIR	-KSVK	VRVEESPSY	----	: 401		
Chicken TAPBPL	:	PNVLT	SSSH	LRSSNGTYS	SFSRFFLLTATLR	--DNG-HTYT	CRVEHSSIQ-APIR	-RSIT	VAVREATS	-----	: 405		
Kiwi TAPBPL	:	KNVL	SSNH	QSSNGTYS	SFSRYFLLTASLE	--DNE-HTYT	CRVDHQSILQ-APIR	-RSVIV	QVRGTS	-----	: 407		
Turtle TAPBPL	:	THVVS	SSSH	HHNSDGTYS	FSSYFLLKASLR	--DDG-RRYT	CRVEHQSILK-HPIK	-KSLT	VKVTESTS	-----	: 406		
Opossum TAPBPL	:	KNVLF	SSSH	KNNGDETF	S	SSYYLLKASLQ--DDG-FRYT	CLVEHEGLR-FPIR	-KSIII	RVTEKSG	-----	: 394		
Salmon TAPBPα#C	:	GQGS	VTGH	FGPDNTYS	QTSRLELDSAKL	DLGRG-GEVTC	VAVHPGGT----	RRASVTL	NVIGINGPSIED	: 405			
Pike TAPBP	:	GQGS	VSGH	FGPDGTYS	QTTTLQLDSAKL	DLGRG-GEVTC	VAVHPGGT----	RRARVTL	NVIGVSAPSIED	: 405			
Zebrafish TAPBP	:	GQGS	VTGH	FRASDGTFS	QSSRLELDSTKLRLARG	-GEIS	CAKH	DGGT----	RRASAAL	NVIGVAAPSIED	: 405		
Gar TAPBP	:	EGATV	S	GH	FG	RAGTFTQTSRLEL	DLGRLEHG-GQLVC	VAKH	DGGT----	RRPGVTL	NIAGVGGPTVED	: 403	
Chicken TAPBP	:	MDSWT	SGH	QAADGTYS	R	TAAARLIPARPQ-HHG-DVY	SCVTH	TALA-KPMR	-VSVRL	LLAGTEGPHLED	: 397		
Human TAPBP	:	GQRWLS	ALRHHS	DGSVSLSGHL	QPPVTTTE-QHG-ARY	ACRIH	PSLP-ASGRSA	EV	TL	EVAGLSG	PSLED	: 412	
		*		340	*		360	*		380	*		

		*	400	*	420	*	440	
Human TAPBPR	:	ALGVIFA--SSLFLLALMFLGLQRRQAPTGLGLLQAERWETTSCADTQSSHLHEDRTARVSQPS-						: 468
Chicken TAPBPR	:	LEGLVGGAIATAIFVSVLFIVLRRKRAAEPKPEQLLTASE-----						: 444
Salmon TAPBPR	:	---SYWMVLGFLVITVLFYQVMK-----						: 434
Zebrafish TAPBPR	:	---AFWTISLMLLSVVFLYQAFK-----						: 458
Gar TAPBPR	:	QLWSMVGGLSFLVFLFGLFMLLR-----						: 443
Medaka TAPBPR	:	---SYWWFLGFLIVTILFFYSFKK-----						: 429
Salmon TAPBPL1a	:	ALWFFSGFGFILVMVATLFLVML-PR	SSARKANQRKPY-----					: 433
Salmon TAPBPL1b	:	FLWLFFGCGFILVMVATLCVML-PR	SSARKANKRKPYP-----					: 435
Zebrafish TAPBPL	:	TMWYIAVFGFIAMLVILCYML-PQ	VLARR--SKKMF-----					: 436
Medaka TAPBPL	:	FNLAVGLVLISLSGVLIIILLY-LNS	ARRPSKIVRNHTESQW-----					: 439
Salmon TAPBPL2	:	WTWHLTLPLVCLIFVIFIAVTRQ	RFF-----					: 410
Gar TAPBPL	:	FWIFVAFIVMMLLLLVLIP--YL	HSAARREAKKKKPY-----					: 437
Frog TAPBPL	:	QEFFLVI--LIIFLSVLLTLILH-----						: 423
Chicken TAPBPL	:	TTWLLLL--LLGLTGCLVASL-H	HLHQVRSTTKPKPY-----					: 440
Kiwi TAPBPL	:	TIWVLL--LLGLSVCLAVVL-H	YFYKVKSTAKPKPY-----					: 441
Turtle TAPBPL	:	ATWFLVI--LLVLLAGCLVVTI-C	YLYRGMGPNKPKPY-----					: 441
Opossum TAPBPL	:	SSWGLLI--ILIIIMVVIFVLL-R	YLHQVKSMNKTKPY-----					: 429
Salmon TAPBP _a 27	:	SMAMVAVALGLYGLIKIVSWTF-SS	GSDDTNSQEKVK-----					: 442
Pike TAPBP	:	SMAMVAVALCLYGLIKIVSWTFI--	SSGSEDAEKKAK-----					: 440
Zebrafish TAPBP	:	SMAMVAALLLYGMIKFLSWTFSS	SDSELNDKKEK-----					: 443
Gar TAPBP	:	SMAMVAALLLYGLIKIVSWTLKT	DSSVPEEPEKKQE-----					: 441
Chicken TAPBP	:	ITGLFLVAFVLCGLIRWLY----	PKAARPKETTKSQ-----					: 430
Human TAPBP	:	SVGLFLSAFLLLGLFKALGWAA--	VYLSTCKDSKKKAE-----					: 448
		400	*	420				
			TAP association		ER retention motif			

References:

1. Dong G, Wearsch PA, Peaper DR, Cresswell P, & Reinisch KM (2009) Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* 30(1):21-32.
2. Hermann C, Trowsdale J, & Boyle LH (2015) TAPBPR: a new player in the MHC class I presentation pathway. *Tissue antigens* 85(3):155-166.
3. Morozov GI, *et al.* (2016) Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *Proceedings of the National Academy of Sciences of the United States of America* 113(8):E1006-1015.
4. Rizvi SM, Del Cid N, Lybarger L, & Raghavan M (2011) Distinct functions for the glycans of tapasin and heavy chains in the assembly of MHC class I molecules. *Journal of immunology* 186(4):2309-2320.
5. Sadasivan B, Lehner PJ, Ortmann B, Spies T, & Cresswell P (1996) Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* 5(2):103-114.
6. Paulsson KM, Jevon M, Wang JW, Li S, & Wang P (2006) The double lysine motif of tapasin is a retrieval signal for retention of unstable MHC class I molecules in the endoplasmic reticulum. *Journal of immunology* 176(12):7482-7488.
7. Neerinx A, *et al.* (2017) TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *eLife* 6(e23049).

Additional file 13: Figure S9. PSME1 and PSME2 data

Table of Contents		Page
S9a	Phylogeny of deduced PSME1, PSME2 and PSME3 amino acid sequences	1
S9b	Alignment of deduced PSME1, PSME2 and PSME3 amino acid sequences	2

S9a. Phylogeny of deduced PSME1, PSME2 and PSME3 amino acid sequences

The evolutionary history was inferred using the Neighbor-Joining method (1). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches (2). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (3) and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair. Sequence accession numbers are shown in parenthesis and Atlantic salmon (Salmon) sequences are shown with red font.

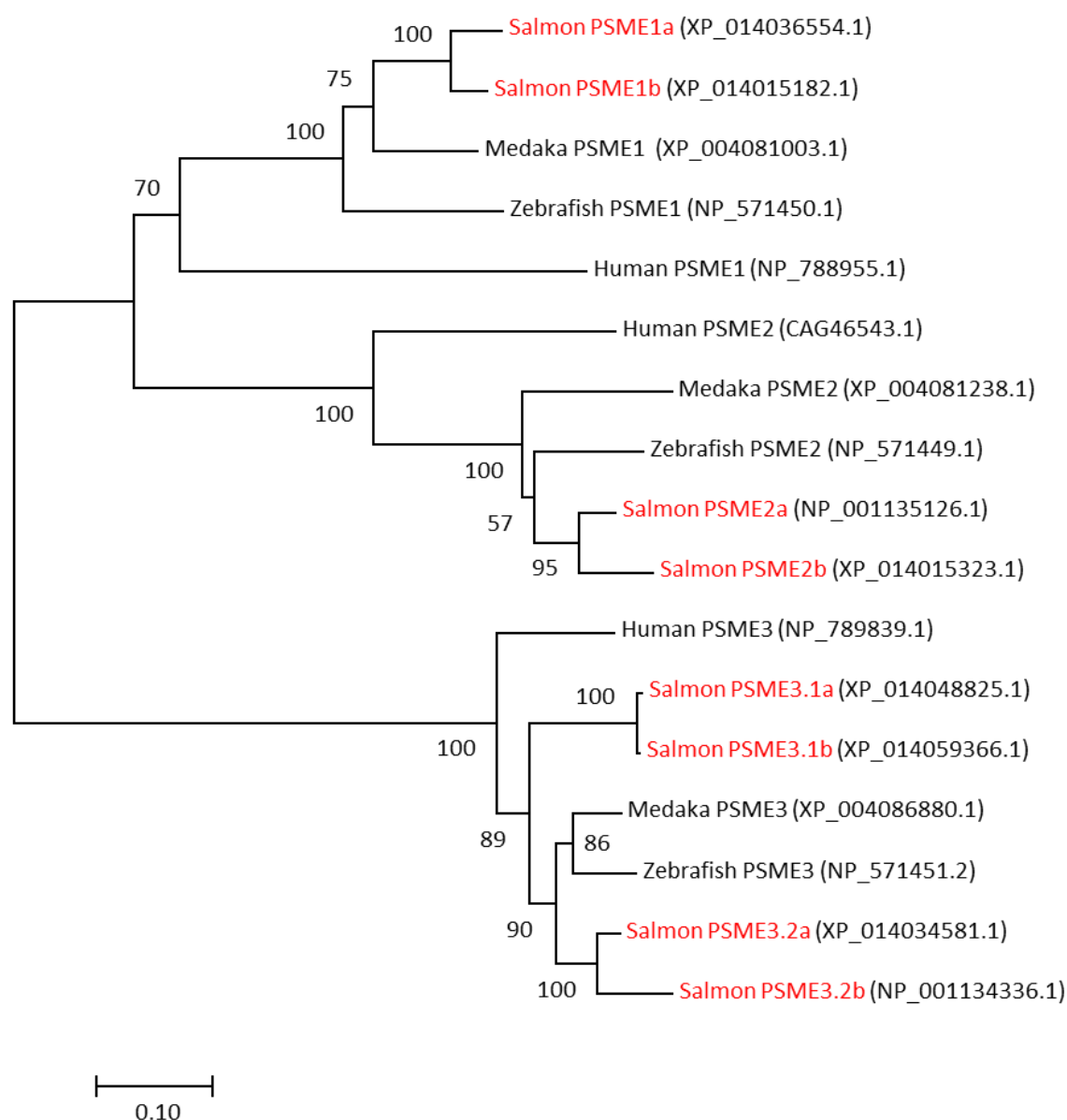


Figure S9b. Alignment of deduced PSME1, PSME2 and PSME3 amino acid sequences

Alignment of deduced PSME1, PSME2 and PSME3 amino acid sequences from Atlantic salmon (Sasa), medaka (Orla), zebrafish (Dare) and human (Hosa). Sequence references are shown in Figure S9a. Amino acid residues are colour coded according to physiochemical properties.

```

      *           20           *           40           *           60
HosaPSME1 : MAM---LRVQPEAQAKVDVFREDLCTKTENLLGSYFPPKKISELDAFLKEPALNEANLSNLIKAPLDIPVP : 66
SasaPSME1a : .TS---IDIR..SKKM..D.CTR.TKEA.T.VT.F..Q..A.MEML..K-SFSTDG.AA..S....M. : 65
SasaPSME1b : .TS---IDIR..SKKQ..D.CTQ.TKEA.T.VT.F..Q..A.M.ML..T-S.STEG.AA.....I. : 65
OrlaPSME1 : .TS---DIRH.SKKQ..E.CQT.SKEA.E..SKF..E..EQ.QML..T-SF.CDD.AS.....I. : 65
DarePSME1 : .TS---DMS.ASKKQ..G.SQKITKEA.Q.ISKI..E..A.M.NV.QG-SCSLKD..VI.....I. : 65
HosaPSME2 : ..KPCGV.LSG.VRKQ.E..QN.FQEA.EF.YRFL.Q..IY.NQL.Q.DS..V.D.TS.R.....I. : 69
SasaPSME2a : .SRSSV.KIKSANAV..EN..QS.YQQA.D.FSN.I.L..TQ..NL...ED.SITD..T.H.....I. : 69
SasaPSME2b : .SKSSVMKIKSVNAV..EH.HHS.YQQADD.FSN.I.L..Q..NL...DD..IPD..T.Q.....I. : 69
OrlaPSME2 : .SKAAT.KIKSASAV..EN..QS.YHEAG..FSNLI.Q..IQ...L.RDD.FILTD--S.Q.....I. : 67
DarePSME2 : .SKANV.KLNSDNAVRIENY.QS.YKQA.D.FSN.I.L...H..NL...GDEF.SITD..S.H.....I. : 69
HosaPSME3 : ..S--L.K.DQ.VKL...S...RITSEA.D.VANF....LL...S....I..IHD.TQIHSDMNL... : 67
SasaPSME3.1a : .SSK-GIK.DNDLKT...A...RITGEA..VADF....LL...S....I..V.D.KEIHSEINVK.. : 68
SasaPSME3.1b : .SSK-GIK.DNDLKT...A...RITGEA..VADF....LL...S....I..V.D.KEIHSEIN.K.. : 68
SasaPSME3.2a : .NS--L.K.DN.LKT...A...RITGEA.E.VA.F..N.LL...H...D.G..ICE.KEIHSEINLT.. : 67
SasaPSME3.2b : .TV--VVF..-----A..QRITGEA..VA.F..N.LL...H...D.I..ICE.KEIHSEINLT.. : 61
OrlaPSME3 : .SS--L.K.DS.IKT...A...RITAEA.D.VANF....LL...S...D.SI.IVE.KDIHSEINLT.. : 67
DarePSME3 : .SS--L.K.DN.IKT...A...RITSEA.D.VANF....LL...Q...D.TI.ITE.QEIHSEINLT.. : 67

```

```

      *           80           *           100          *           120          *           1
HosaPSME1 : DPVKEKEKEE---RKKQQEKEDKDEKK---KGEDEDKGPFGPVCNCNEKIVVLLQRLKPEIKDVIEQL : 128
SasaPSME1a : ..A..EA.-----..KE...A.EG..EKDSK...A..A...IP...RVES..KEI..Q.QILK.K. : 129
SasaPSME1b : ..A..EA.-----Q..KE...A.EG..DKDSEK...A.....IP...RVES..KEI..Q.QLLK.K. : 129
OrlaPSME1 : ..A..E..-----..KE..DA.EG..DKDSK...A.....IYS..RVES...EV...QTLK.K. : 129
DarePSME1 : ....EL.-----..KE...A.EG..---DK...A.....IA...TVEK.IKQI...QTLK.C. : 125
HosaPSME2 : ..P-P.DD-----EMETD.-----QEKKV.K..FLPG..VLS..ALV...VWTLK.KC : 118
SasaPSME2a : ..PTPEDE-----EMETD.N.-----DD.KKK.A.K..FIKG....K..D.V...IALR.TI : 123
SasaPSME2b : ..PTAEDE-----EMETD.N.-----DKKKK.A.S..LIK....K..D.V...LSLR.TI : 123
OrlaPSME2 : E.PSPE.E-----DMETD.DE-----D.KKK.A.K..FIKG...MT..E.V...VALR.TI : 120
DarePSME2 : ..PAPEDE-----EMETD.NE-----DD.KKK.A.K..FIKG..R..K..DIV...MGLK.TC : 123
HosaPSME3 : ..ILLTNSHDGLDGPTYKKRRLDEC--EEAFQ-GTKVFVM.N.MLKS.QQL.DIIEKV...RLL.KC : 133
SasaPSME3.1a : ..IILNNSHDGVDVQNSRKRKME.GLEDDNCQDGPVFAM.G.MIKS.GQL.D.IE.V...RTL.KC : 137
SasaPSME3.1b : ..IILNNSHDGVDVQNSRKRKME.GLDDNCQDGPVFAM.G.MIKS.GQL.D.IE.V...RTL.KC : 137
SasaPSME3.2a : ..IILSNLHDGLEAQNNAKKRME.GSGEDKVA-GTKVFVM.G.MMKS.A.L.D.IEKV...RTL.KC : 135
SasaPSME3.2b : ..IILSNLHDGLEAQNNAKKRME.GTGEDKVA-GTKVLIM.S.MMKS.A.L.G.IEKV...RTL.KC : 129
OrlaPSME3 : ..IILPNLHDGLEAQNNAKKRLE.GSGEDKMT-GTKVFVM.G.MMKS.GNL.D.IEKV...RTL.KC : 135
DarePSME3 : ..ILLTDIHDGIEGQNAKKRKF.E.GAGDDKVG-GTKVFVM.G.MMKS.G.L.E.IEKV...RTLM.KC : 135

```

```

      40           *           160          *           180          *           200
HosaPSME1 : NL-----VTTWLQLQIPRIEDGNNFGVAVQEKVFEELMTSLHTKLEGFHTQISKYFSERGDA : 184
SasaPSME1a : .T-----SM.V....K.....L.NTR..I..Q.....Y..... : 185
SasaPSME1b : .T-----SM.V....K.....L.NTR..I..E.Q.....Y..... : 185
OrlaPSME1 : .T-----SM.V....K.....L.NTR..I..AIQ.....Y..... : 185
DarePSME1 : .T-----SM.I...V.....NTR..I..Q.....Y..... : 181
HosaPSME2 : I.-----I..I.HL..K....D...I...L.RVNAVK..V.A.Q.T..... : 174
SasaPSME2a : IT-----SC.I.HL..K....D...I...IL.RIAVAVK..VD...N.N..... : 179
SasaPSME2b : IT-----SC.I.HL..K....D...I...IL.RIAVAVK..VD...N.N..... : 179
OrlaPSME2 : IT-----SC.I.HL..K....D...I...IL.RIAVAVK..VD..Q.N.N..... : 176
DarePSME2 : IT-----SC.IAHL..K....D...I...IL.RI.AVK..V...Q.N.N..... : 179
HosaPSME3 : .TPSGKGPHICFDLQ.KM.V..L.....SI..ETVAELRTVEGEAASYLD...R.YIT.AKL : 202
SasaPSME3.1a : .T-----KM.V..L.....SI..ETVAELRTVEGEAASYLD...R.YIT.AKL : 193
SasaPSME3.1b : .T-----KM.V..L.....SI..ETVAELRTVEGEAASYLD...R.YIT.AKL : 193
SasaPSME3.2a : .T-----KM.V..L.....SI..ETVAELRTVEGEAASYLD...R.YIT.AKL : 191
SasaPSME3.2b : .T-----KM.V..L.....SI..ETVAELRTVEGEAASYLD...G.YIT.AKL : 185
OrlaPSME3 : .T-----KM.V..L.....SI..ETVAELRTVEGEAASYLD...R.YIT.AKL : 191
DarePSME3 : .T-----KM.V..L.....SI..ETVAELRTVEGEAAS.LD...R.YIT.AKL : 191

```



```

      *      220      *      240      *      260      *
HosaPSME1 : VTKA AKQPHVGDYRQLVHELD EAEYRDIRLMVMEIRNAYVRRQGQGRGGQRQLSQATHSLTLQARG- : 250
SasaPSME1a : .D..S.E.....QYQ.CEL.IV.L...T.AVLLDIINKNYDKIKKPRGDC--K.LIY : 250
SasaPSME1b : .A..S.....QYQ.CEL.IV.L...T.ALLFDIINKNYDKIKKPRGDC--K.LIY : 250
OrlaPSME1 : .A..S.....QYQ.CEL..HILD...T.AVLFDIINKNYDKIKRPRGDG--K.LIY : 250
DarePSME1 : .A..S.....F.....QHQC.E.L.II.L...T.AMLYDVITKNFDKIKKPRGD.SSK.LIY : 248
HosaPSME2 : .A..S.ET..M...A...R...A.VEL.A..LDL.AF.AELYHIISNLEKIVNPKGEE--KPSMY : 239
SasaPSME2a : .S..S.IT..M...S...K...V.S...VILLD..GF.AELYDIISKNI EKVTNPKGEE--KPSMY : 244
SasaPSME2b : .A..S.ST..M...S...K...V.S...VILLD..GF..ELYDIISKNI EKVTNPKGEE--KPSMY : 244
OrlaPSME2 : .A..S.ET..M...S...K.R.LFS.L.VILLD..GF.AELYDIINKNI EKVTNPKGEE--KSSMY : 241
DarePSME2 : .A..S.DT..M...S...K...A.SE..VI.LD..GF.AELYDVISKNI EKVTNPKGEE--KPSMY : 244
HosaPSME3 : .S.I..Y...E...RT.T.I..K..ISL..IIS.L..Q..TLHDMILKNI EKIKRPRS.N--AETLY : 267
SasaPSME3.1a : .S.IT.Y...E..QRT.T.I..K..ISLKII.S.L..Q..TLHDMILKNI EKIKRPRS.N--NEALY : 258
SasaPSME3.1b : .S.IT.Y...E..QRT.T.I..K..ISLKII.S.L..Q..TLHDMILKNI EKIKRPRS.N--NEALY : 258
SasaPSME3.2a : .S.I..Y...E...RT.T.I..K..ISLKII.S.L..Q..TLHDMILKNI DKIKKPRS.N--AEALY : 256
SasaPSME3.2b : .S.I..Y...E...RT.T.IG.KK.ISLKVI.S.L..Q..TLHDMILKNI DKIKKPRS.N--AEALY : 250
OrlaPSME3 : .S.I..Y...E...RT.T.I..K..ISLKII.S.L..Q..TLHDMILKNI EKIKRPRS.N--SDALY : 256
DarePSME3 : .S.I..Y...E...RT.T.I..K..ISLKII.S.L..Q..TLHDMILKNI EKIKRPRS.N--TDALY : 256

```

References:

1. Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 4(4):406-425.
2. Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution; international journal of organic evolution* 39(4):783-791.
3. Zuckerkandl E & Pauling L (1965) Molecules as documents of evolutionary history. *Journal of theoretical biology* 8(2):357-366.

Additional file 14. Figure S10. ERAP data

Table of Contents		Page
S10a	Phylogeny of deduced ERAP1 and ERAP2 amino acid sequences	1
S10b	Alignment of deduced ERAP1 and ERAP2 amino acid sequences	2

Figure S10a. Phylogeny of deduced ERAP1 and ERAP2 amino acid sequences.

The evolutionary history was inferred using the Neighbor-Joining method (1). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches (2). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (3) and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair. Sequence accession numbers are shown in parenthesis and Atlantic salmon sequences are shown with red font.

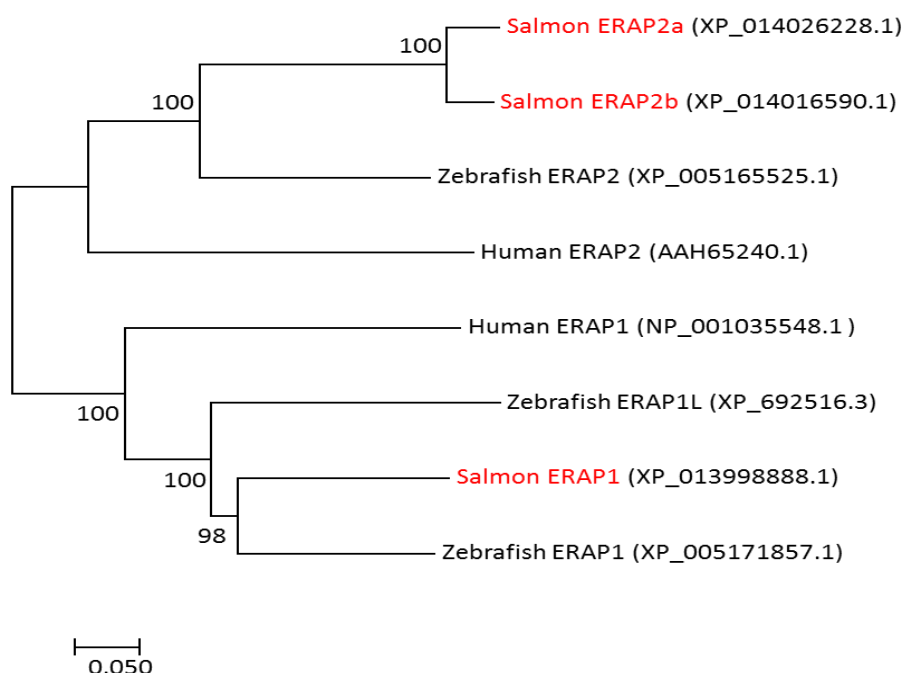


Figure S10b. Alignment of deduced ERAP1 and ERAP2 amino acid sequences

Alignment of deduced ERAP1 and ERAP2 amino acid sequence from Atlantic , medaka, zebrafish and human. Sequence references can be found in Figure S10a. Amino acid residues are colour coded according to their physiochemical properties and numbering above the alignment is consecutive.

		*	20	*	40	*	60	*	80		
SasaERAP2a	:	-----	MVRFVLVALLSLAGV	-	TQTSASPTQASEPPNPT	--	EEQPPGLGTGSLSFPSW	SHLRLPGYIVPLHYHL	:	63	
SasaERAP2b	:	-----	MFW.....F.....	-----	S.....S.....	-----	N.....N.....	E.....E.....	:	65	
OrlaERAP2	:	-----	MASIKL.E.L.C.FFL	GCHS.Q.SQ.V.SA.H.K	-	G.PS..EDN.....	R.....R.....	I.....I.....	:	66	
DareERAP2	:	-----	MHL.F.I..F.ESL	-----	CYV.S..SNS.....	-----	GDN--.A.DG.P...	KV...N...V...	:	56	
HosaERAP2	:		MFHSSAMVNSHRKPMFNI	IHRGFYC.TAILPQICIC	QFSVPS.YHFTED	-	PGAF.VA.NGER...	QE...SVVI...D.	:	80	
SasaERAP1	:	-----	RTLTL.TI.VL.HVS	FAPSL.AQLPGDHD	T.DKSSSLL...IA	NGQP...I.M...	ETVS.I...D.		:	66	
OrlaERAP1.1	:	-----	MFPL..L.FALLPSS	HGAQIPSPDQ	KE-----	I.VA.NGQL..	DRM...KT.S...	D.	:	56	
DareERAP1	:	-----	MATMTR.PV.LNL	LVIPHLWAAVTPE	TS-----	F.ISSSGEP...	NKM...DT.S...	N.	:	58	
DareERAP1L	:	-----	M.F.AKYFHFTL	VALFCA	LFGNHAMT-----	VNKNKTV.P...	DKM...EIVK.Q...	D.	:	55	
HosaERAP1	:	-----	MVFLPLKWSLATMSF	LSSLLALLTVS	.PSWCQSTEAS-----	KRSDGTP...NKI...	E.VI.V...D.		:	65	
		*	100	*	120	*	140	*	160		
SasaERAP2a	:	LLHPNLTMLS	YSYSGTVRIELQ	VQNNTN	NWVVLHSGKGLR	ITATTATMLDQ	NLAHLS--	DRVLPVLHNP	THEQTAFISPRVLSG	-GQK	: 142
SasaERAP2b	:	I...I.....			V.....	-----	Q.....	V.....	A.T....		: 144
OrlaERAP2	:	V.Q.....T.RFT	S.Q.QID.....	S.....	D.Q.SK..I..H.F.	I..T--	Q.....	S.....	I.....	S....	: 145
DareERAP2	:	.I.....T..FT	S.K..ID.K.....	N.K.Y..V..EHE	--	EKT.S..EY.L..I...		KI.TS..E.			: 135
HosaERAP2	:	FV.....S.DFVASEK	.VL.S.A.QFII...	D.E..N..LQSEEDS	RYMKPGKE.K..SY.A...	I.LLV.EK.TP	-HL.				: 161
SasaERAP1	:	.V.....S.DFT	E.Q.Q..FED	STII...	D.Q.AK.EL	APEGPGSL--	PVP.Q..EY	AFH.L.LM	DVL.VR..GM		: 145
OrlaERAP1.1	:	S.....T.DFT	V...D.D.HED	SV...A.KMY	SS.LL	APAG-----	T...Q..EY	RFH.L.LM	DV..TK..R.		: 131
DareERAP1	:	.I.....S.DFT	S.Q.QIE	LQD.KTII...	N.Q.QS.RL..A.I.Q	----	QQP.K..EY	YFQ.I.LV	DKA.LKR	HV	: 136
DareERAP1L	:	.I.....S.TFT	E.Q.QIE	KQD.RAII...	N.QVSK.LL	GSRQH.H--	HQD.QISE	FEAN..I.L	EGFTF	-KGS	: 133
HosaERAP1	:	.I.A...T.TFW..TKV	.ITASQP	STII..HH.Q	SR..LRKGAGER..	--	EEP.Q..EH	RQ..I.LLA	EP.LV..LP		: 144
		*	180	*	200	*	220	*	240		
SasaERAP2a	:	YFLFLEFGAEL	GEGFYGYRSTYRTS	AGETRNLASTH	FEPTSARMAFP	PCFDEPSFKANYSIS	IRRSLAHTALS	NMPVEQ	TVV	:	224
SasaERAP2b	:	D.....	T...T.....		I.....		P.....		E.	:	226
OrlaERAP2	:	...YI.....FA.....	K...K...K...T.....		F..Q...	PQYIS.....	IVK..E			:	227
DareERAP2	:	...Y.....P.SD..H..K...K...	KS...V.....L.....	TVR...GPS.I.....	L...E					:	217
HosaERAP2	:	.YVAMD.Q.K..D.E..K...	LG...I..V.D...Q.....	L...F..K..ESR	I.....KVK..E					:	243
SasaERAP1	:	.KVR...S.N.SDS	H..K.S...TK..V.FM..Q..A..A...	A...A...FT.Q...	ESR.I.....KVK..E					:	227
OrlaERAP1.1	:	.EVQ...A.N.SDS	YH..K.S...T..V.VM..Q..A.F..G.....	A...A...FT.R.I	EPR.I.I...KVR..E					:	213
DareERAP1	:	.SVE.H.A.N.S.S.H..K...	K.DV.VV..Q..A..A...A...A...F.VQ...	EAK.I...KLR	LE					:	218
DareERAP1L	:	HVVH...Y.N.SDS	H..KGQ.T.NS..V.M...Q...H..A...	A...F.RV..ESR	ISI...KLR..E					:	215
HosaERAP1	:	.TVVIHYAGN	S.T.H..K...KE..L.I...Q...A.....	A...A...SF..K...EPR	L.I...LVKS	T				:	226
		*	260	*	280	*	300	*	320		
SasaERAP2a	:	LDDGLMEDR	FAVSVRMSSYL	VAFIVCDFRS	VSATTASGVKVS	VYAAPEKWQQ	THYALKA	AAVKLLEF	YKYNIKYPL	PKQDL	: 306
SasaERAP2b	:	S.....	T...T.....	L...K.....	S.....						: 308
OrlaERAP2	:	.HG..L..H..P..K..T...	VI..K..T...I..I.....	E.....EV...M.D..E...	P.....						: 309
DareERAP2	:	ISN..F..H.EA..K...L...	I...K...GL..T.INI	I..V...H..Q...E..LR...	Q...L...L...						: 299
HosaERAP2	:	.EG..L..H.ETT.K..T...	Y...H.L.GF.S...I..S.D	RN...Q.SL..D...D.Y...S.L...							: 325
SasaERAP1	:	.PG.VL..H.DTT...T...	S...Q...K..SH...I..V.D	IN.DF.N...R..D.DD..D.P...							: 309
OrlaERAP1.1	:	.PGD..L..H.DTT...T...	Y..S..Q...R..QH...I.I..V..IN	AF..D...D.DD..D.P...							: 295
DareERAP1	:	.KN..F..Q.D...K..T...	Y..S..L.I.K.SQH..QI...V..ID	AEF..D...D.DD..D							

```

SasaERAP2a : YPKLRVEDYLVDTCFAAIGRDSLNSSRPISSAAESPTQIAEMFDTVSYDKGACVLHMLRHYLTDQVFQSGIVRYLRYSYSN : 470
SasaERAP2b : .....L.....V.....V.....K.....N.....M..... : 472
OrlaERAP2 : ..N.K..E..LH.....H.....P..N...K.....F...E.....K..FG.. : 473
DareERAP2 : ..E.K...HFL...G.....L.....K.....E...I.....F...EG....I....F..C.. : 463
HosaERAP2 : ..E.QFD..FLNV..EV.TK.....KP..T...Q...E...N...I.N..KDF.GEEK..K..IQ..KKF..R.. : 489
SasaERAP1 : N.E.Q.N.FFLGK..E.MEV...S..H.V..QVDT...Q...D.....I.N...DF..PEA.KI.....Q.. : 473
OrlaERAP1.1 : ..E.Q.D.FFLGK..E.MEV...S..H.V.TPV.N...Q...D.....I.N...DF..PEA.EI...I...K.F..Q.. : 459
DareERAP1 : N.E.Q...FLGK..E.MEV...S..H.V.TPV.N.A..Q...D.....I.N...EF..PKI.KL...K..KH...Q.. : 464
DareERAP1L : H.E.Q.N..FLEK..T.LSV...S..H.V.TPV.N.AE.S...D...R...I.N...DF..PE..KY..IN..KKH...Q.. : 461
HosaERAP1 : H.E.K.G..FFGK..D.MEV.A...H.V.TPV.N.A..R...D.....I.N...E..SADA.K...Q..QKH...K.. : 472

500 * 520 * 540 * 560 *
SasaERAP2a : AHNQDLWDSLANTCSEEEFTSG----EHCYSSRQAANKNAYLYAGEHLDLTTMMNTWMLQGTGVPLVTTRQGSRLLLKQERF : 547
SasaERAP2b : .R.....P.....G.....NS.....T.....A.....V..... : 549
OrlaERAP2 : .Q.....I.....K...NRD.SN.....F.....AI...T..K.I.....K.PL...R.H.. : 550
DareERAP2 : .R.E.....IK.....D.AA----Y...A..T...RF...V..KK...T..K.I.....K..RK.HIG... : 540
SasaERAP2 : .K.DD...S..S.S.L.SD...--GV.H.DPKMTS.MLAF.L.NAEVKE..T..T..K.I..LV.KQD.CS.R.Q... : 566
SasaERAP1 : TV.SH..E..T.I.QSDDLDE.RLKDEGF.SQEKAGSGAPKW.S.DQ..VRAI.D..T..E.F....EVR.RQVR.S...Y : 555
OrlaERAP1.1 : TV.SH..E..TDI..SDDLDE.RMKHK.F.SKRKDESGASKW.S.DE..VRAI.D..T..E.F....EVK.REVR.S... : 541
DareERAP1 : TV.SN..E..T.V.DSDGLDE.RLKGD.S.RH.ASNTGASKWHSDE..VKA..E..T..E.F..I.IAVK.REVR.S...Y : 546
DareERAP1L : TV.SH..E..T.I.TSDGLD..RLKLDGF.SKHTAETPAKWFKEDSV.VGAI.D..T..E.F..I..EVK.QEVT... : 543
HosaERAP1 : TK.E....M.SI.PTDGVKG---MDGF.SR.QHSSSSSHWHQEG-V.VK.....T..K.F..I.I.VR.RNVHM...HY : 549

580 * 600 * 620 * 640 *
SasaERAP2a : LRTAHPSDPAWPSLQQGFLWHIPLTYRTDTSTSIHRHLMTTLTDSDVDGEEVGWVKVNVDMAGYYLVHYDGSQWNLILQLLK : 629
SasaERAP2b : ...T.....Y.....E.....D..... : 631
OrlaERAP2 : .K.VL...Q.S.....A.SI.....STS..IHI..AS...I.S..T..V...EDD...VMAK.MR : 632
DareERAP2 : .KIVL.D..S.H...D.Y.....K.SH.DHEVK.IIDKKS.VLLLD.H.D...L.T..N...I...EE..NA.TE... : 622
HosaERAP2 : .QGVFQE..E..RA..ERY.....S.SS.NV...ILKSK..TL.LP.KTS..F...SN...I...E.H...Q..TQ.N : 648
SasaERAP1 : .K.DD-----QTH...QV...I.SS.STV..F.IK.R..VLYLP...I.F...S...M...E.E..RS.TS..L : 631
OrlaERAP1.1 : ..DN-----TE...Q...M.SA.R.V..F.LK.KS.VLYLP...D..F..N.S...M...G.G..NSI.S..H : 617
DareERAP1 : .KSDDL-----QTSS...QV...I.SD..TV..F.IK.K..VLYLP...D.I.F...S...I...E...D..T... : 622
DareERAP1L : .KG.ES-----NSSS...QV...I.SG.NAVQ.F.LK.ER.VFYLP.K.E.I.F...LR...I...ESG...C..KQ.R : 619
HosaERAP1 : MKGSDG-----APDT.Y...V...FI.SK.DMV..F.IK.K..VLILP...E.I.F..G.N...I...EDD...S.TG... : 625

660 * 680 * 700 * 720 * 7
SasaERAP2a : DNHTALSFMDRTHLIHNAFQLTTAGRLSLDKALDLIGLYRSESHTVPLLEGLGYLEAFYRMVERRDIPDVTQNLRTYILWYF : 711
SasaERAP2b : N.....S.....Q.A.....M.....S.....HI.. : 713
OrlaERAP2 : E.....YK.....VS..H...N..M.....QL.K.....NLI.KI.ELEL.KD.GR...YF.. : 714
DareERAP2 : V.....K..AS.....V.....R.....A..K..T.N...Q...QS..KLI.K.K.A...H..K...Q.. : 704
HosaERAP2 : Q...L.RPK..VG...DV...VG...T...MTY..QH.TSSPA...S...S..H.MD..N.S.ISE..KR.L.Q.. : 730
SasaERAP1 : T..R...SN..AS..N...VSV.KVG..T...SL..SK.TDIM.VTQ...E.VPL.KLM.K..MEGLENQMKG..VEL.. : 713
OrlaERAP1.1 : H.....GN..AS...V...VSV.EKVR..T..E.SL..SR.TEIMAVWQ.F.E.VPL.KLM.K..M.ELENQMKVCGVSQ- : 698
DareERAP1 : H.....SN..AS..N...VSV.K.P...TL..SK.TEIM.VTQ.FNE.VPL.KLM.K..MVELENQMKG...QL.. : 704
DareERAP1L : M...VF.SN..AS...DI...VSIKVP...N.SL..SK..EIM.VTQ.FSE.VPL.KLM.K..MQELENQ.KSHLVKL.. : 701
HosaERAP1 : GT...V.SN..AS..N...VSI.K..IE...SL..KH.TEIM.VFQ..NE.IPM.KLM.K..MNE.ETQFKAFLIRLL : 707

680 * 700 * 720 * 740 * 760 * 780 * 800 * 820
SasaERAP2a : RDVIDRQTWSDKGSVSERRRLRSELLSLACHLGLDPLCKLQAQRSFTHWLDSNSTLNLNLPADVAETVYSVGAQEDTGWASLLQTY : 793
SasaERAP2b : .G.....H.....E.....S.....T...F.....S.....HI.. : 795
OrlaERAP2 : .A...Q.....S.....V.....DYP..ER.NQH.QE..Q..G.....T.....H.H..N...D.. : 796
DareERAP2 : K...K.S...D.M..D...EDV...D.DYP...EK.E.L.VS.VK..G.IS..T..S...M...D.S.VY..RQ.. : 786
HosaERAP2 : KP....S.....WD.M...A..K...D.NHA..IQK.AEL.SQ.ME.SGK..I.T..LKI.....TTA..NY..EQ.. : 812
SasaERAP1 : .GL.....D...Q.V...Y..LFGSVRNHP.VAT.THL.NK.RA.DGNMS..S..SLA.FAI..RDPE..DF.FEK.. : 795
OrlaERAP1.1 : .EHKQNSVFNCRFDLGSSTIS--H.VAGVTNQYS----TREMLHEIRSFFDS.TEE.GSQMRCIQQSYETIEDNIRWMDRN : 773
DareERAP1 : QKL.EQ.L.T.E.....M...Y..LFG.VR.HP..VSN.TQL.NL.R..DGNMS..N..TMA.F...RTED..DF.FEK.. : 786
DareERAP1L : QPL...S...N.....M..NY..LF..VRRYPS.VST.TQL.HK.KE.DGKMW..T..SLV..TE..RT.D..DF..EK.. : 783
HosaERAP1 : ..L..K...T.E.....M...Q..L...VHNYQ..VQR.EGY.RK.KE..GN.S..V..TLA.FA...STE..DF.YSK.. : 789

800 * 820 * 840 * 860 * 880 * 900
SasaERAP2a : THSLSETHKRKILSALASSRDTNKLTRLLELGVGEVIRTQDLDSLIVMVARNPRGHHLAWSYVQKYWSTLVDFKQLGSFSI : 875
SasaERAP2b : .L.....H.....H.....L..... : 877
OrlaERAP2 : KI...AQ.N...Q..T...G..L...Q..L..K...S...S...L...K..Y...NF.K.N.DE..E.L...C.. : 878
DareERAP2 : GV.MC..E.S.F...T..K.SE..S...Q..M..T..K..N.P...Y...V..F...DF.K.H.NE..E..P...G.. : 868
HosaERAP2 : EL.M.SAEQN...Y..ST.KHQE..LK..I...M..K..K..N.AA.LHAI..R.K.QQ...DF.REN.TH.LK..D...YD.. : 894
SasaERAP1 : R..QHTSV.SR.K...SI.PLQH..K..M.QSLA..MK...PHVVIS.S...K.YK...DFLRHN.H...K..D...H.. : 877
OrlaERAP1.1 : LPL.QAWLN-----KHG.RNHHQE..... : 793
DareERAP1 : KE.MYVSM.SR.KQ..MT.PLDH..KWM.M.QSL...L.K...PYVVT.S...KAYKH..DFL.AN.DS.IK..D...H.. : 868
DareERAP1L : KR.V.PSE.WM.KA..SY.PLAH..QW...RSS...IMK...P.MLIS.SK..K.FK...DFLKS.N.GK..K..D...SA.. : 865
HosaERAP1 : QF...S.E.SQ.EF..CRTQNKE..QW..DESF.K.DK.K.EFFQILTLIG...V.YP...QFLR.N.NK..Q..E...S.. : 871

```

```

          *      920          *      940          *      960          *      980
SasaERAP2a : RNIIGTTAQFSSTEELTEVRVFFESIH-EQASQLRVTEVAMDNVQKNILWLQRNLGTLRSWLDQQID----- : 942
SasaERAP2b : .....G..F.....-.....Q.....I.....M.....N.H.----- : 944
OrlaERAP2  : .H.L....G....PDD....QL....K-.....A.QI.L..MK...R.I...E...K..NE..K----- : 945
DareERAP2  : .E..V...VT.....R..ES..K..Q-..V....IIQ..TE..E...V..K...E.M.T..QRRLN----- : 935
HosaERAP2  : .M..S....H...KDK.Q..KL...LE-A.G.H.DIFQTVLETIT...K..EK..P...T..MVNT----- : 960
SasaERAP1  : SGMVT.V.N.Y.TR.M.D...G...D.LSE.TG.G..CIQQTYESIEE..R.MDQH.PQ.KA...R.AQGARTETQGHEDL-- : 957
OrlaERAP1.1 : ----- : -
DareERAP1  : AHMVV.V.N.Y.TR.M.A...N..SLLQP.TGAE..CIQQ.VE.IEE..R.MDK..PL.KA..HRY.Y.QQKNT----- : 942
DareERAP1L : SRVVV.V.D.Y.TK.M.D..EL..G.LAQD.G.G..SIQQ.LEKI.Q...MD..VPL.KA...HHTS----- : 933
HosaERAP1  : AHMVM....N...TRTR.E..KG...S.LK-.NG....CVQQTIEETIEE..G.MDK.FDKI.V..QSEKLERM----- : 941

```

References:

1. Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 4(4):406-425.
2. Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution; international journal of organic evolution* 39(4):783-791.
3. Zuckerkandl E & Pauling L (1965) Molecules as documents of evolutionary history. *Journal of theoretical biology* 8(2):357-366.