Comparative genomic analysis of the proteasome β5t subunit gene: implications for the origin and evolution of thymoproteasomes

Yoichi Sutoh • Mizuho Kondo • Yuko Ohta • Tatsuya Ota • Utano Tomaru • Martin F. Flajnik • Masanori Kasahara

Yoichi Sutoh • Mizuho Kondo • Utano Tomaru • Masanori Kasahara (✉)

Department of Pathology, Hokkaido University Graduate School of Medicine, North-15 West-7, Sapporo 060-8638, Japan

e-mail: mkasaha@med.hokudai.ac.jp

Yuko Ohta • Martin F. Flajnik

Department of Microbiology and Immunology, University of Maryland, 685 West Baltimore Street, Baltimore, Maryland 21201, USA

Tatsuya Ota

Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies, Hayama 240-0193, Japan

Sequence data reported in this paper have been submitted to the DDBJ/EMBL/NCBI databases under accession numbers: AB624351 and AB624352.

Abstract The thymoproteasome is a recently discovered, specialized form of 20S proteasomes expressed exclusively in the thymic cortex. Although the precise molecular mechanism by which the thymoproteasome exerts its function remains to be elucidated, accumulating evidence indicates that it plays a crucial role in positive selection of T cells. In the present study, we analyzed the evolution of the β5t subunit, a β-type catalytic subunit uniquely present in thymoproteasomes. The gene coding for the β5t subunit, designated *PSMB11*, was identified in the cartilaginous fish, the most divergent group of jawed vertebrates compared to the other jawed vertebrates, but not in jawless vertebrates or invertebrates. Interestingly, teleost fish have two copies of apparently functional *PSMB11* genes, designated *PSMB11a* and *PSMB11b*, that encode β5t subunits with distinct amino acids in the S1 pocket. BLAST searches of genome databases suggest that birds such as chickens, turkey, and zebra finch lost the *PSMB11* gene, and have neither thymoproteasomes nor immunoproteasomes. In mammals, reptiles, amphibians, and teleost fishes, the *PSMB11* gene (the *PSMB11a* gene in teleost fish) is located next to the *PSMB5* gene coding for the β5 subunit of the standard 20S proteasome, indicating that the *PSMB11* gene arose by tandem duplication from the evolutionarily more ancient *PSMB5* gene. The general absence of introns in *PSMB11* and an unusual exon-intron structure of jawed vertebrate *PSMB5* suggest that *PSMB5* lost introns and duplicated in tandem in a common ancestor of jawed vertebrates, with *PSMB5* subsequently gaining two introns and *PSMB11* remaining intronless.

Keywords thymoproteasome • immunoproteasome • β5t subunit • *PSMB11* • *PSMB5*

Introduction

Proteasomes are the multi-subunit protease responsible for the generation of peptides presented by major histocompatibility complex (MHC) class I molecules (Goldberg 1995; Groettrup et al. 2010; Pamer and Cresswell 1998; Tanaka and Kasahara 1998). The 20S proteasome, a catalytic core of the larger 26S proteasome, is composed of 14 non-identical subunits that form four stacked rings of seven subunits each, $\alpha_{1\text{-}7}\beta_{1\text{-}7}\beta_{1\text{-}7}\alpha_{1\text{-}7}$ (Tanaka 2009). In standard 20S proteasomes, proteolysis is conducted by three β subunits, β1, β2, and β5, which have caspase-like, trypsin-like, and chymotrypsin-like activities, respectively. Accumulated evidence indicates that there are two specialized forms of 20S proteasomes involved in antigen presentation: immunoproteasomes and thymoproteasomes. The former contain three interferon-γ-inducible catalytic subunits, β1i (LMP2), β2i (MECL-1), and β5i (LMP7), in place of β1, β2, and β5 subunits, respectively (Belich and Trowsdale 1995; Griffin et al. 1998; Groettrup et al. 2010; Tanaka and Kasahara 1998). These subunit substitutions alter the cleavage specificities of the 20S proteasome so that the peptides binding to MHC class I molecules are produced more efficiently. On the other hand, the latter contain β1i, β2i, and β5t subunits instead of β1, β2, and β5 subunits, respectively (Murata et al. 2007); incorporation of the β5t subunit reduces the chymotrypsin-like activity of the 20S proteasome, thereby decreasing the production of peptides binding to MHC class I molecules with high affinity. β5t is expressed exclusively in cortical thymic epithelial cells (Murata et al. 2007; Tomaru et al. 2009), and thymoproteasomes appear to play a crucial role in positive selection of T cells (Murata et al. 2008; Nitta et al. 2010; Takahama et al. 2010).

Phylogenetically, immunoproteasomes occur only in jawed vertebrates, indicating that they originated from more ancient standard 20S proteasomes concommitant with the emergence of MHC- and T/B-cell-receptor-based adaptive immunity (Kandil et al. 1996). It has been suggested that, along with many other genes involved in adaptive immunity, the genes coding for β1i, β2i, and β5i subunits arose through two rounds of whole-genome duplication (WGD) (Clark et al. 2000; Danchin et al. 2004; Flajnik and Kasahara 2010; Kasahara et al. 1996; Kasahara et al. 1997). Interestingly, in both humans and mice, the gene coding for the β5t subunit (*PSMB11* in human and *Psmb11* in mouse) lacks introns and is located adjacent to the gene coding for the β5 subunit (*PSMB5* in human and *Psmb5* in mouse) (Murata et al. 2007). To gain insights into the evolution of thymoproteasomes, we analyzed publicly available databases focusing on the genes coding for β5t and β5 subunits.

Materials and methods

Bioinformatic analysis

NCBI and Ensembl databases were searched using human and mouse β5t protein sequences as queries. For the identification of shark β5t, we used the elephant shark genome database (Venkatesh et al. 2007). Genomic organization of the genes coding for β5t and β5 subunits was determined by comparing cDNA and genomic sequences when cDNA sequences were available, or by comparing genomic and predicted cDNA sequences when cDNA sequences were not available. For the construction of phylogenetic trees, amino acid sequences of β5t, β5i, and β5 subunits were aligned

using the version 6.0 MAFFT program (Katoh et al. 2002). The alignment was made using the default Auto setting, excluding the poorly conserved N-terminal region thought be removed by post-translational processing. Trees were constructed using the neighbor-joining algorithm implemented in MEGA version 4.1 (Kumar et al. 2008). The distance matrix was obtained by calculating p-distances for all pairs of sequences. Gaps were excluded using the pairwise-deletion option. The reliability of branching patterns was assessed by bootstrap analysis (5,000 replications). The average number of amino acid substitutions per site among human, mouse, and lizard sequences were obtained by MEGA5 (Tamura et al. 2011) using the JTT+Γ model (Γ parameter = 0.05-0.39). Standard errors were estimated by a bootstrap procedure (5,000 replications). Three-dimensional models of β5t, β5i, and β5 subunits were generated using the SWISS-MODEL server (Arnold et al. 2006) based on the crystal structure of the bovine β5 subunit (accession number: 1IRU) and edited by PyMol version 1.1 (http://www.pymol.org/).

Isolation of two zebrafish *PSMB11* cDNA clones

Tissues containing thymus were obtained from zebrafish purchased from a local dealer. Total RNA isolated from the tissues was converted to cDNA using M-MLV reverse transcriptase (Invitrogen, Carlsbad, CA). Based on the genomic sequences retrieved from the Ensembl database, primers were designed in the predicted 5'- and 3'-untranslated regions of the genes coding for β5t subunits. The primer sequences were 5'-ATCAGGCCTCGGAGGAGTATT-3' and 5'-ATCTGCTGTCGAATGACCAGC-3' for *PSMB11a*, and 5'-TCGACACAGTTACAATTCACTGT-3' and

5'-ACGACTTTCCCCAGAAGAACA-3' for *PSMB11b*. PCR products were cloned into the pGEM-T Easy vector (Promega, Madison, WI), and their sequences were determined using an automated sequencer. Potential PCR errors were eliminated by setting up PCRs in triplicate and sequencing multiple clones for each reaction.

Results and discussion

Phylogeny of the β5t subunit

tBLASTn searches of the NCBI, Ensembl, and the elephant shark genome databases using human and mouse β5t proteins as queries identified *PSMB11* sequences in cartilaginous fish, bony fish, amphibians, reptiles, and mammals, but not in jawless vertebrates such as lampreys or invertebrates such as tunicates and amphioxus for which draft genome sequences are available (Fig. 1; see supplementary Fig. 1 for the alignment of protein sequences used for tree construction). Thus, like β1i, β2i, and β5i subunits of the immunoproteasome, β5t appears to have emerged in a common ancestor of jawed vertebrates when primitive thymus-like tissues of jawless vertebrates (Bajoghli et al. 2011) were transformed to full-fledged thymic tissues with clear cortical and medullary differentiation. The co-emergence of β5t, β1i, β2i, and β5i in a common ancestor of jawed vertebrates is also consistent with the fact that two of the catalytic subunits of thymoproteasomes, β1i and β2i, are components of immunoproteasomes.

Previously, we pointed out that β5i subunits evolve faster than β5 subunits (Kandil et al. 1996). It is notable that the branch lengths connecting individual subunits are longer in β5t subunits than in β5i or β5 subunits (Fig. 1). This is due to faster

evolution of β5t subunits than β5i or β5 subunits, which is evident from the comparisons of human, mouse, and lizard sequences showing larger average distances (the number of amino acid substitutions per site) for β5t (0.512 ± 0.099) than β5i (0.192 ± 0.043) or β5 (0.115 ± 0.056). The observed differences among the three β5 family subunits were statistically significant at the 99% level.

β5 subunit occurs in jawless vertebrates, invertebrates, fungi, and plants. Thus, the origin of β5 is more ancient than that of β5t or β5i, indicating that β5t and β5i emerged from β5 by gene duplication in a common ancestor of jawed vertebrates. In the phylogenetic tree, the three groups of β5 family sequences, β5, β5t, and β5i, form a trifurcation (Fig. 1). Therefore, the tree topology does not provide information as to the order of gene duplication that gave rise to β5t and β5i subunits. The observation that β5i is more closely related to β5 than β5t is to β5 can be accounted for by faster evolution of β5t relative to β5i.

Thymoproteasomes and immunoproteasomes in birds

The MHC region of jawed vertebrates generally encodes at least two components of the immunoproteasome, β1i and β5i (Flajnik and Kasahara 2001). An exception is the avian MHC; sequence analysis of the chicken, quail, turkey and zebra finch MHC regions showed that they encode none of the genes coding for β1i, β2i, and β5i (Balakrishnan et al. 2010; Chaves et al. 2009; Kaufman et al. 1999; Shiina et al. 2004). In fact, birds appear to lack immunoproteasomes altogether because none of the avian genome and cDNA sequences currently available contain the sequences coding for β1i, β2i, or β5i.

Interestingly, we were unable to identify the *PSMB11* gene or its cDNA in

chicken, turkey, or zebra finch despite extensive searches of their genome databases (Dalloul et al. 2010; Hillier et al. 2004; Warren et al. 2010). Thus, birds appear to represent natural knock-out animals deficient in both immunoproteasomes and thymoproteasomes. By contrast, *Anolis* lizards have not only the *PSMB11* gene (Fig. 1), but also the MHC-encoded *PSMB8* gene coding for β5i (Fig. 2S) and the unmapped *PSMB9* gene coding for β1i (Ensembl contig AAWZ02039800, also supported by cDNA sequence FG737470.1). Therefore, loss of thymoproteasomes and immunoproteasomes seems to have taken place in an avian lineage. In mammals, medullary thymic epithelial cells involved in negative selection of T cells predominantly express immunoproteasomes (Nil et al. 2004); this ensures tolerance against self peptides generated by immunoproteasomes in the periphery. In cortical thymic epithelial cells, expression of the β5t subunit presumably blocks the formation of immunoproteasomes and allows for the production of a distinct repertoire of peptides that positively select T cells. Such differential display of peptides on the cortical and medullary thymic epithelial cells appears important in attaining efficient T cell selection (Hogquist and Xing 2010; Klein et al. 2009). Therefore, a jawed vertebrate ancestor that evolved immunoproteasomes presumably needed to have thymoproteasomes. The coexistence of the two forms of proteasomes in phylogeny supports this supposition (Fig. 1). In birds, loss of immunoproteasomes is likely to have decreased the needs to keep thymoproteasomes, resulting in the eventual loss of both forms of proteasomes. It would be interesting to examine whether thymic selection of CD8+ T cells in birds has any unusual features or birds have a mechanism that enables the production of distinct pools of MHC class I-binding peptides in the cortex and medulla independent of immunoproteasomes and thymoproteasomes.

Peptides bound to chicken MHC class I molecules show anomalous properties. First, peptides from the MHC class I molecule from the B4 haplotype generally have negatively charged amino acids at the C-terminus, a feature not found in mammals (Wallny et al. 2006). This has been suggested to be due to the lack of immunoproteasome components (Wallny et al. 2006). Second, chicken MHC class I molecules, BF2*2101 from the B21 haplotype, bind peptides more promiscuously than mammalian MHC class I molecules (Koch et al. 2007). This may also be related to the absence of immunoproteasomes and/or thymoproteasomes.

Recent evidence indicates that, besides antigen presentation, immunoproteasomes play important roles in the maintenance of protein homeostasis by preventing the accumulation of harmful protein aggregates during inflammation (Seifert et al. 2010). Thus, it would be interesting to examine whether the ability to resolve protein aggregates is altered in birds.

Teleost fish have two copies of *PSMB11* genes

We identified a single copy of the *PSMB11* gene in mammals, reptiles, amphibians, and cartilaginous fish. By contrast, two copies of *PSMB11* genes with no apparent defects were identified in all teleost species for which whole genome sequences were available, including zebrafish (*Danio rerio*), fugu (*Takifugu rubripes*), *Tetraodon nigroviridis*, medaka (*Oryzias latipes*), and stickleback (*Gasterosteus aculeatus*) (Fig. 1). We confirmed by cDNA cloning from thymic tissues that zebrafish express both copies of *PSMB11* genes (accession numbers: AB624351 and AB624352). One copy is located adjacent to the *PSMB5* gene, whereas the other copy is located on a different

chromosome or contig. We named the former and latter copies *PSMB11a* and *PSMB11b*, respectively (Fig. 2), and the encoded subunits were named β5ta and β5tb, respectively (Fig. 1). *PSMB11a* and *PSMB5* are located next to each other in all the teleost species subjected to whole genome sequencing. Before the discovery of the β5t subunit, a proteasome subunit gene located in the zebrafish MHC class I region was named *PSMB11* (Michalova et al. 2000; Murray et al. 1999). We suggest that the name *PSMB11* should be reserved for the gene coding for the β5t subunit and that the fish gene, which was originally described as *PSMB11* but is actually closely related to *PSMB9* coding for β1i (Clark et al. 2000; Hansen et al. 1999), should be renamed.

Several data and analyses indicate that the genome of a jawed vertebrate ancestor experienced two rounds of WGD (Flajnik and Kasahara 2010; Kasahara 2007; Ohno 1970; Putnam et al. 2008; Van de Peer et al. 2009), and that the genome of a teleost ancestor underwent an additional round of WGD (Meyer and Van de Peer 2005). Thus, the teleost genome often contains paralogous copies resulting from the fish-specific WGD (Postlethwait 2007). Comparison of the genomic regions containing *PSMB11a* and *PSMB11b* in *Tetraodon* indicates that, besides sharing the paralogs (paralogous copies) of the β5 family of proteasome genes, they share paralogs of at least three other gene families: *CDH24* (cadherin-like 24), *LRRC16B* (leucine rich repeat containing 16B), and *TGM1* (transglutaminase 1, K polypeptide) (Fig. 3). These observations suggest that *PSMB11a* and *PSMB11b* diverged by *en bloc* duplication, presumably through the fish-specific WGD. The extent of amino acid sequence identity between teleost β5ta and β5tb is ~50-60%. This is also consistent with the idea that they diverged by the fish-specific WGD. Notably, the teleost genes coding for β5 and β5i subunits do not have paralogs thought to have originated by the fish-specific WGD.

Thus, the *PSMB11*gene is unique in that it escaped diploidization.

Cleavage specificities of 20S proteasomes are determined by the nature of the amino acids that constitute the S1 pocket (Groll et al. 1997; Unno et al. 2002). In humans and mice, the S1 pockets of β5 and β5i are mainly made up of hydrophobic residues (Fig. 4). By contrast, the pocket of β5t is mainly composed of hydrophilic residues, accounting for its weaker chymotrypsin-like activity (Murata et al. 2007). This tendency is generally conserved in β5, β5i, and β5t subunits of non-mammalian vertebrates (Fig. 4a). In this regard, it is interesting to note that the S1 pockets of β5ta and β5tb show considerable differences in amino acid composition. The S1 pocket of β5tb is typical of β5t subunits in that it is made up of hydrophilic residues, and is quite similar to that of human β5t (Fig. 4b). However, like β5 and β5i, β5ta has A20 instead of S20 characteristic of β5t. Likewise, β5ta has G49 instead of S49 typically found in β5t. Thus, while retaining hydrophobic residues at positions 31 and 45, the S1 pocket of β5ta is less hydrophilic than that of β5tb, suggesting that β5ta and β5tb might differ in their ability to reduce the chymotrypsin-like activity of 20S proteasomes. Therefore, the duplicate *PSMB11* genes in teleost fish might have achieved some functional specialization.

Teleost β5ta and β5tb are also unique in that they have insertions of 20 (*Tetraodon* β5ta) to 65 (zebrafish β5ta) amino acid residues, the sequences of which are poorly conserved across different teleost species (Fig. 1S). These insertions appear to be located between the regions presumed to form β-strands S7 and S8 (Fig. 1S). Thus, they presumably do not affect the structure of the cavity inside the 20S proteasomes.

Linkage of *PSMB11* and *PSMB5* genes in vertebrates

As in mammals, the genes coding for β5t and β5 are located next to each other in *Anolis* lizards and *Xenopus tropicalis* (Fig. 2). Sharks have the *PSMB5* gene as evidenced by the availablity of corresponding cDNA sequences (Kandil et al. 1996); however, information is not available on its linkage to the *PSMB11* gene. Nevertheless, the observation that *PSMB5* and *PSMB11* are located in tandem in all bony vertebrates suggests strongly that, unlike the genes coding for the β1i, β2i, and β5i subunits that are thought to have emerged by WGD (Flajnik and Kasahara 2010; Kasahara et al. 1996; Kasahara et al. 1997), *PSMB11* arose by tandem duplication from the evolutionarily more ancient *PSMB5* gene ubiquitously distributed in eukaryotes. It is notable that the relative orientation of the two genes differs between bony fish and other classes of vertebrates; *PSMB11a* and *PSMB5* are located in a tail-to-head orientation in bony fish, whereas *PSMB11* and *PSMB5* occur in a head-to-head orientation in mammals, *Anolis* lizards, and amphibians. Simple tandem duplication would create two genes oriented in the same direction. Thus, it is likely that *PSMB11* underwent an inversion relative to *PSMB5* in a tetrapod ancestor.

PSMB11 lacks introns in vertebrates ranging from the cartilaginous fish to mammals

A unique property of the mammalian *PSMB11* gene is the absence of introns. Comparative analysis indicates that *PSMB11* basically lacks introns in all classes of jawed vertebrates (Fig. 2). The only exception is the *PSMB11b* gene of zebrafish; it has a single intron of 2,604 bps. Because all the other bony fish species examined have intronless *PSMB11a* and *PSMB11b* genes, zebrafish *PSMB11b* appears to have acquired

an intron secondarily. Also unique is the structure of mammalian *PSMB5* genes. Most β-type subunit genes including those coding for β1, β1i, β2, β2i, and β5i subunits are made up of six to eight exons, with similar exon-intron organization (Hayashi et al. 1997; Tanaka and Kasahara 1998). However, the mammalian *PSMB5* gene is made up of three exons, and its organization differs radically from that of typical β-type proteasome subunit genes (Abdulla et al. 1996; Kohda et al. 1997). Although information is not available in the cartilaginous fish, the *PSMB5* gene of teleost fish, amphibians, and reptiles all share the same three-exon structure (Fig. 2). Interestingly, in lampreys and amphioxus, the *PSMB5* gene has a six-exon structure typical of β-type proteasome subunit genes (the organization is slightly modified in *Ciona intestinalis*, resulting in a five-exon structure). None of the exon-intron boundaries is shared between jawed-vertebrate-type and jawless-vertebrate/invertebrate-type *PSMB5* genes. We suggest, therefore, that *PSMB5* probably lost all of its introns in a jawed vertebrate ancestor after its separation from a jawless vertebrate ancestor, and that this intronless *PSMB5* gene duplicated in tandem in a jawed vertebrate ancestor. After this duplication, *PSMB11* has remained intronless, but *PSMB5* acquired two introns at the latest by the emergence of bony fish (Fig. 5). An alternative possibility is that the *PSMB5* gene retained the jawless-vertebrate/invertebrate-type organization when it duplicated in tandem; however, this scenario requires independent reorganization of two genes: major exon-intron reorganization in the *PSMB5* gene and a complete loss of introns in the *PSMB11* gene. Hence, this possibility appears less likely. The *PSMB5* gene is located in the region thought to be paralogous to the MHC region encoding the β5i subunit (Flajnik and Kasahara 2010; Kasahara 2007; Kasahara 2010), suggesting that it has remained in its original position. Thus, *PSMB5* might have lost its introns in a jawed

vertebrate ancestor by homologous recombination between the genomic copy of the gene and a reverse transcriptase product of a spliced mRNA (Derr 1998; Fink 1987; Kaessmann et al. 2009; Mourier and Jeffares 2003).

In summary, our analysis suggests that the gene coding for β5t, a unique component of the thymoproteasome, emerged in a common ancestor of jawed vertebrates by tandem duplication from the *PSMB5* gene. Thus, similar to other key components of MHC- and T/B-cell-receptor-based adaptive immunity, thymoproteasomes appear unique to jawed vertebrates. Remarkably, birds appear to have only standard 20S proteasomes; thus, they may be regarded as natural knock-out animals deficient in both immunoproteasomes and thymoproteasomes, offering an unique opportunity to study the functions of immunoproteasomes and thymoproteasomes *in vivo*. Our results also show that teleost fish have two copies of *PSMB11* genes that are apparently functional. It would be interesting to examine whether they are both expressed in cortical thymic epithelial cells, and whether the existence of two copies of *PSMB11* genes offers any advantage to the positive selection of T cells in the teleost fish.

Acknowledgments This work was supported by grants from The Ministry of Education, Culture, Sports, Science and Technology of Japan, and National Institutes of Health grant AI027877. Yoichi Sutoh and Mizuho Kondo are supported by the Research Fellowship of the Japan Society for the Promotion of Science.

References

Abdulla S, Beck S, Belich M, Jackson A, Nakamura T, Trowsdale J (1996) Divergent intron arrangement in the *MB1/LMP7* proteasome gene pair. Immunogenetics 44*:*254-258

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22*:*195-201

Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, McCurley N, Bockman DE, Schorpp M, Cooper MD, Boehm T (2011) A thymus candidate in lampreys. Nature 470*:*90-94

Balakrishnan CN, Ekblom R, Volker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC, Edwards SV (2010) Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol 8*:*29

Belich MP, Trowsdale J (1995) Proteasome and class I antigen processing and presentation. Mol Biol Rep 21*:*53-56

Chaves LD, Krueth SB, Reed KM (2009) Defining the turkey MHC: sequence and genes of the *B* locus. J Immunol 183*:*6530-6537

Clark MS, Pontarotti P, Gilles A, Kelly A, Elgar G (2000) Identification and characterization of a β proteasome subunit cluster in the Japanese pufferfish (*Fugu rubripes*). J Immunol 165*:*4446-4452

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le A, Bouffard P, Burt DW, Crasta O, Crooijmans RP, Cooper K, Coulombe RA, De S, Delany ME,

Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MA, Harkins TT, Herrero J, Hoffmann S, Megens HJ, Jiang A, de Jong P, Kaiser P, Kim H, Kim KW, Kim S, Langenberger D, Lee MK, Lee T, Mane S, Marcais G, Marz M, McElroy AP, Modise T, Nefedov M, Notredame C, Paton IR, Payne WS, Pertea G, Prickett D, Puiu D, Qioa D, Raineri E, Ruffier M, Salzberg SL, Schatz MC, Scheuring C, Schmidt CJ, Schroeder S, Searle SM, Smith EJ, Smith J, Sonstegard TS, Stadler PF, Tafer H, Tu ZJ, Van Tassell CP, Vilella AJ, Williams KP, Yorke JA, Zhang L, Zhang HB, Zhang X, Zhang Y, Reed KM (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol 8*:*e1000475

Danchin E, Vitiello V, Vienne A, Richard O, Gouret P, McDermott MF, Pontarotti P (2004) The major histocompatibility complex origin. Immunol Rev 198*:*216-232

Derr LK (1998) The involvement of cellular recombination and repair genes in RNA-mediated recombination in Saccharomyces cerevisiae. Genetics 148*:*937-945

Fink GR (1987) Pseudogenes in yeast? Cell 49*:*5-6

Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: Glimpses into the evolution of the adaptive immune system. Immunity 15*:*351-362

Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11*:*47-59

Goldberg AL (1995) Functions of the proteasome: The lysis at the end of the tunnel. Science 268*:*522-523

Griffin TA, Nandi D, Cruz M, Fehling HJ, Kaer LV, Monaco JJ, Colbert RA (1998)

Immunoproteasome assembly: cooperative incorporation of interferon γ (IFN-γ)-inducible subunits. J Exp Med 187:97-104

Groettrup M, Kirk CJ, Basler M (2010) Proteasomes in immune cells: more than peptide producers? Nat Rev Immunol 10:73-78

Groll M, Ditzel L, Loewe J, Stock D, Bochtler M, Bartunik HD, Huber R (1997) Structure of 20S proteasome from yeast at 2.4 Å resolution. Nature 386:463-471

Hansen JD, Strassburger P, Thorgaard GH, Young WP, Du Pasquier L (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J Immunol 163:774-786

Hayashi M, Ishibashi T, Tanaka K, Kasahara M (1997) The mouse genes encoding the third pair of β-type proteasome subunits regulated reciprocally by IFN-γ: structural comparison, chromosomal localization, and analysis of the promoter. J Immunol 159:2760-2770

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobbson L, Kerje S, Andersson L, Crooijmans RP, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown

WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SM, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyras E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695-716

Hogquist KA, Xing Y (2010) Why CD8+ T cells need diversity when growing up. Immunity 32:5-6

Kaessmann H, Vinckenbosch N, Long M (2009) RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet 10:19-31

Kandil E, Namikawa C, Nonaka M, Greenberg AS, Flajnik MF, Ishibashi T, Kasahara M (1996) Isolation of low molecular mass polypeptide complementary DNA

clones from primitive vertebrates: Implications for the origin of MHC class I-restricted antigen presentation. J Immunol 156:4245-4253

Kasahara M (2007) The 2R hypothesis: an update. Curr Opin Immunol 19:547-552

Kasahara M (2010) Genome duplication and T cell immunity. Prog Mol Biol Transl Sci 92:7-36

Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T (1996) Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci USA 93:9096-9101

Kasahara M, Nakaya J, Satta Y, Takahata N (1997) Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet 13:90-92

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 30:3059-3066

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken *B* locus is a minimal essential major histocompatibility complex. Nature 401:923-925

Klein L, Hinterberger M, Wirnsberger G, Kyewski B (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. Nat Rev Immunol 9:833-844

Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny HJ, van Hateren A, Hunt L, Jacob JP, Johnston F, Marston DA, Shaw I, Dunbar PR, Cerundolo V, Jones EY, Kaufman J (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity 27:885-899

Kohda K, Matsuda Y, Ishibashi T, Tanaka K, Kasahara M (1997) Structural analysis and chromosomal localization of the *Psmb5* gene coding for the constitutively expressed β-type proteasome subunit. Immunogenetics 47*:*77-87

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9*:*299-306

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

Michalova V, Murray BW, Sultmann H, Klein J (2000) A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny. J Immunol 164*:*5296-5305

Mourier T, Jeffares DC (2003) Eukaryotic intron loss. Science 300*:*1393

Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, Tanaka K (2007) Regulation of CD8+ T cell development by thymus-specific proteasomes. Science 316*:*1349-1353

Murata S, Takahama Y, Tanaka K (2008) Thymoproteasome: probable role in generating positively selecting peptides. Curr Opin Immunol 20*:*192-196

Murray BW, Sultmann H, Klein J (1999) Analysis of a 26-kb region linked to the Mhc in zebrafish: Genomic organization of the proteasome component β/tansporter associated with antigen processing-2 gene cluster and identification of five new proteasome β subunit genes. J Immunol 163*:*2657-2666

Nil A, Firat E, Sobek V, Eichmann K, Niedermann G (2004) Expression of housekeeping and immunoproteasome subunit genes is differentially regulated in positively and negatively selecting thymic stroma subsets. Eur J Immunol

34:2681-2689

Nitta T, Murata S, Sasaki K, Fujii H, Ripen AM, Ishimaru N, Koyasu S, Tanaka K, Takahama Y (2010) Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. Immunity 32:29-40

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, New York

Pamer E, Cresswell P (1998) Mechanisms of MHC class I-restricted antigen processing. Annu Rev Immunol 16:323-358

Postlethwait JH (2007) The zebrafish genome in context: ohnologs gone missing. J Exp Zool B Mol Dev Evol 308:563-577

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064-1071

Seifert U, Bialy LP, Ebstein F, Bech-Otschir D, Voigt A, Schroter F, Prozorovski T, Lange N, Steffen J, Rieger M, Kuckelkorn U, Aktas O, Kloetzel PM, Kruger E (2010) Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. Cell 142:613-624

Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, Hanzawa K, Beck S, Kulski JK, Inoko H (2004) Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol 172:6751-6763

Takahama Y, Nitta T, Ripen AM, Nitta S, Murata S, Tanaka K (2010) Role of thymic cortex-specific self-peptides in positive selection of T cells. Semin Immunol 22:287-293

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol:10.1093/molbev/msr1121

Tanaka K (2009) The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci 85:12-36

Tanaka K, Kasahara M (1998) The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-γ-inducible proteasome activator PA28. Immunol Rev 163:161-176

Tomaru U, Ishizu A, Murata S, Miyatake Y, Suzuki S, Takahashi S, Kazamaki T, Ohara J, Baba T, Iwasaki S, Fugo K, Otsuka N, Tanaka K, Kasahara M (2009) Exclusive expression of proteasome subunit β5t in the human thymic cortex. Blood 113:5186-5191

Unno M, Mizushima T, Morimoto Y, Tomisugi Y, Tanaka K, Yasuoka N, Tsukihara T (2002) The structure of the mammalian 20S proteasome at 2.75 Å resolution. Structure 10:609-618

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10:725-732

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*)

genome. PLoS Biol 5:e101

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjodt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci USA 103:1434-1439

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TA, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, Nam K, Backstrom N, Smeds L, Nabholz B, Itoh Y, Whitney O, Pfenning AR, Howard J, Volker M, Skinner BM, Griffin DK, Ye L, McLaren WM, Flicek P, Quesada V, Velasco G, Lopez-Otin C, Puente XS, Olender T, Lancet D, Smit AF, Hubley R, Konkel MK, Walker JA, Batzer MA, Gu W, Pollock DD, Chen L, Cheng Z, Eichler EE, Stapley J, Slate J, Ekblom R, Birkhead T, Burke T, Burt D, Scharff C, Adam I, Richard H, Sultan M, Soldatov A, Lehrach H, Edwards SV, Yang SP, Li X, Graves T, Fulton L, Nelson J, Chinwalla A, Hou S, Mardis ER, Wilson RK (2010) The genome of a songbird. Nature 464:757-762

Figure legends

Fig. 1 Phylogenetic tree of the β5 family of proteasome subunits. Nodes supported by bootstrap values over 90% are indicated by open circles. DDBJ/EMBL/NCBI accession numbers or Ensembl gene identification numbers are as follows: human β5i, P28062; mouse β5i, NP_034854.2; *Anolis* lizard β5i, ENSACAT00000006461; *X. tropicalis* β5i, NP_001103512.1; zebrafish β5i, AAB87679.1; nurse shark β5i, BAA10934.1; human β5, NP_002788.1; mouse β5, BAA24916.1; chicken β5, BAA19471.1; *Anolis* lizard β5, FG667020.1; *X. tropicalis* β5, XP_002941560.1; zebrafish β5, AAI53590.1; nurse shark β5, BAA10935.1; lamprey β5, BAA10932.1; hagfish β5, BAA10931.1; *C. intestinalis* β5, XP_002132107.1; Florida lancelet β5, AAM18885.1; human β5t, BAF63540.1; mouse β5t, BAF63539.1; *Anolis* lizard β5t, ENSACAT00000001900; *X. tropicalis* β5t, XP_002941563.1; zebrafish β5ta, AB624351; zebrafish β5tb, AB624352; fugu β5ta, ENSTRUT00000004609; fugu β5tb, ENSTRUT00000034202; stickleback β5tb, ENSGACT00000002260; *Tetraodon* β5ta, ENSTNIT00000001930; *Tetraodon* β5tb, ENSTNIT00000009972; medaka β5tb, ENSORLT00000020386; and elephant shark β5t, AAVX01606762.1. Medaka and stickleback β5ta genes were not annotated in the databases.

Fig. 2 Genomic organization of *PSMB11* and *PSMB5* genes. Teleost fish have two *PSMB11* genes located on different chromosomes or contigs. Only partial sequence data are available for the medaka *PSMB5* gene. No avian *PSMB11* sequence was identified in the NCBI/Ensembl databases. The Ensembl database contains only a partial genomic sequence of chicken *PSMB5*. Hence, its structure is not shown. Note that lamprey,

Ciona, and Florida lancelet *PSMB5* genes are drawn in scales different from those used for the remaining genes.

Fig. 3 The genomic regions containing *PSMB11a* and *PSMB11b* share paralogs of at least three other gene families. The *Tetraodon* genome contains two sets of closely related *CDH24*, *LRRC16B*, and *TGM1* genes. One set, designated with suffix *a*, is located immediately downstream from the *PSMB11a/PSMB5* gene cluster on choromosome 15. Another set of *CDH24*, *LRRC16B*, and *TGM1* genes, designated with suffix *b*, is distributed between 11.20 and 12.09 Mbp on chromosome 1. *PSMB11b* is located at ~11.29 Mbp on choromosome 1. The information is based on the TETRAODON 8.0 assembly of the Ensembl database. Numbers starting with ENSTNIG are Ensembl gene identification numbers.

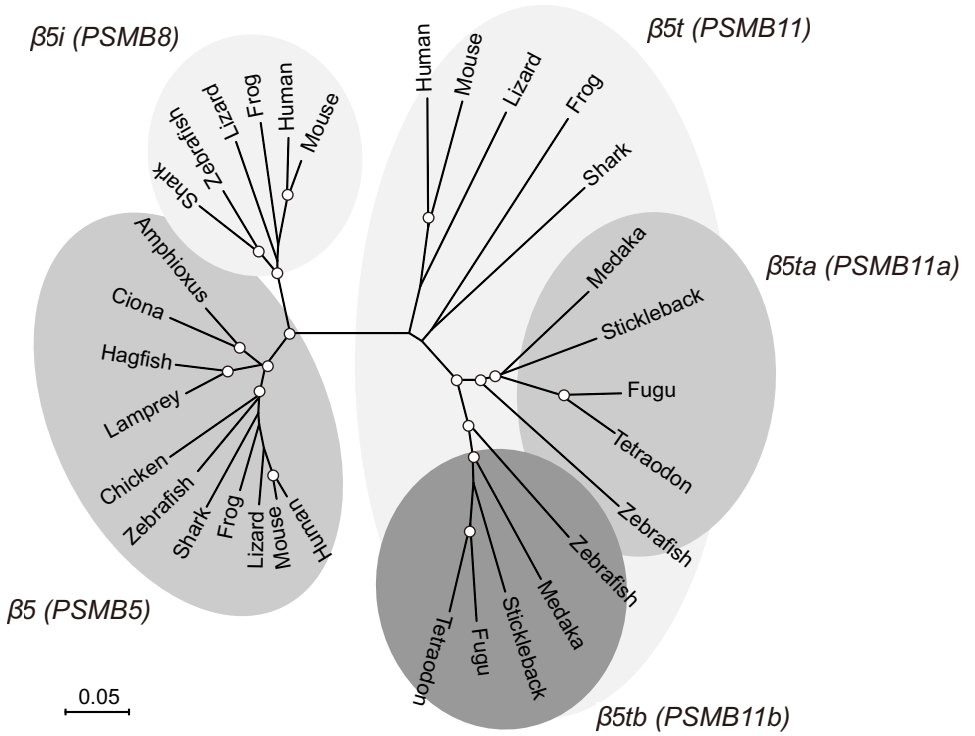
Fig. 4 Amino acid sequences forming the S1 pockets of β5i/β5/β5t subunits (**a**) and the predicted three-dimensional structures of the S1 pockets (**b**). In **a**, hydrophobic and hybrophilic residues are boxed in black and white, respectively. In **b**, hydrophobic and hybrophilic residues are indicated in red and blue, respectively. Amino acids are shown in a standard single-letter code.

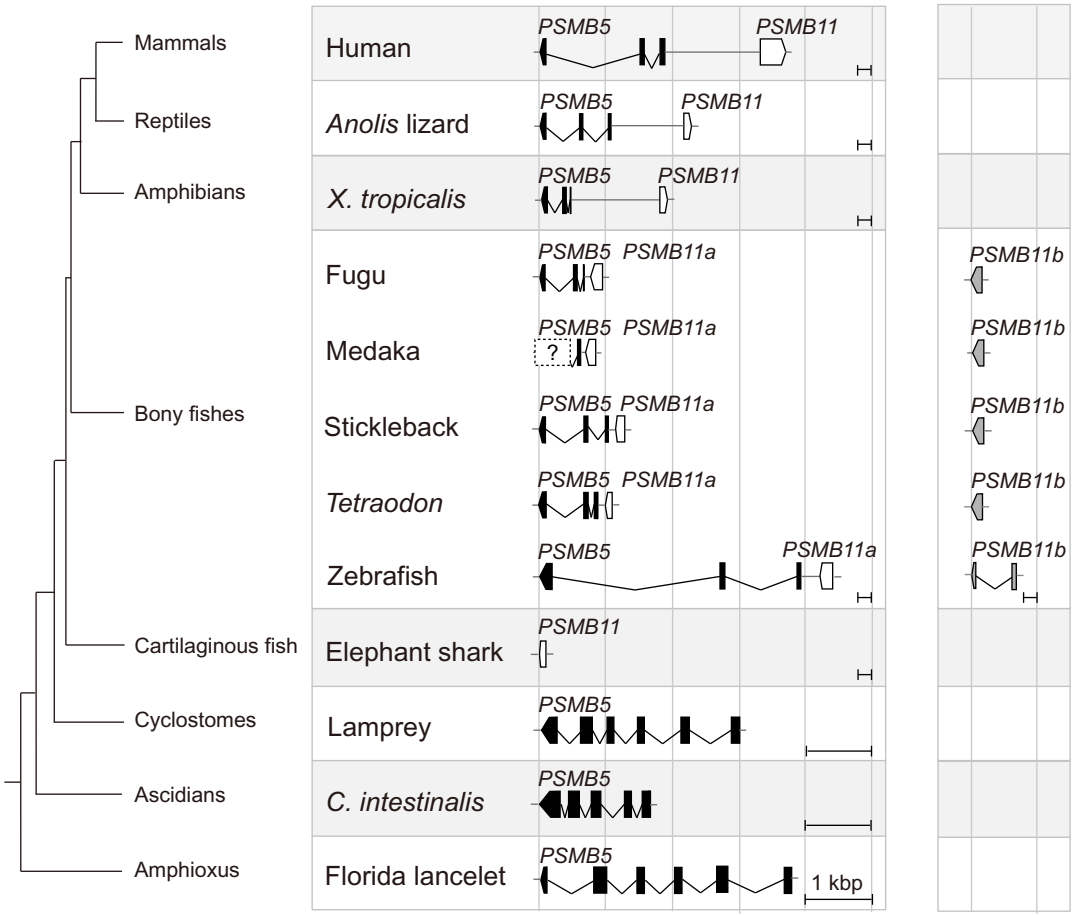
Fig. 5 Evolution of β5 and β5t subunit genes. Exons and introns are indicated by filled boxes and lines, respectively. *Exon-intron organization of chicken *PSMB5* is not known.

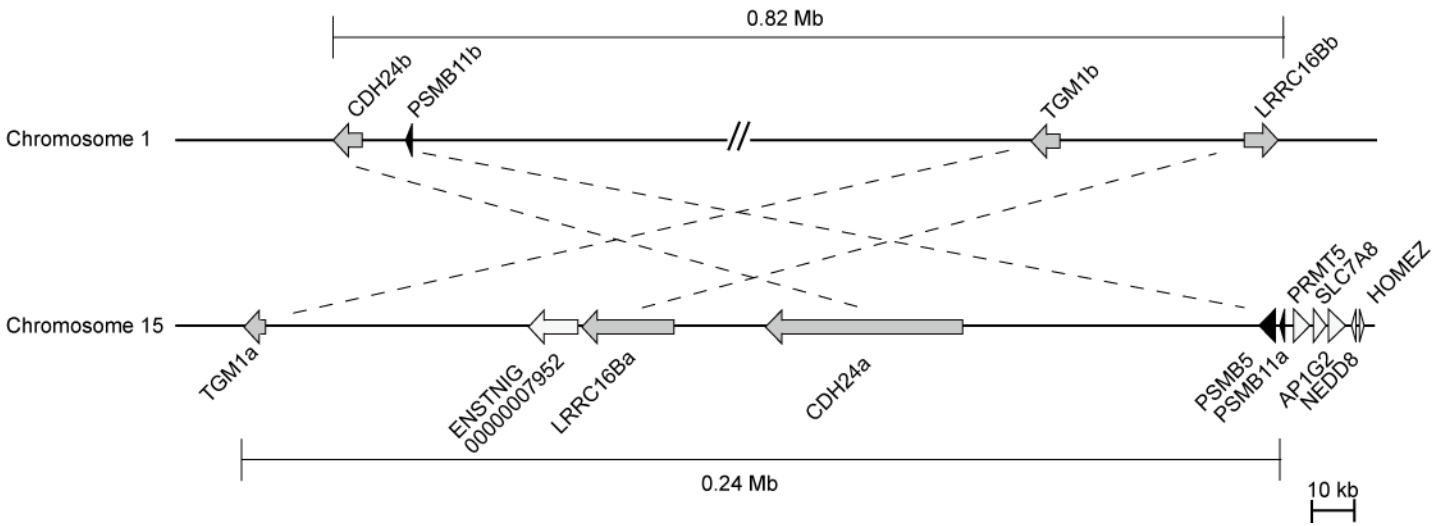
Fig. 1S Amino acid sequence alignment of β5i, β5, and β5t subunits from various

species. H1-H5 and β1-β12 stand for helices and β-strands, respectively. The locations of secondary structures are based on the information deposited in the RCSB Protein Data Bank. This alignment was used for the construction of the tree shown in Fig. 1.

Fig. 2S *Anolis* Ensembl scaffold GL344036.1 encodes MHC class I, *TAP1*, *TAP2*, and *PSMB8*. The MHC class I gene shown here appears to be a classical class I gene as its predicted α1 and α2 domains have eight conserved residues that are involved in the anchoring of peptide termini. The draft genome sequence contains gaps upsteam of the *PSMB8* gene (indicated by a broken line) and lacks the presumed exon 1 sequence of *PSMB8*.


β5i (PSMB8)
Mouse
Human
Frog
Lizard
Zebrafish
Shark
β5t (PSMB11)
Human
Mouse
Lizard
Frog
Shark
β5ta (PSMB11a)
Medaka
Stickleback
Fugu
Tetraodon
Zebrafish
β5tb (PSMB11b)
Zebrafish
Medaka
Stickleback
Fugu
Tetraodon
β5 (PSMB5)
Amphioxus
Ciona
Hagfish
Lamprey
Chicken
Zebrafish
Shark
Frog
Lizard
Mouse
Human
0.05

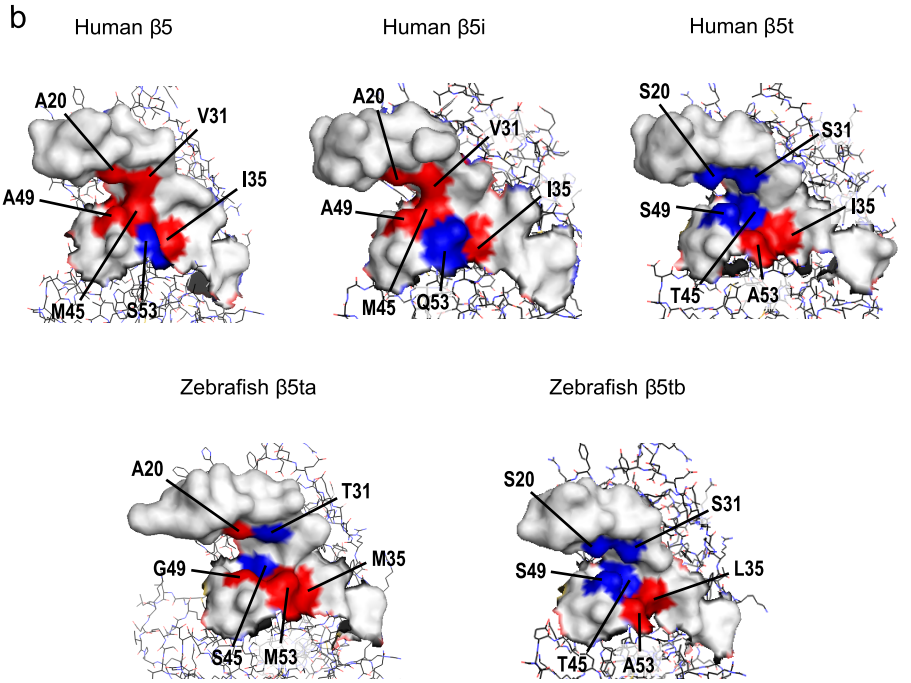

Mammals
Reptiles
Amphibians
Bony fishes
Cartilaginous fish
Cyclostomes
Ascidians
Amphioxus
Human
Anolis lizard
X. tropicalis
Fugu
Medaka
Stickleback
Tetraodon
Zebrafish
Elephant shark
Lamprey
C. intestinalis
Florida lancelet
PSMB5
PSMB11
PSMB11a
PSMB11b
?
1 kbp

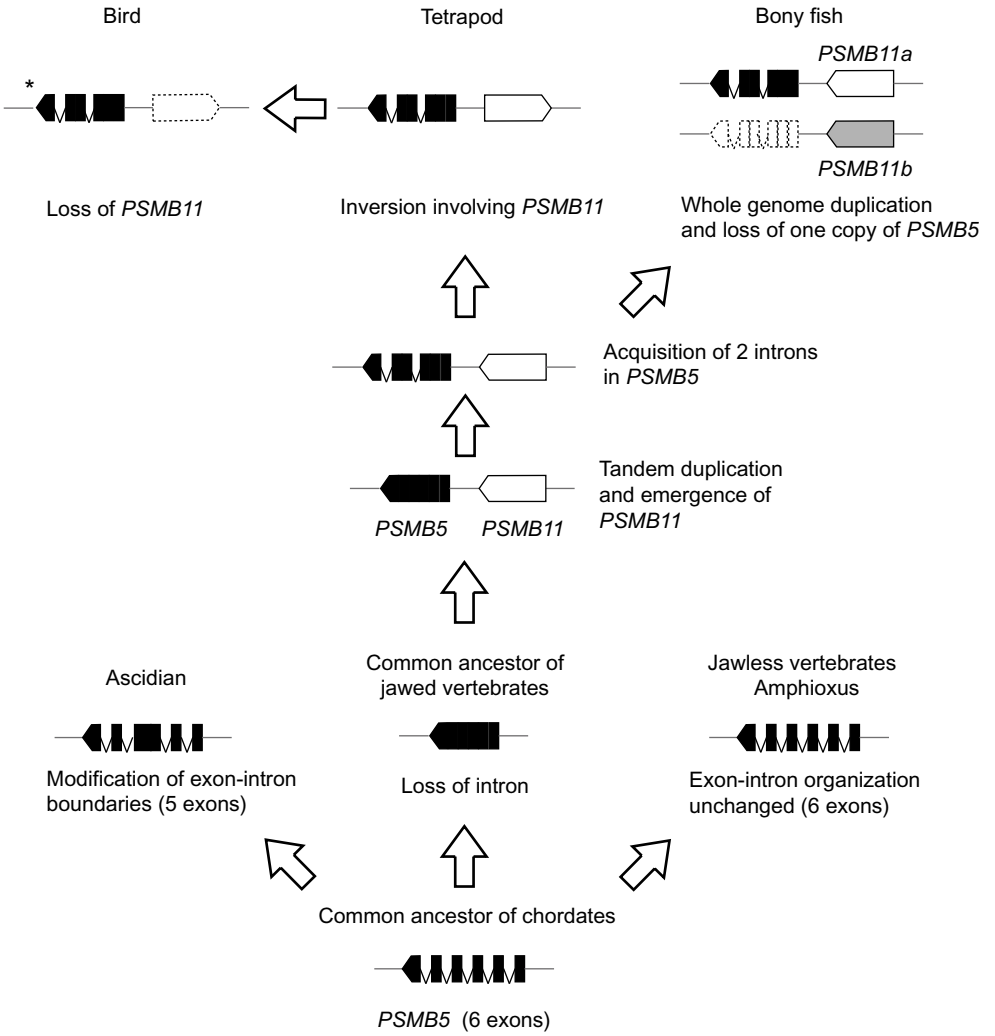

0.82 Mb
CDH24b
PSMB11b
TGM1b
LRRC16Bb
Chromosome 1
Chromosome 15
PRMT5
SLC7A8
HOMEZ
TGM1a
ENSTNIG
00000007952
LRRC16Ba
CDH24a
PSMB5
PSMB11a
AP1G2
NEDD8
0.24 Mb
10 kb

a

	20	31	35	45	49	53
Human β5i	A	V	I	M	A	Q
Mouse β5i	A	M	I	M	A	Q
Lizard β5i	A	C	I	M	A	Q
Frog β5i	A	F	I	M	A	Q
Zebrafish β5i	A	A	I	M	A	Q
Shark β5i	A	A	I	M	A	Q
Human β5	A	V	I	M	A	S
Mouse β5	A	V	I	M	A	S
Chicken β5	A	V	I	M	A	S
Lizard β5	A	V	I	M	A	S
Frog β5	A	V	I	M	A	S
Zebrafish β5	A	V	I	M	A	S
Shark β5	A	V	I	M	A	S
Lamprey β5	A	V	I	M	A	M
Hagfish β5	A	V	I	M	A	M
Ciona β5	A	V	I	M	A	S
Amphioxus β5	A	V	I	M	A	S
Human β5t	S	S	I	T	S	A
Mouse β5t	S	S	I	T	S	A
Lizard β5t	S	S	I	T	S	A
Frog β5t	S	S	T	T	S	Q
Fugu β5ta	A	T	S	C	G	M
Stickleback β5ta	A	V	T	S	G	M
Medaka β5ta	A	V	T	S	G	M
Tetraodon β5ta	A	T	C	C	G	M
Zebrafish β5ta	A	T	M	T	G	M
Fugu β5tb	S	S	V	T	S	V
Stickleback β5tb	S	S	R	T	S	A
Medaka β5tb	S	S	L	T	S	A
Tetraodon β5tb	S	S	V	T	S	V
Zebrafish β5tb	S	S	L	T	S	A
Shark β5t	S	S	V	T	A	A

b




Bird
Tetrapod
Bony fish
*
PSMB11a
PSMB11b
Loss of PSMB11
Inversion involving PSMB11
Whole genome duplication
and loss of one copy of PSMB5
Acquisition of 2 introns
in PSMB5
Tandem duplication
and emergence of
PSMB11
PSMB5
PSMB11
Common ancestor of
jawed vertebrates
Ascidian
Jawless vertebrates
Amphioxus
Modification of exon-intron
boundaries (5 exons)
Loss of intron
Exon-intron organization
unchanged (6 exons)
Common ancestor of chordates
PSMB5 (6 exons)