学位論文

Evolution of the dimorphism of the immunoproteasome subunit beta type 8 (*PSMB8*) gene in reptiles

（爬虫類における免疫プロテアソームサブユニット *PSMB8* 遺伝子の二型性の進化）

平成24年12月博士（理学）　申請

東京大学大学院理学系研究科
生物科学専攻

黄　慶輝

Contents

Abstract

The proteasome subunit beta type 8 gene (*PSMB8*) encodes one of the beta subunits of the immunoproteasome responsible for the generation of peptides presented by major histocompatibility complex (MHC) class I molecules. Dimorphic alleles of the *PSMB8* gene, termed A and F types, based on the deduced 31st amino acid residue of the mature protein have been reported from various vertebrates. Phylogenetic analysis indicated the presence of two ancient lineages, one comprising the F type *PSMB8* of sharks and basal actinopterygians such as Polypteriformes, Cypriniformes and Salmoniformes, and the other comprising the A type *PSMB8* of these animals. These lineages were termed the ***PSMB8F*** lineage and the ***PSMB8A*** lineage, respectively. Interestingly, both the F and A type *PSMB8* of acanthopterygians and *Xenopus* were included in the ***PSMB8A*** lineage, indicating that evolution of the *PSMB8* dimorphism was not straightforward. I analyzed the *PSMB8* gene of five reptiles: gecko, alligator, crocodile, turtle, and soft shell turtle, and found both the A and F types from all five. Phylogenetic analysis indicated that the ***PSMB8F*** lineage was lost in a common ancestor of tetrapod, and the apparent functional equivalents of the ***PSMB8F*** lineage members with Phe or Tyr at the 31st position of the mature protein was regenerated within the ***PSMB8A*** lineage. The A and F type *PSMB8* sequences of newt, *Xenopus*, gecko, Crocodilia (alligator and crocodile), and Testudines (turtle and soft shell turtle) formed taxon-specific monophyletic clades, indicating that recurrent regeneration of *PSMB8* F type occurred at least five times independently during tetrapod evolution. However, once established, these apparently

independent dimorphisms seem to have been evolutionarily stable and some of them show long-term trans-species dimorphism. Genomic typing of wild individuals of geckos indicated that the frequencies of the A and F type alleles are not highly biased. Phylogenetic analysis of each exon of the reptile *PSMB8* gene suggested interallelic sequence homogenization as a possible evolutionary mechanism for the apparent recurrent regeneration of *PSMB8* dimorphism in tetrapods. An extremely strong balancing selection acting on *PSMB8* dimorphism was implicated in an unprecedented pattern of allele evolution.

Introduction

Proteasomes are the multi-subunit protease responsible for the generation of peptides presented by major histocompatibility complex (MHC) class I molecules (Tanaka and Kasahara 1998). The 20S proteasome, a catalytic core of the larger 26S proteasome, is a large complex composed of four stacks of two outer α-rings and two inner β-rings containing seven α and seven β subunits, respectively (Groll et al. 1997; Unno et al. 2002). In standard 20S proteasomes, proteolytic activity depends on three β subunits, PSMB5, PSMB6, and PSMB7, which have chymotrypsin-like, caspase-like, and trypsin-like specificity, respectively (Tanaka and Kasahara 1998). The 20S immunoproteasome, an altered form of proteasome, is formed by replacing the PSMB5, PSMB6, and PSMB7 of the 20S proteasome with the IFN-γ-inducible β subunits, PSMB8, PSMB9, and PSMB10, respectively (Fig.1) (Akiyama et al. 1994; Früh et al. 1994). These subunit substitutions change the cleavage specificity of the 20S proteasome, leading to production of peptides more suitable for binding to MHC class I molecules (Tanaka and Kasahara 1998). The antigen processing pathway that generates the antigenic peptide for MHC class I begins in the cytoplasm (Fig. 2). The immunoproteasome fragments proteins into smaller polypeptides. These proteolytic intermediates are then transported into the endoplasmic reticulum (ER) by TAP (heterodimer of ABCB2 and ABCB3), the transporter associated with antigen processing. The pool of proteolytic intermediates entering the ER possess the proper C-terminal anchor residue required to bind MHC class I molecules but contain extra N-terminal residues. The extra N-terminal residues are trimmed in the ER by ERAAP (ER aminopeptidase associated with antigen processing) to generate the final products for binding the MHC class I (Rock and Goldberg 1999; Blanchard and Shastri 2008).

These antigenic peptides are loaded onto the MHC class I molecules with the help of the TAP binding protein associated with antigen presentation (TAPBP). The MHC-peptide complex is then transported to the cell surface, and is recognized by the TCR of $CD8^+$ T cells (killer T cell). PSMB8 subunit is essential for the production of the MHC class I binding peptides, and PSMB8-deficient mice show reduced cell surface expression of MHC class I molecules and increased susceptibility to pathogens (Fehling et al. 1994). In contrast, two missense mutations of the human *PSMB8* gene, one affects only the chymotrypsin-like activity and the other affects all three peptidase activities, are reported to cause autoinflammation and lipodystrophy in homozygotic carriers without increasing their susceptibility to pathogens (Agarwal et al. 2010; Arima et al. 2011; Kitamura et al. 2011). These reports suggest that PSMB8 plays some roles outside of the class I antigen processing pathway and that the full physiological function of PSMB8 is still to be understood.

The proteasome subunit beta type 8 (*PSMB8*) gene in the MHC region is believed to have arisen by gene duplication of *PSMB5* in a common ancestor of jawed vertebrates, simultaneously with the appearance of MHC and the canonical adaptive immunity (Kasahara 1997). Highly diverged dichotomous alleles of *PSMB8* showing only ~80 % amino acid sequence identity have been reported in several bony fish and amphibian *Xenopus* species (Nonaka et al. 2000; Miura et al; 2010, Fujito and Nonaka 2012; Tsukamoto et al. 2012). These dimorphic alleles share a similar substitution at the 31^{st} position of the mature protein, and are termed as the F and A types, possessing Phe or Tyr and Ala or Val, respectively at this position. These amino acid substitutions most likely affect the size of the inner space of the S1 pocket, resulting in a difference in cleavage specificity (Fig. 3) (Unno et al. 2002; Tsukamoto et al. 2012). Phylogenetic analysis indicated the presence of two ancient lineages, one comprising the F type

PSMB8 of sharks and basal actinopterygians such as Polypteriformes, Cypriniformes and Salmoniformes, and the other comprising the A type *PSMB8* of these animals. These lineages were termed the ***PSMB8F*** lineage and the ***PSMB8A*** lineage, respectively. Interestingly, both the F and A type *PSMB8* of acanthopterygians and *Xenopus* were included in the ***PSMB8A*** lineage. The A and F types of *PSMB8* of bichir and zebrafish segregated as alleles, providing an extreme example of long-term trans-species polymorphism (Fujito and Nonaka 2012; Tsukamoto et al. 2012). Because the bichir and zebrafish diverged about 400 million years (My) ago, an exceptionally strong and long-lasting balancing selection is postulated for the *PSMB8* dimorphism of basal actinopterygians. However, the ***PSMB8F*** lineage was lost in tetrapods and acanthopterygians and apparent revival of F type *PSMB8* within the ***PSMB8A*** lineage has occurred at least twice independently in teleost *Oryzias* and amphibian *Xenopus* species (Nonaka et al. 2000; Miura et al. 2010). Interestingly, the origins of the PSMB8 dimorphism of *Oryzias* and *Xenopus* species were independent of each other despite the similar substitution at the 31st position of the mature protein (Tsukamoto et al. 2005). The apparently revived *PSMB8* dimorphisms of *Oryzias* and *Xenopus* were also considered to be under a strong balancing selection, and showed long-term trans-species polymorphism lasting for up to 100 My (Nonaka et al. 2000; Miura et al. 2010). The allelic status of the A and F types of *PSMB8* was proven in *Oryzias* species not only by segregation analysis but also by physical analysis of the genomic regions harboring the A or F types of the *PSMB8* gene (Tsukamoto et al. 2005). In contrast, the F and A types are reported to be pseudoalleles in sharks (Ohta et al. 2002). On the other hand, all mammalian *PSMB8* sequences thus far reported from various species including human are A type, suggesting that the F type *PSMB8* was lost in a common ancestor of mammals. Moreover, the *PSMB8* gene itself seems to have

been lost in the common ancestor of birds, because it cannot be found in the chicken, turkey or zebra finch genomes (Hillier et al. 2004; Dalloul et al. 2010; Warren et al. 2010). In reptile, only the PSMB8 with C^{31} of green anole lizard is recorded in the genome database. To clarify the evolution of the *PSMB8* dimorphism in tetrapods, I searched for dimorphic alleles of *PSMB8* in several reptilian species.

Materials and Methods

Samples

Wild individuals (N=43) of the Japanese gecko (*Gekko japonicus*) were sampled in Yanaka Cemetery, Tokyo, Japan. Liver, spleen and muscle samples of the spectacled caiman (*Caiman crocodilus*) (N=4) and saltwater crocodile (N=1) (*Crocodylus porosus*) were purchased from a crocodile farm in Pingdong, Taiwan. A blood sample was collected from a red-eared slider (*Trachemys scripta*) (N=5) caught at Sanshiro Pond at the University of Tokyo, Tokyo, Japan. Chinese soft-shell turtles (*Pelodiscus sinensis*) (N=5) were purchased from a soft-shell turtle farm in Kyushu, Japan.

Additional samples of amphibian, bird, and lungfish species were also applied in this study: A wild individual of Tago's brown frog (*Rana tagoi*) was sampled in mountain area of Yamanashi, Japan. A cultured African lungfish (*Protopterus dolloi*) was purchased from aquarium shop and liver sample from a cultured Australian lungfish (*Neoceratodus forsteri*) was kindly provided by Dr. M. Takechi. An embryo of African penguin (*Spheniscus demersus*) was provided by Ueno Zoo, Tokyo, Japan. A spleen sample of common ostrich (*Struthio camelus*) was purchased from an ostrich farm in Ibaraki, Japan.

The phylogenic relationship of these species is described in Fig. 4 together with divergence times of these taxon based on Shen et al. 2011 and Slack et al. 2006.

RNA extraction

Total RNA was isolated from the liver, spleen, muscle or blood by using ISOGEN (NIPPON GENE, Tokyo, Japan).

RT-PCR

Total RNA of these species was reverse-transcribed at 42 ℃ for 1.5 h using ReverTra Ace (Toyobo Corp., Osaka, Japan). The degenerate primers for PCR amplification of the *PSMB8* cDNA were designed based on the amino acid sequences, which are highly conserved among known *PSMB8* sequences of various species and are shared by both the A and F types of *PSMB8*. The forward primers were 5′–TGGTACCACCACCCTNGCNTTYACNTT–3′ (FP1), 5′–TGGTACCACCACRYTNGCCTTYAANTT–3′ (FP2) and 5′–TGGTACCACCACRYTNGCCTTYAARTT–3′ (FP3), where N, R and Y represent a mix of G, A, T and C, a mix of G and A, or a mix of T and C, respectively, which correspond to the amino acid sequence GTTTLAFKF, and the reverse primer was 5′–CAAGAGNCKYTCCCARTAYTGRCARTC–3′ (RP1), where K represents a mix of G and T, corresponding to the amino acid sequence DCQYWERLL. cDNA synthesized from the total RNA of samples was used as a template for PCR, with 35 cycles of amplification (denaturing at 94 ℃ for 30 s, annealing at 54 ℃ for 30 s and extension at 72 ℃ for 1 min) using Ex-Taq DNA polymerase (TaKaRa Bio Inc.). Partial cDNA products of the expected size (about 180 bp) were obtained and sequenced directly after treatment with Exonuclease I (New England Biolabs, Ipswich, MA, USA) and Antarctic phosphatase (New England Biolabs).

DNA sequencing

DNA sequence analysis was performed using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems Japan Ltd., Tokyo, Japan) with a BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems Japan Ltd.).

Rapid amplification of cDNA ends (RACE)-PCR

The SMART RACE cDNA amplification kit (Clontech Laboratories) was used to obtain the mature protein coding sequences from *PSMB8* cDNA.

Phylogenetic analysis of the *PSMB8* genes

The deduced nucleotide sequences corresponding to the PSMB8 mature protein from my samples were aligned with those of the medaka (*Oryzias latipes*), rainbow trout (*Oncorhynchus mykiss*), Atlantic salmon (*Salmo salar*), Japanese loach (*Misgurnus anguillicaudatus*), zebrafish (*Danio rerio*), nurse shark (*Ginglymostoma cirratum*), bichir (*Polypterus senegalus*), Japanese fire belly newt (*Cynops pyrrhogaster*, newt), *Xenopus laevis*, *X. tropicalis*, green anole lizard (*Anolis carolinensis*), platypus (*Ornithorhynchus anatinus*), tammar wallaby (*Macropus eugenii*) and human (*Homo sapiens*), using MUSCLE algorithms (Edgar 2004) running in MEGA 5 (Tamura et al. 2011). Based on this alignment, the phylogenetic trees were constructed using the neighbor-joining (NJ) method (Saitou and Nei 1987) using the p-distance model (Nei and Kumar 2000) and reliabilities were estimated with 500 bootstrap replications (Felsenstein 1985). The DNA sequences corresponding to the *PSMB5* mature protein of sea lamprey (*Petromyzon marinus*) and/or human were used as an outgroup. The additional phylogenetic trees were constructed using the maximum likelihood (ML) method based on the Kimura's 2-parameter model (Kimura 1980) using MEGA 5 (Tamura et al. 2011) and bootstrap possibilities were determined with 500 bootstrap replications. The DDBJ/EMBL/NCBI accession numbers of the nucleotide sequences used here are shown in the legend of Fig. 6.

Genomic DNA extraction

Genomic DNA was extracted from 1–3 mg of the muscle samples of Japanese gecko, spectacled caiman, and red-eared slider using the Puregene Genomic DNA Purification Kit (Gentra Systems) according to the manufacturer's instructions, and was finally dissolved in 30–50 μL TE (10 mM Tris, 1 mM EDTA).

Genomic PCR amplification of the *PSMB8* alleles for estimating gene frequency

To estimate allelic frequency of the *PSMB8* A and F types in wild populations, 43 wild samples of Japanese gecko were analyzed by genomic PCR. The following primer sets designed at the nucleotide sequences of exon 2 to exon 3 common to both the A and F types were used: Forward, 5'-CAGCACGGGGTCATCGTGGCCACCGAC-3' and Reverse, 5'- GGCCAGGAGCCGCTCCCAGTACTGGCA-3'. The PCR conditions were as follows: denaturation at 94 ℃ for 1 min, 40 cycles of denaturation at 98 ℃ for 10 s, annealing at 65° C for 30s and elongation at 72 ℃ for 2 min, and final elongation at 72 ℃ for 5 min with LA Taq (TaKaRa Bio Inc.). PCR products were separated on 2.0% agarose gel.

Genomic PCR amplification of the *PSMB8* alleles of *Xenopus* species

To compare the intron 4 sequences of the *PSMB8* gene of *X. laevis* and *X. tropicalis*, the F type sequences of these species were downloaded from the respective genome project pages. The A type sequences not available from genome databases were determined by genomic PCR using genomic DNA from laboratory-reared *X. laevis* and *X. tropicalis* individuals. I used genomic DNA samples from a *X. laevis* individual shown previously to possess the A type sequence (Nonaka et al. 2000); *X. tropicalis* DNA was kindly provided by Dr. M. Taira and was also shown to possess

the A type. The PCR primers were designed to recognize the A type-specific *PSMB8* gene sequences shared by *X. laevis* and *X. tropicalis* in exons 4 and 5: forward, 5′–TGATGCTCCAGTATCGCGGGTCT–3′ and reverse, 5′–TGCGYAGCGTAGCTAAYTGCAYGC–3′. The PCR conditions were as follows: denaturation at 94 ℃ for 1 min, 40 cycles of denaturation at 98 ℃ for 10 s, annealing at 63° C for 30 s, elongation at 72 ℃ for 1 min 40 s and final elongation at 72 ℃ for 2 min with LA Taq DNA polymerase (TaKaRa Bio Inc.). The nucleotide sequences were aligned using MUSCLE algorithms (Edgar 2004) running in MEGA 5 (Tamura et al. 2011).

Genomic PCR amplification of the *PSMB8* alleles of reptile species

The intron sequences of *PSMB8* gene in gecko, alligator and turtle were amplified by genomic PCR. The *PSMB8* gene sequence spanning from exon 2 to exon 3, exon 3 to exon 4, exon 4 to exon 5, and exon 5 to exon 6 was amplified using genomic DNA as a template. The primers sets designed to recognize the *PSMB8* gene sequences are listed in Table 1. The PCR conditions were as follows: denaturation at 94 ℃ for 1 min, 40 cycles of denaturation at 98 ℃ for 10 s, annealing at 62 to 68 ℃ for 30 s and elongation at 72 ℃ for 1 to 4 min and final elongation at 72 ℃ for 5 min with LA Taq DNA polymerase (TaKaRa Bio Inc.). PCR products were separated on 2.0% agarose gels and analyzed by direct sequencing.

The nucleotide sequences corresponding to the introns of the *PSMB8* gene from alligator and turtle were aligned using MUSCLE algorithms (Edgar 2004) running in MEGA 5 (Tamura et al. 2011).

Sequence comparison among *PSMB8* introns of tetrapods

Nucleotide sequences of introns were compared using the dot plot and percent identity plot generated by PipMaker (http://bio.cse.psu.edu) (Schwartz et al. 2000). Total length of the aligned sequences and their nucleotide sequence identity among haplotypes were calculated by summing up the aligned gap-free segments. In the dot plot and percent identity plot, gap-free segments with more than 50% identity between two sequences were plotted.

Results

Dimorphisms of *PSMB8* gene in tetrapods

The nucleotide sequences of the *PSMB8* cDNA encoding the mature protein were obtained by the SMART RACE cDNA amplification method. Both the A and F type alleles of *PSMB8* gene were found from all five reptile species used in this study, Japanese gecko (*Gekko japonicus*, gecko), spectacled Caiman (*Caiman crocodilus*, alligator), saltwater crocodile (*Crocodylus porosus*, crocodile), red-eared slider (*Trachemys scripta*, turtle), and Chinese soft-shell turtle (*Pelodiscus sinensis*, soft shell turtle).

All the deduced amino acid sequences of the mature protein of both the A and F type PSMB8 of these reptile species were composed of 204 residues, and were aligned without any insertion/deletion (Fig. 5). The 31st residue of all A type PSMB8 was Ala except for gecko A type possessing Val at this position. Similarly, the 31st residue of all F type PSMB8 was Phe except for F type of turtle and soft shell turtle possessing Tyr at this position. Except for the 31st position, there is no perfectly diagnostic site for the A and F types. The neighbor-joining (NJ) tree constructed based on the nucleotide sequences of the mature peptide region of various animals, using human and lamprey *PSMB5* sequences as an outgroup, indicated the presence of two dichotomous clades supported by bootstrap percentages of 76% and 78%, respectively (Fig. 6). One clade contained only the *PSMB8* sequences of shark and basal teleosts with F^{31} termed the ***PSMB8F*** lineage, and all other sequences belonged to the other clade termed the ***PSMB8A*** lineage. All *PSMB8* sequences of amphibian and reptile species belonged to the ***PSMB8A*** lineage. If the evolution of these *PSMB8* genes had followed an ordinary process through the accumulation of point mutations, without any extraordinary events

potentially disturbing phylogenetic analyses such as exchange of partial sequences between the analyzed genes, then this tree indicates the following points. The ***PSMB8F*** lineage was lost in a common ancestor of tetrapod, and the apparent functional equivalents of the ***PSMB8F*** lineage members with Phe or Tyr at the 31st position of the mature protein was regenerated within the ***PSMB8A*** lineage. The A and F type *PSMB8* sequences of newt, *Xenopus*, gecko, Crocodilia (alligator and crocodile), and Testudines (turtle and soft shell turtle) formed taxon-specific monophyletic clades, indicating that recurrent regeneration of *PSMB8* F type occurred at least five times independently during tetrapod evolution (Fig. 6). The alligator and crocodile genomes share the same dimorphism, indicating that this dimorphism was maintained in a trans-species manner. A similar trans-species dimorphism was also observed with the turtle and soft shell turtle genomes. The maximum likelihood (ML) tree constructed using the Kimura's 2-parameter model (Fig. 7) also indicated the presence of two dichotomous lineages, although the supporting bootstrap percentages were lower than the NJ tree (61% and 62 %). Taxon-specific clustering was basically conserved by the ML tree too, except that the A and F types of turtle and soft shell turtle did not form a clade. However, the bootstrap percentage to support this topology was only 34% and 50 %, and I concluded that the *PSMB8* dimorphism was revived at least five times independently in Urodela, Anura, Squamata, Testudines and Crocodilia.

Interestingly, newt and *Xenopus* sequences did not form a monophyletic cluster in both NJ tree (Fig. 6) and ML tree (Fig. 7). Whereas the newt sequences were located at the reasonable position, the *Xenopus* sequences formed a clade with bichir and nurse shark sequences although bootstrap percentage to support this clustering was very low. The A type of *PSMB8* mature protein coding sequence of Tago's brown frog (*Rana tagoi*) is obtained and the peptide sequence length is 205 a. a., identical with A type of

two *Xenopus* species (Fig. 5). When the only available A type *PSMB8* sequence of *Rana tagoi* was added to this analysis, it also formed a clade with *Xenopus* sequences without affecting the other tree topology (Fig. 8). The A type of *PSMB8* gene is downloaded from the genome database of West Indian Ocean coelacanth (*Latimeria chalumnae*, EMBL No.: ENSLACG00000014047, Fig. 5) and added to phylogenetic tree. The coelacanth A type sequence formed a cluster with bichir and nurse shark A type sequences (Fig. 8). The *PSMB8* gene cannot be identified from cDNA samples of African lungfish (*Protopterus dolloi*) and Australian lungfish (*Neoceratodus forsteri*) by PCR in my preliminary experiments.

The *PSMB8* gene or its cDNA cannot be identified from genome database of Neognathae birds: chicken, turkey, and zebra finch (Hillier et al. 2004; Dalloul et al. 2010; Warren et al. 2010). I tried to isolate *PSMB8* gene from the cDNA of a Palegnathae bird, common ostrich (*Struthio camelus*) and an aquatic and flightless Neognathae bird, African penguin (*Spheniscus demersus*). The fragment of *PSMB8* gene cannot be amplified in both bird species by degenerate primers, suggesting that the *PSMB8* gene might be lost in the common ancestor of birds.

The supposed evolutionary history of the *PSMB8* gene from the above results is summarized in Fig. 9.

The discordance between the tree topology of *PSMB8* gene and the phylogenetic relationship of vertebrates

The divergence times between the amphibian and reptiles, between the Anura and Urodela, between the Squamata and other reptiles, and between the Crocodilia and Testudines are reported to be 344 million years (My) ago, 295 My ago, 285 My ago, and 257 My ago, respectively (Shen et al. 2011). However, the tree topology

within reptile sequences as well as among the bichir, nurse shark and *Xenopus* sequences of the ***PSMB8A*** lineage shows discordance with the phylogenetic relationships among these species (Fig. 6 and Fig. 7), and the above-mentioned divergence times cannot be applied for calibrating evolutionary events of the *PSMB8* gene. This suggests that the evolution of the *PSMB8* gene belonging to the ***PSMB8A*** lineage was not straightforward and that unusual evolutionary events such as interallelic sequence homogenization might have distorted tree topology (see below).

Allelic frequency of the dichotomous *PSMB8* types in the Squamata

Although an allelic frequency of the A and F types of *PSMB8* in wild populations is an important factor for inferring a kind of selective pressure acting on dimorphism, it has been estimated only in a few species thus far because of the difficulty in obtaining suitable samples. Allelic frequencies of both the A and F type of loach species were close to 0.5, whereas allelic frequencies biased to the A type were observed in the *Oryzias* species sharing the dimorphism apparently revived tens of millions years ago in their common ancestor (Tsukamoto et al. 2009; Tsukamoto et al. 2012). The allelic frequencies of the A and F types in Japanese fire belly newt were 0.34 and 0.66, respectively (Tanaka et al. unpublished data). Allelic frequencies estimated with wild frog populations were also biased to the A type in *X. laevis* and to the F type in *X. tropicalis* (Ohta et al. 2003). I investigated allelic frequencies of the A and F types of *PSMB8* of the gecko, because many wild individuals were available for this species. Forty-three wild individuals of the gecko were analyzed by genomic PCR amplification using specific primers corresponding to the exon 2 and 3 sequences. Twelve individuals had only the A type, 20 had both the A and F types and 11 had only the F type (Fig. 10). Therefore, the allelic frequencies of the A and F types were 0.51 and

0.49, respectively.

Interallelic sequence homogenization of the *PSMB8* gene in reptiles

A close examination of amino acid substitutions among the Crocodilian and Testudine sequences shown in Fig. 11 detected two characteristic substitution patterns. One comprised taxon-specific substitutions found at four positions compatible with the topology shown in Fig. 6: $G^{88}C$, $L^{139}M$, $T^{148}S$ and $A^{154}D$. The other comprised type-specific substitutions found at five positions: $V^{16}T$, $C^{130}S$, $K^{186}R$, $E^{188}G$ and $I^{196}L$. The latter pattern implies that the A and F types were established prior to differentiation between the Crocodilia and Testudines. Interestingly, three of these five type-specific substitutions are located in exon 6, whereas no taxon-specific substitutions are found in this region. These results indicate that the evolutionary history of the Crocodilian and Testudine *PSMB8* gene is not straightforward and that different parts of the gene experienced different evolutionary events. When the amino acid sequence identities between the A and F alleles were compared separately for each exon in the Crocodilia and Testudines, exon 6 showed a notably lower identity (71–79%) than the other exons (85–98%) (Table 2). When the exon 6 sequences were compared within the same type, identities between Crocodilia and Testudines (79–96%) were higher (Table 3) than those between the A and F types of the same species, suggesting that the type-specific sequence difference of exon 6 was established before the divergence of Crocodilia and Testudines. Not only the amino acid sequence comparison, but the nucleotide sequence comparison also showed similar pattern (Table 4 and Table 5). To clarify these points, NJ trees were constructed separately for exons 2 to 5 (Fig. 12a) and for exon 6 (Fig. 12b) based on the p-distance of nucleotide sequences. In the exon 2–5 tree, taxon-specific clustering

of the sequences of Crocodilia and Testudines was supported by bootstrap percentages of 99% and 97%, respectively. On the other hand, in the exon 6 tree, the A and F type-specific clusterings were supported with bootstrap percentages of 96% and 85%, respectively. These topologies were completely reproduced by the ML trees based on the Kimura's 2-parameter model, although the bootstrap percentages supporting each clustering were lower than for the NJ tree (Fig. 13).

These results suggest that the A and F types of the *PSMB8* gene had diverged significantly from each other in the common ancestor of Crocodilia and Testudines and that sequence homogenization events occurred at the exon 2 to exon 5 region independently in these lineages. The two types of *PSMB8* sequences of the gecko are very close to each other throughout their length. In particular, no amino acid substitution was observed in the exon 6 region. Both A and F type sequences of the gecko were included in the F type cluster of Crocodilia and Testudines, suggesting that sequence homogenization to the F type sequence excluding the region encoding the 31st residue occurred recently in the gecko.

Interallelic sequence homogenization of the *PSMB8* gene in *Xenopus* species

Another example of interallelic sequence homogenization was observed for the *X. laevis PSMB8* gene. When the entire sequences were compared, type-specific clusterings of the *X. laevis* and *X. tropicalis PSMB8* genes were observed, indicating that an apparent revival of the F type occurred before the divergence of these two species (Fig. 6). However, nucleotide sequence alignment showed that the type-specific sites supporting this evolutionary history are missing from position 319 near the 3′ end of exon 4 to position 467 in the middle of exon 5 (Fig. 14a). There has been no nucleotide substitution between the A and F type sequences of *X. laevis* in

this region, strongly suggesting that interallelic sequence homogenization occurred recently in the *X. laevis PSMB8* gene. Comparison with the A and F type sequences of *X. tropicalis* indicated that the sequence of this region of the A type *PSMB8* of *X. laevis* was substituted with the corresponding sequence of the F type *PSMB8*. To clarify the 5′ boundary of homogenization, intron 4 sequences of the A and F type *PSMB8* genes of these two species were compared (Fig. 14b). Distribution of the type-specific sites indicated that the 5′ boundary of homogenization resides near the 3′ end of intron 4. The lengths of *PSMB8* intron 4 in *Xenopus* species are shown in Table 6a and the dot plot and percent identity plot between the two dichotomous alleles of the *Xenopus* species are shown in Fig. 15. In the intron regions, diagonal lines are only observed when comparing the same type of different *Xenopus* species, indicating the type-specific patterns.

PipMaker comparison of *PSMB8* dichotomous allele in reptile species

The intron lengths of the *PSMB8* gene in reptile species are shown in Table 6. The intron lengths range from 147 to about 4000 bp (Table 6b), 134 to 1401 bp (Table 6c), and 131 to 458 bp (Table 6d) in gecko, alligator, and turtle, respectively. Because of the poly-G and repeated sequences, only part (about 600 bp) of F type intron 3 in gecko could be sequenced and the accurate length of this intron is unknown.

To visualize the nucleotide sequence diversity between the dichotomous alleles, the A and F types of *PSMB8* were compared by PipMaker. By the dot plot analysis, diagonal lines were observed in all exon regions shown by shading, indicating that the exons were conserved between the two dichotomous alleles of gecko (Fig. 16a). The diagonal lines with some interruptions were found from all intron regions. Three independent diagonal lines (M1, M2 and M3, 140 to 200 bp) from the intron region

were short tandem and inverted repeats and showed high identities (about 90%) with the squamate microsatellite (Accession number: AY221140.1 and DQ923603.1) by BLAST analysis (http://blast.ncbi.nlm.nih.gov/). Notable tandem repeats were also found in intron 4 region. The percent identity plot (Fig. 16b) showed a very high identity between intron 5 region of dichotomous alleles of gecko.

The exons showed a high degree of conservation between the two dichotomous alleles of alligator, and of turtle (Fig. 17a, and Fig. 18a, respectively), whereas the diagonal lines with a lot of gaps were observed from intron regions. Between the dichotomous alleles of alligator (Fig. 17), two big gaps (800 and 1200 bp) are observed from intron 2 region and intron 4 region, respectively. Between the dichotomous alleles of turtle (Fig. 18), no big gap (> 200 bp) could be observed. The percent identity plots between the dichotomous alleles of alligator, and of turtle (Fig. 17b, and Fig. 18b, respectively) indicated low identities from intron regions.

The PipMaker analysis indicated the intron regions between dichotomous *PSMB8* alleles of alligator, and of turtle were highly diverged. There is no obvious evidence of exceptional evolutionary events that can be observed in these dot plots and percent identity plots.

Interspecies comparison of *PSMB8* dichotomous alleles

The examination of amino acid substitutions (Fig. 11), sequence identity data of each exon (Table 2 to Table 5), and phylogenetic analysis of partial *PSMB8* sequence (Fig. 12 and Fig. 13) suggested that exon 2 to 5 is taxon specific and exon 6 is type-specific in reptile species. To check if there is any possible boundary between taxon-specific exon 5 and type-specific exon 6, interspecies comparison of intron 5 was performed by PipMaker. The dot plots and percent identity plots comparing

gecko sequences with alligator sequences, and comparing gecko sequences with turtle sequences are shown in Fig. 19, and Fig 20, respectively. Although diagonal lines are observed from intron 5 region (Fig. 19a and Fig. 20a, top-left and bottom-right), they are discontinuous and often interrupted. The percent identity plot (Fig. 19b and Fig. 20b, top-left and bottom-right) also indicated that the identity between intron regions of different reptile species are low. To conclude, there is no obvious boundary of sequence homogenization that can be detected from intron 5 regions in these comparisons. Diagonal lines also could not be observed from intron 2, intron 3, and intron 4 regions between gecko and alligator sequences, or between gecko and turtle sequences (data not shown).

The dot plot and percent identity plot between alligator with turtle sequences are shown in the Fig. 21. The diagonal lines are observed in most intron regions but often interrupted by gaps (Fig. 21a), and the identity is low (Fig. 21b). From alligator and turtle intron 5 regions, there is no obvious boundary of taxon-specific exon 5 and type-specific exon 6 that could be observed by PipMaker analysis. The intron sequences seemed to be highly diverged in these species.

I tried to find the evidence of sequence homogenization directly in the alignment of intron 5 region (Fig. 22). As the results of PipMaker analysis, the intron 5 sequences of two *PSMB8* alleles in alligator and turtle seemed to be highly diverged. These intron sequences could not be aligned very well and hard to decide that the boundary between taxon-specific sequence and species-specific sequence from the alignment.

Discussion

In this thesis, I showed that both the A and F type *PSMB8* sequences are present in all five analyzed reptile species, and the phylogenetic analysis indicated that the *PSMB8* dimorphisms were established by five independent evolutionary events in tetrapods. In all cases, the 31st position of the mature protein is occupied either by Phe or Tyr with a bulky aromatic side chain (F type), or by Ala or Val with a smaller side chain (A type). Thus, if these dimorphisms were actually established independently, they provide a unique example of convergent dimorphisms, indicating the presence of an extremely strong selection for the *PSMB8* dimorphism. However, once established, these apparently independent dimorphisms seem to have been evolutionarily stable and some of them show long-term trans-species dimorphism. The persistence times for the *PSMB8* dimorphisms of Crocodilia and Testudines are estimated to be at least 110 and 181 My, respectively, based on divergence time between the alligator and crocodile or between the turtle and soft shell turtle (Shen et al. 2011). The *PSMB8* dimorphisms in *Oryzias* and *Xenopus* were also reported to have been maintained for up to 100 My in a trans-species manner (Nonaka et al. 2000; Miura et al 2010). These exceptionally long-term trans-species dimorphisms provide additional evidence for the extremely strong balancing selection.

The human PSMB8 protein, possessing Val at the 31st position and therefore classified as A type, shows a chymotrypsin-like specificity (Agarwal et al. 2010). Because the 31st position is located at the entrance of the S1 pocket, the F type PSMB8 possessing a bulky amino acid at this position has been predicted to show elastase-like specificity (Tsukamoto et al. 2012), although experimental evidence for this is still to be obtained. Obviously, dual specificity of PSMB8 could contribute to expansion of the

class I antigen recognition repertoire, increasing the fitness of heterozygous individuals by providing them with a chance to cope with various kinds of pathogens. Therefore, the *PSMB8* dimorphism could be under the overdominance-type balancing selection.

In this study, interallelic sequence homogenization is suggested as a possible mechanism to explain the evolutionary pattern of the *PSMB8* genes of the Crocodilia and Testudines. The exon 6 sequences of the A and F types of *PSMB8* of these animals formed type-specific clades upon phylogenetic analysis (Fig. 12b and 13b), whereas the taxon-specific clades of the Crocodilia and Testudines were observed in the exon 2–5 tree (Fig. 12a and 13a). These two trees suggest that the A and F type sequences of Crocodilia and Testudines once diverged to a certain extent, after which sequence homogenization occurred and then was repeated in the exon 2 to exon 5 region independently in the two groups. The exact boundaries of sequence homogenization events are difficult to determine because of the high degree of sequence identity, even between the once diverged A and F types; however, the boundaries seem to differ between the Crocodilia and Testudines, given that type-specific residues were observed at the 54^{th} and 71^{st} positions in Crocodilia and at the 67^{th}, 87^{th} and 97^{th} positions in Testudines. Amino acid sequence identities between the A and F types of each species, calculated separately for each exon (Table 2 and 4), were 71–79% for exon 6 and 85–98% for exons 2–5, providing further support for this explanation. Because the evolutionary rate is not so different between exon 6 and the other exons of the *PSMB8* gene in reptiles (Table 3 and 5), a different history of homogenization seems to be the most plausible explanation for the observed difference in sequence identities between exon 6 and the other exons. The A and F types of gecko *PSMB8* showed extremely high sequence identities in all exons (Table 2 and Table 4) and phylogenetic analysis of exon 6 sequences located both the A and

F type sequences of the gecko *PSMB8* into the F type clade of Crocodilian and Testudines (Fig. 12b and 13b). These results suggest that considerable sequence diversification between the A and F types of *PSMB8* occurred in a common ancestor of the Squamata, Crocodilia and Testudines and that sequence homogenization to the F type sequence, except for the 31st residue, occurred in the gecko lineage, most probably followed by frequent homogenization events, until recently. A clearer sign of interallelic sequence homogenization was found in the *X. laevis PSMB8* gene, although the range of homogenization was short and did not affect the topology of the phylogenetic tree based on the entire coding sequence. This homogenization is likely to be a recent event because no nucleotide substitution between the A and F type sequences is observed in the homogenized region. Identification of multiple interallelic homogenization events in reptile and amphibian *PSMB8* genes suggests that homogenization was not uncommon during evolution of the *PSMB8* gene.

I speculate that interallelic sequence homogenization of the *PSMB8* gene also occurred outside the reptilian and amphibian lineages. A straightforward explanation of the evolutionary history of the *PSMB8* dimorphism shown in Fig. 6 is that the ancient dichotomous lineages of *PSMB8*, the ***PSMB8A*** and ***PSMB8F*** lineages, have been conserved by basal actinopterygians such as the Polypteriformes, Cypriniformes and Salmoniformes for more than 400 My since the divergence of bichirs from the other Actinopterygii (Fujito and Nonaka 2012) and that the ***PSMB8F*** lineage has been lost at least twice in the acanthopterygian and tetrapod lineages. It seems that the F type *PSMB8* has since been revived multiple times independently by *de novo* mutations at the 31st amino acid position of the ***PSMB8A*** lineage. However, this scenario necessitates the absence of the dimorphism for a certain period from the loss of the ***PSMB8F*** lineage to the revival of the F type *PSMB8* within the ***PSMB8A*** lineage,

showing an apparent discrepancy with the extremely strong balancing selection suggested by the long-term trans-species dimorphism and recurrence of the dimorphism. An alternative scenario is that the apparent loss of the ***PSMB8F*** lineage is the result of sequence homogenization arising from interallelic recombination or gene conversion by which most parts of the ***PSMB8F*** lineage sequence were replaced by the ***PSMB8A*** lineage sequence except for the close vicinity of the F^{31} encoding region. As the ***PSMB8A*** and ***PSMB8F*** lineage sequences seem to have diverged so much from each other prior to the appearance of the Osteichthyes (Fig. 6), the barrier for interallelic sequence homogenization between the ***PSMB8A*** and ***PSMB8F*** lineage alleles should have been very strong. However, once such sequence homogenization had occurred, the following interallelic gene homogenization between the A and F type *PSMB8* alleles should have been much easier thanks to the high degree of sequence identity at most parts of the gene. Thus, the A and F type sequences could have kept a high degree of sequence identity for a long time through frequent sequence homogenization events. Nevertheless, it is possible that homogenization did not occur for a while simply by chance and that the A and F type sequences diverged to a certain degree, at which point interallelic gene homogenization would have been difficult. If so, then the A and F alleles would have followed the usual diversification process simply depending on the time elapsed. If the actual evolution of *PSMB8* dimorphism followed this scenario, the open circles in Fig. 6 would correspond to the timing when the last interallelic homogenization occurred in that animal lineage. The hypothetical evolutionary process of the dimorphism of the reptilian and amphibian *PSMB8* gene based on the homogenization scenario is shown schematically in Fig. 23. In this scenario, I postulate the sequence homogenization by which most parts of the ***PSMB8F*** lineage sequence were replaced by the ***PSMB8A*** lineage sequence except in

the close vicinity of the F^{31} encoding region. Sequence homogenization continued in the amphibian lineage beyond the divergence between newts and *Xenopus*, stopped earlier in the *Xenopus* lineage, and then later in the newt lineage. In contrast, sequence homogenization stopped in the reptile lineage, and the A and F type sequences had diverged before divergence of the Squamata from other reptiles. However, homogenization revived in the Squamata and continued until recently. Sequence homogenization was also revived in the common ancestor of Crocodilia and Testudines. However, exon 6 was not homogenized in this case. It is still uncertain whether the actual evolution of the *PSMB8* dimorphism followed this scenario based on sequence homogenization, or the convergent dimorphism scenario based on typical mutation processes. In either case, the evolutionary pattern of the *PSMB8* dimorphism is totally unprecedented, suggesting the presence of extremely strong balancing selection.

There are still many interesting features about immunoproteasomes and *PSMB8* gene that are not fully explored:

Besides to increase the generation of peptides appropriate for MHC class I presentation, the further function of immunoproteasome is very puzzling. Recently, it was reported that immunoproteasomes degrade ubiquitinated proteins faster than the constitutive proteasomes, polyubiquitin conjugates accumulate after interferon-γ treatment but then are preferentially degraded by immunoproteasomes, and that immunoproteasome deficiency causes the formation of inclusions and more severe experimental autoimmune encephalomyelitis (EAE) (Seifert et al. 2010). In contrast, results from another report (Nathan et al. 2013) do not support all conclusions above and suggest that immunoproteasomes do not degrade ubiquitinated proteins more efficiently than constitutive proteasomes. Thus, the full physiological function of

immunoproteasome is not very clear yet and experimental evidences are still to be obtained.

The Neognathae birds appear to represent natural knock-out animals deficient in immunoproteasomes (Sutoh et al. 2012). In this study, the *PSMB8* gene also cannot be identified in a basal Palegnathae bird, ostrich. However, the *PSMB8* gene with dimorphism was identified in the closest extant relatives of bird, alligator and crocodile. These results suggested that birds lost the *PSMB8* gene after the divergence from the most recent common ancestor of birds and Crocodilia, about 241 My ago (Fig. 1). Although birds have no immunoproteasome, their constructive proteasome may cover the function of immunoproteasome and more researches in cell physiology field are required to prove it.

In this thesis, I speculate that interallelic sequence homogenization of the *PSMB8* gene also occurred outside the reptilian and amphibian lineages. Recently, another evidence was found in teleost, pufferfishes (Tetraodontidae). The dichotomous *PSMB8* alleles with extreme high identity (98-99%) were established by four independent evolutionary events in four species of Tetraodontidae: green spotted puffer (*Tetraodon nigroviridis*), grass puffer (*Takifugu niphobles*), black spotted puffer (*Arothron nigropunctatus*), and Valentinni's sharpnose puffer (*Canthigaster valentini*) (Genjima et al. unpublished data). The divergence time between these four pufferfish species and other Tetraodontidae species is only about 100 My ago (Yamanoue et al. 2011). It is hard to consider that the F type *PSMB8* could be revived four times independently only by *de novo* mutations at the 31st amino acid position in such a short period. I think the independent *PSMB8* dimorphism in pufferfishes is also an indirect evidence for the sequence homogenization hypothesis.

The dichotomous *PSMB8* alleles with high sequence identity (about 95%) also

established by two independent evolutionary events in two eel species (Anguilliformes): Kidako moray (*Gymnothorax kidako*), and Japanese eel (*Anguilla japonica*) (Noro et al. unpublished data). Surprisingly, both A and F type of the *PSMB8* gene in two Anguilliformes species belonged to the ***PSMB8F*** lineage and indicated again that both A and F types of *PSMB8* are under strong balancing selection force.

More data is still needed to clarify the evolutionary history of the dichotomous *PSMB8* alleles in tetrapods. One of my conclusions is that the ***PSMB8F*** lineage was lost in a common ancestor of tetrapod. Coelacanths and lung fishes, the closest extant relatives of tetrapods, should be appropriate materials for exploring the disappearance of the ***PSMB8F*** lineage in tetrapods, although the *PSMB8* gene cannot be identified from cDNA samples of lungfishes in my preliminary results.

Until now, the dichotomous *PSMB8* alleles are only found in ectothermic vertebrates and mammals only have A type of *PSMB8*. In the preliminary study, only A type of *PSMB8* could be identified from genomic DNA samples of eight marsupial species and two monotreme species (Yu et al. unpublished data). Interestingly, the 31st position of PSMB9 of mammals is occupied by Phe (F) whereas that position of most other animals is occupied by Met (M). It is possible that mammalian PSMB9 has a similar specificity as F type PSMB8, but needs more experimental evidences to prove it.

Since the PSMB8 dimorphism is only found in ectothermic vertebrates, another possible function is that two types of the PSMB8 are used for regulating the immune activity triggered in different surrounding temperature. Temperature is a principle cue that could stimulate changes in the immune response in teleosts (Bly and Clem 1992). Moreover, seasonal change affects lymphocytes distribution and leukocytes functions

in turtle (Muñoz et al. 2000, Muñoz and De La Fuente 2004). Thus, two type of PSMB8 may have different level of enzyme activity in different surrounding temperature but more experimental evidences are needed.

Not only the *PSMB8* gene, but also many immunity related genes of reptile have not been fully explored yet. Recently, a few studies of the reptile immunity have been reported: Hepcidin in Simensis crocodile (*Crocodylus siamensis*) (Hao et al. 2012), the immunoglobulin heavy chain locus of the anole lizard (*Anolis carolinensis*) (Gambón Deza et al. 2009), cathelicidins in elapid snakes (*Naja atra*, *Bungarus fasciatus* and *Ophiophagus hannah*) (Zao et al. 2008) and Myeloid differentiation factor 88 (MyD88) in Chinese soft shell turtle (*Pelodiscus sinensis*) (Li et al. 2010). Further researches are needed in this field, and I wish this thesis could be a modest spur to induce others to come forward with valuable contributions.

Table 1. The primer sets used for *PSMB8* intron amplification in reptile species

(a) Gecko

	Forward primer	Reverse primer
Intron 2V & F	CAGCACGGGGTCATCGTGGCCACCGAC	GGCCAGGAGCCGCTCCCAGTACTGGCA
Intron 3V	CTGGGCACCATGTCCGGA	TATTGGCCAAGAGCTTGGAG
Intron 3F	AAAGTGATCGAGATCAACCCCTACC	CCCCGGTACTCGTACAGCATATTG
Intron 4V & F	CTCCAAGCTCTTGGCCAATA	AGGCATACGAGTTTCCACTT
Intron 5V & F	AAGTGGAAACTCGTATGCCT	CTCTGCGTACTGGTCGAGCA

(b) Alligator

	Forward primer	Reverse primer
Intron 2A & F	TTCAAGTTCCAGCACGGCGTGGTC	CGGCGCTGCCCGACATGGTGCCCA
Intron 3A & F	CTGGGCACCATGTCGGGCAGCGCCGC	CACGGTACTCGCCCAGCATGTTGGCC
Intron 4A	GAGCGGATCAGCGTCTCGGCTGCC	AGCACCCCATAAGCATAGGTGTTT
Intron 4 F	GAGCGGATCAGCGTTTCAGCTGCT	AGAACCCCATAGGCGTAAGTATTC
Intron 5A	GATGACAATGGGACCCGCTTCTCT	CTGACGTCCGTCCGCTCCACCTTG
Intron 5 F	GATGACAATGGGACCCGCCTCTCC	CTGACATCTGTCCGTCCCACCCGA

(c) Turtle

	Forward primer	Reverse primer
Intron 2A & Y	CAGCACGGGGTCGTCGTGGC	CATGGTGCCCAGCAGGTAGG
Intron 3A	TACCTGCTGGGCACCATGTCGGGG	AGCCCCATGCCCCGGTACTCGCAC
Intron 3Y	TACCTGCTGGGCACCATGTCGGGC	AGCCCCATGCCCCGGTACTCGCAG
Intron 4A	CCTCCAAGCTGCTGTCGAACATGA	TGTCCATCACCCCGTAGGCGTAGC
Intron 4Y	CCTCCAAGCTGCTGTCGAACATGC	TGTCCATCACCCCGTAGGCGTAGG
Intron 5A	GGCCCCGGGCTGTACTACGTGGAC	GGCTGACGTCTGTCTTTTCCACCT
Intron 5Y	GGCCCCGGGCTGTACTACGTGGAC	AGCAGGTCACTGACGTCCGTCCGC

Table 2. The amino acid sequence identity between two dichotomous alleles of reptile species calculated for each exon as well as whole mature protein of PSMB8.

Species	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Total
Alligator	88%	92%	98%	94%	79%	91%
Crocodile	85%	89%	98%	96%	71%	90%
Turtle	96%	92%	93%	93%	71%	90%
Soft shell turtle	96%	92%	88%	88%	71%	88%
Gecko	100%	94%	100%	99%	100%	98%

Table 3. The amino acid sequence identity between Crocodilian and Testudine species calculated for each exon as well as whole mature protein of PSMB8.

Alligator A v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle A	96%	97%	88%	90%	82%	91%
Soft shell turtle A	96%	94%	88%	91%	86%	91%
Crocodile A v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle A	96%	94%	88%	91%	82%	90%
Soft shell turtle A	96%	92%	88%	93%	79%	90%
Alligator F v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle Y	96%	92%	93%	90%	89%	91%
Soft shell turtle Y	96%	92%	93%	87%	96%	91%
Crocodile F v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle Y	92%	92%	93%	93%	89%	92%
Soft shell turtle Y	92%	92%	93%	87%	96%	91%

Table 4. The nucleotide sequence identity between two dichotomous alleles of reptile species calculated for each exon as well as whole mature protein of PSMB8.

Species	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Total
Alligator	85%	91%	94%	85%	77%	87%
Crocodile	86%	95%	98%	96%	76%	92%
Turtle	95%	91%	95%	92%	78%	91%
Soft shell turtle	95%	91%	91%	92%	79%	90%
Gecko	97%	96%	99%	98%	100%	98%

Table 5. The nucleotide sequence identity between Crocodilian and Testudine species calculated for each exon as well as whole mature protein of PSMB8.

Alligator A v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle A	87%	93%	88%	84%	85%	87%
Soft shell turtle A	90%	94%	87%	84%	83%	87%
Crocodile A v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle A	87%	94%	89%	89%	84%	89%
Soft shell turtle A	90%	94%	88%	91%	84%	90%
Alligator F v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle Y	85%	92%	90%	87%	85%	88%
Soft shell turtle Y	89%	89%	88%	87%	87%	88%
Crocodile F v.s.	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Total**
Turtle Y	84%	93%	92%	92%	85%	90%
Soft shell turtle Y	89%	90%	90%	90%	85%	89%

Table 6. The intron length (bp) of the *PSMB8* gene in three reptile and *Xenopus* species

(a) *Xenopus*

	Intron 4
X. laevis A	4010
X. tropicalis A	1538
X. laevis F	911 (from Xenbase)
X. tropicalis F	6482 (from Xenbase)

(b) Gecko

	Intron 2	**Intron 3**	**Intron 4**	**Intron 5**
Gecko V	945	874	1859	147
Gecko F	1606	about 4000	1146	149

(c) Alligator

	Intron 2	**Intron 3**	**Intron 4**	**Intron 5**
Alligator A	134	163	1401	303
Alligator F	999	202	272	408

(d) Turtle

	Intron 2	**Intron 3**	**Intron 4**	**Intron 5**
Turtle A	334	208	352	271
Turtle Y	247	131	458	251

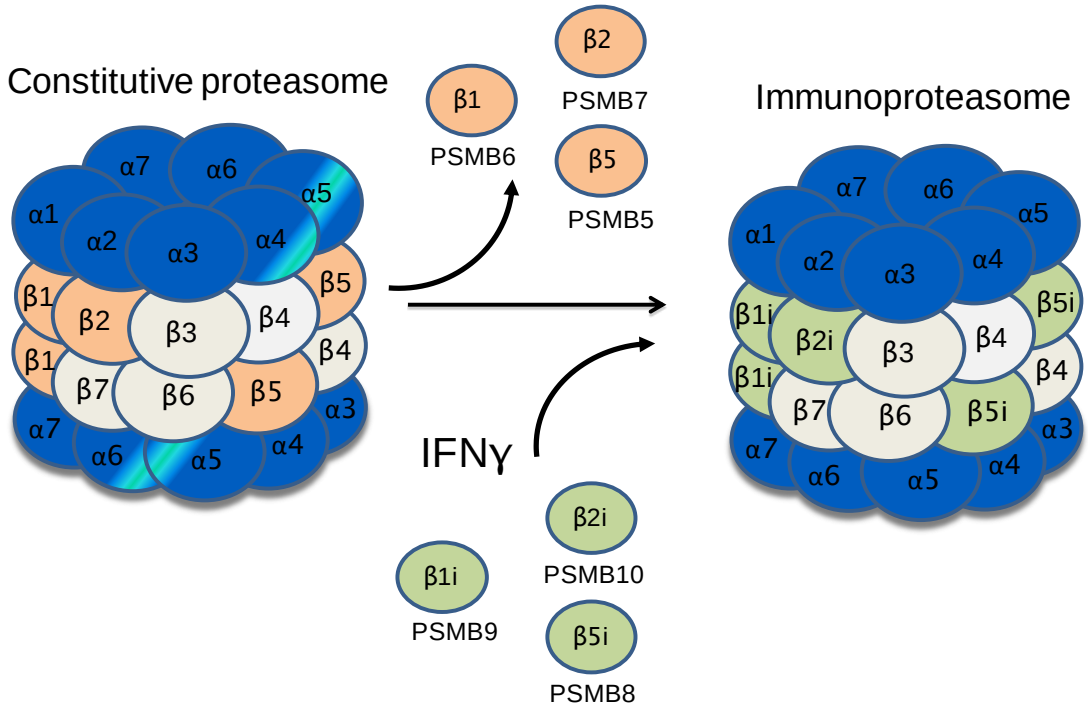


Figure 1. Schematic diagram of constitutive proteasome and immunoproteasome

Alpha and beta subunits are shown in blue and gray, respectively, except for the beta subunit with peptidase activities. These active beta subunits of a constructive proteasome are PSMB5 (β5), PSMB6 (β1), and PSMB7 (β2) shown in orange. They are replaced by IFNγ inducible beta subunits with green color, PSMB8 (β5i), PSMB9 (β1i), and PSMB10 (β2i), respectively, forming an immunoproteasome.

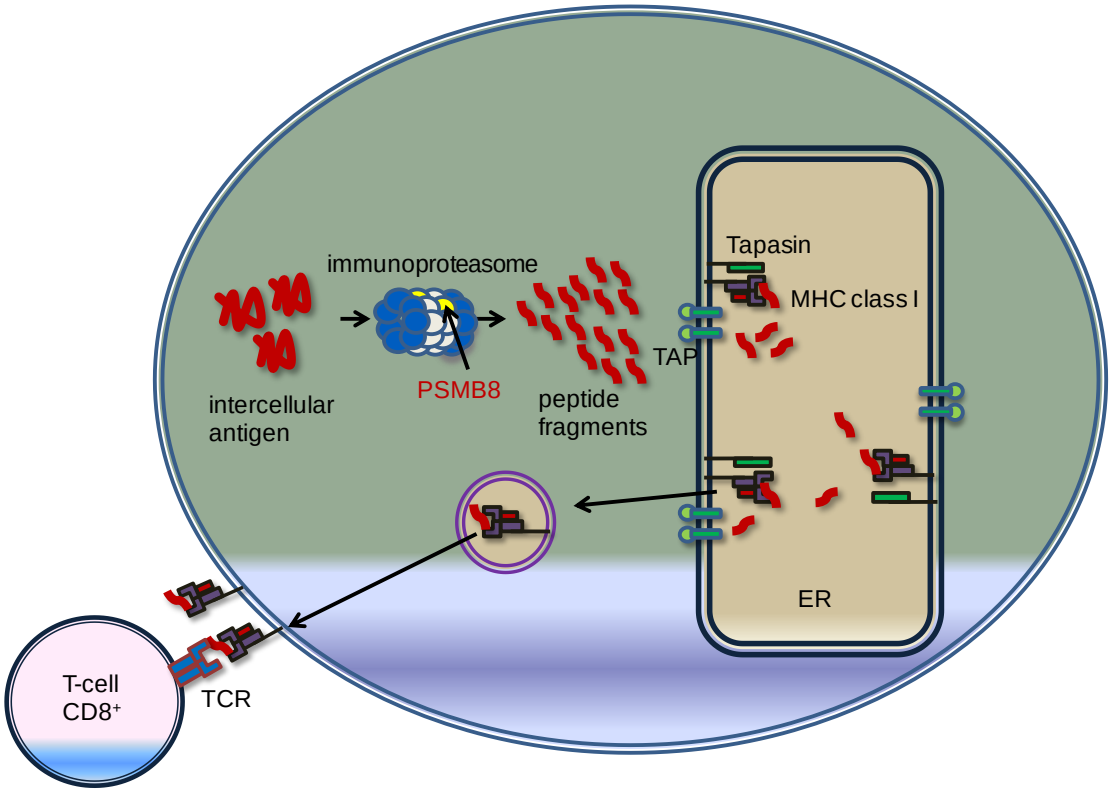


Figure 2. Schematic diagram of MHC class I antigen presentation pathway

First, intercellular antigens are cleaved into small antigen peptides in the cell by the immunoproteasome. Second, antigen peptides generated by the immunoproteasome are transferred into the ER lumen by TAP. Third, the MHC class I molecule in a complex with TAP and tapasin makes peptide binding easier. Finally, antigen peptides connected with MHC class I molecules are transported through the secretory pathway to the cell surface and is recognized by the T-cell receptor (TCR) of $CD8^+$ T cells.

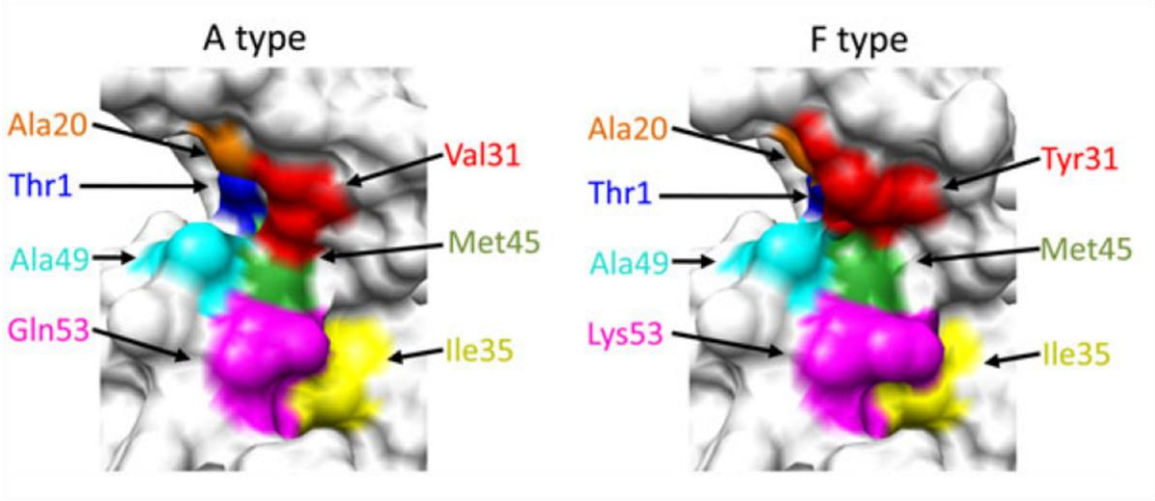


Figure 3. Predicted 3D structures of the A and F type PSMB8 molecules in medaka

This figure is cited from Tsukamoto et al. 2012. The 3D structures of the A and F type PSMB8 molecule in medaka were predicted based on the steric structure of the bovine PSMB5 molecule (PDB ID; 1iruL) by SWISS-MODEL server (http://swissmodel.expasy.org/). The six residues forming the S1 pocket are indicated by orange (20th position), red (31st), yellow (35th), green (45th), cyan (49th), and magenta (53rd), respectively. The threonine residue acting as the general base for the hydrolysis reaction is shown in blue. These views are from the inside of the S1 pocket.

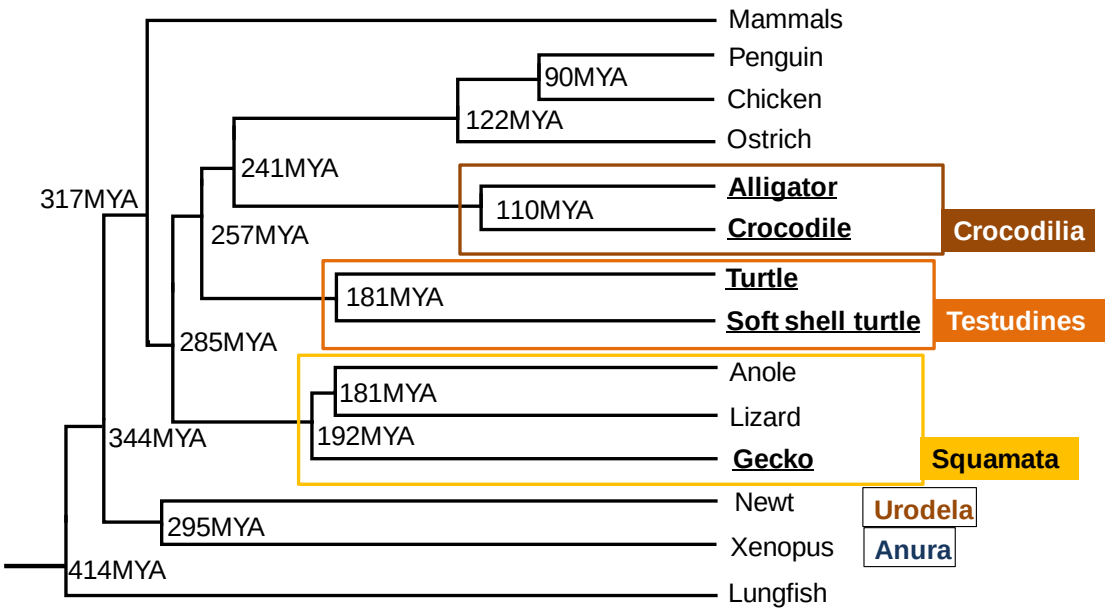


Figure 4. The phylogeny tree of tetrapods

This tree was modified from Shen et al. 2011 and Slack et al. 2006. The reptile species used for this study are underlined and three major reptile orders are marked.

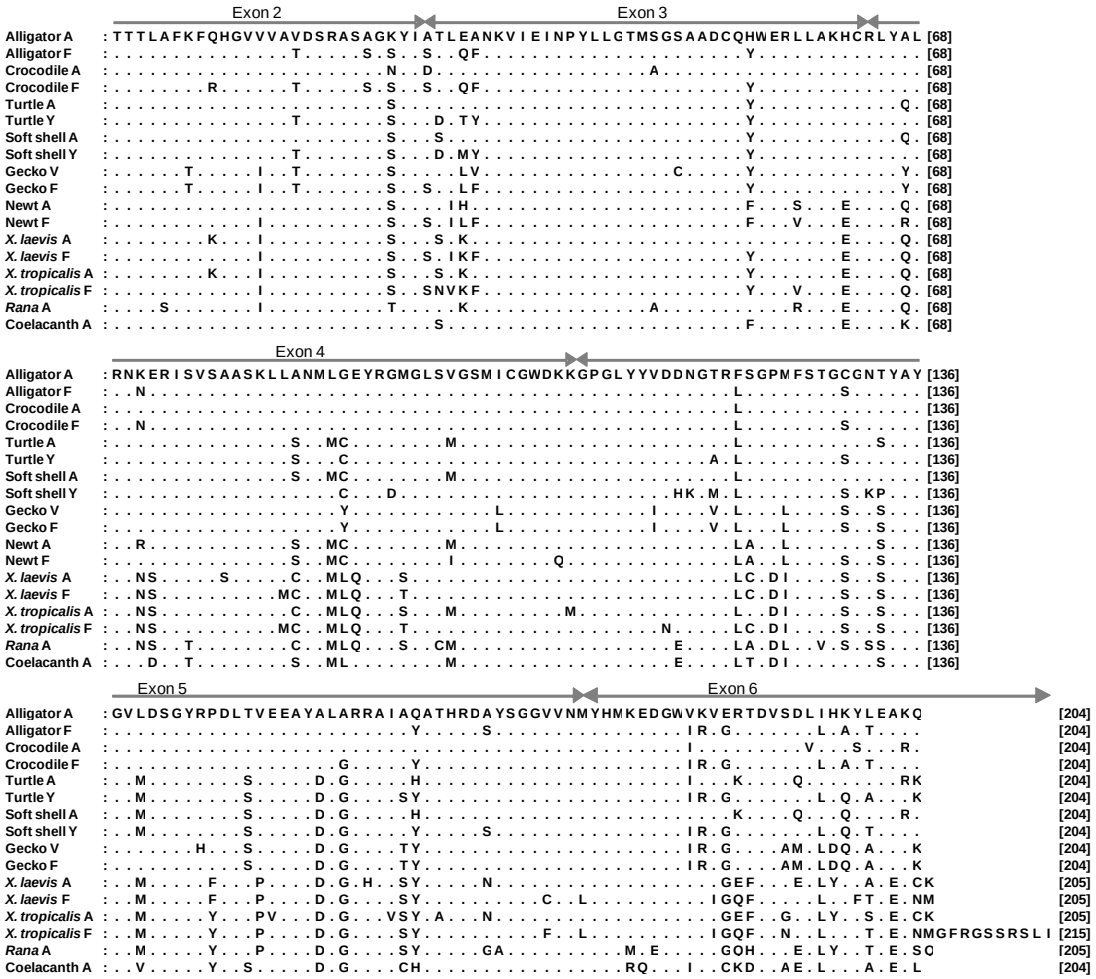


Figure 5. Alignment of the amino acid sequences of the mature proteins of tetrapod PSMB8.

The deduced amino acid sequences of A and F types of PSMB8 mature proteins of tetrapods were aligned using MUSCLE (Edgar 2004). Names of sequences are shown by the animal name followed by the single letter code for the 31st amino acid residue. Identity to the uppermost sequence is shown by dots. The ranges of the exons are indicated by the double-headed arrows above the alignment. Soft shell is soft shell turtle, *Rana* is *Rana tagoi*, *X. laevis* is *Xenopus laevis*, and *X. tropicalis* is *Xenopus tropicalis*.

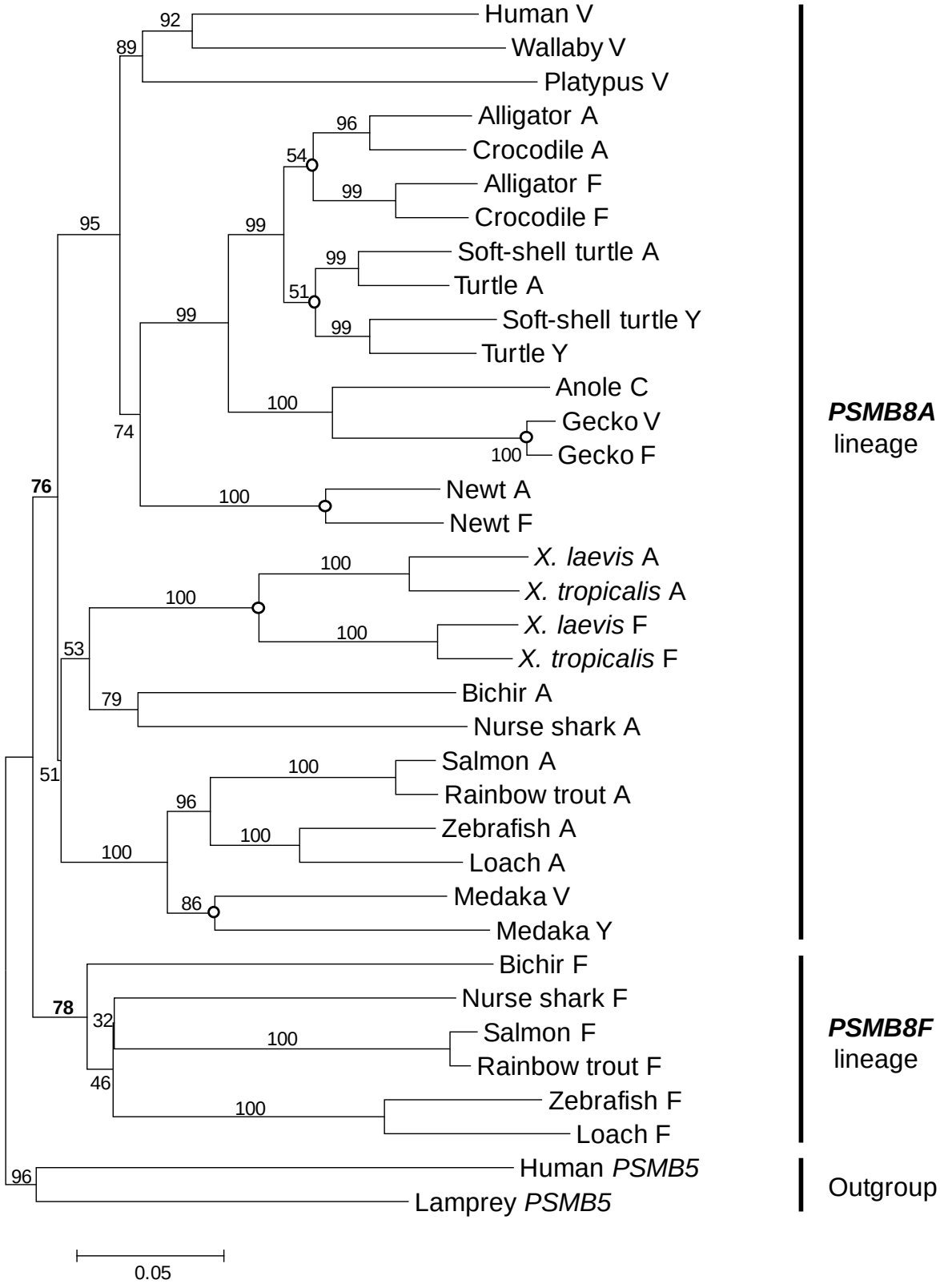


Figure 6. NJ phylogenetic tree based on the nucleotide sequences corresponding to the PSMB8 mature protein sequences.

A neighbor-joining (NJ) tree was constructed based on the p-distance method mounted in MEGA5 (Tamura et al. 2011). The human and sea lamprey *PSMB5* sequences were used as an outgroup. Bootstrap percentages based on 500 replicates, are shown on each branch. The 31st residue of the mature protein is shown after the name of the species and the open circles correspond to the timing when the apparent revival of dimorphism occurred in that animal lineage. The scale bar represents 0.05 substitutions per nucleotide site. DDBJ/EMBL/NCBI accession numbers of the nucleotide sequences used here are as follows: alligator *PSMB8*, AB775808 and AB775809; crocodile *PSMB8*, AB775810 and AB775811; turtle *PSMB8*, AB775812 and AB775813; soft shell turtle *PSMB8*, AB775814 and AB775815; gecko *PSMB8*, AB775816 and AB775817; newt *PSMB8*, AB775818 and AB775819; *Oryzias latipes* (medaka) *PSMB8*, AB183488 and BA000027; rainbow trout *PSMB8*, CA360946 and GE824997; Atlantic salmon (salmon) *PSMB8*, DY720791 and DW555236; Japanese loach (loach) *PSMB8*, BJ830309 and BJ838416/AB602774; zebrafish *PSMB8*, BC066288 and BC092889; nurse shark *PSMB8*, D64056 and D64057; bichir *PSMB8* AB686528 and AB686530; *X. laevis* (*Xenopus laevis*) *PSMB8*, D44540 and D44549; *X. tropicalis* (*Xenopus tropicalis*) *PSMB8*, DT453345 and DN016480; green anole lizard (anole) *PSMB8*, NW_003339585; platypus *PSMB8*, ABU86903.1, wallaby *PSMB8*, ENSMEUG00000007328; human *PSMB8*, NP_004150; human *PSMB5*, NM_002797; and sea lamprey *PSMB5*, D64055.

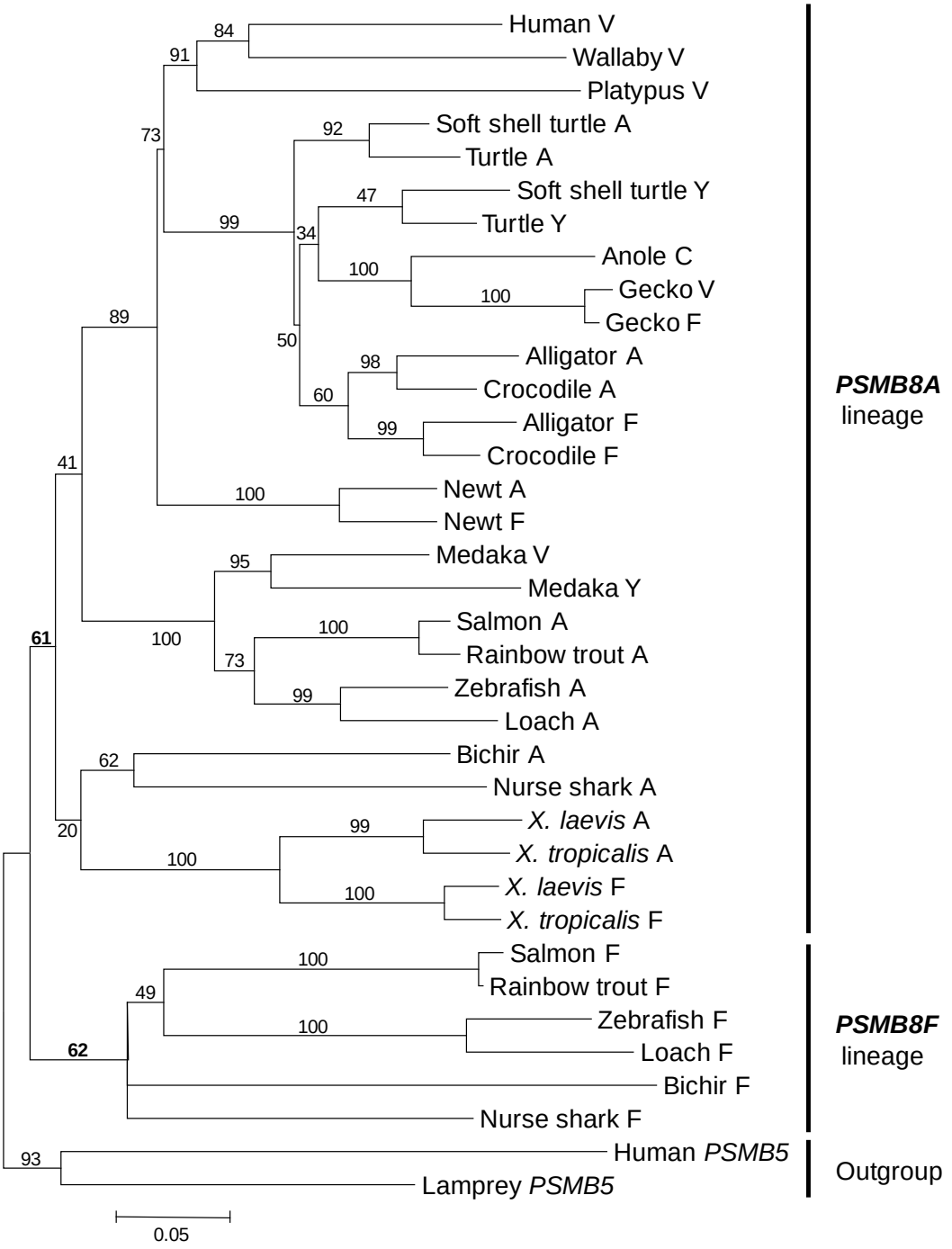


Figure 7. ML phylogenetic tree based on the nucleotide sequences corresponding to the PSMB8 mature protein.

A maximum-likelihood (ML) tree was constructed based on the Kimura's 2-parameter model mounted in MEGA5 (Tamura et al. 2011). The human and sea lamprey *PSMB5* sequences were used as an outgroup. Bootstrap percentages based on 500 replicates, are

shown on each branch. The 31st residue of the mature protein is shown by the single letter code after the name of the animal. The scale bar represents 0.05 substitutions per nucleotide site. DDBJ/EMBL/NCBI accession numbers of the nucleotide sequences used here are summarized in the legend of Fig. 6.

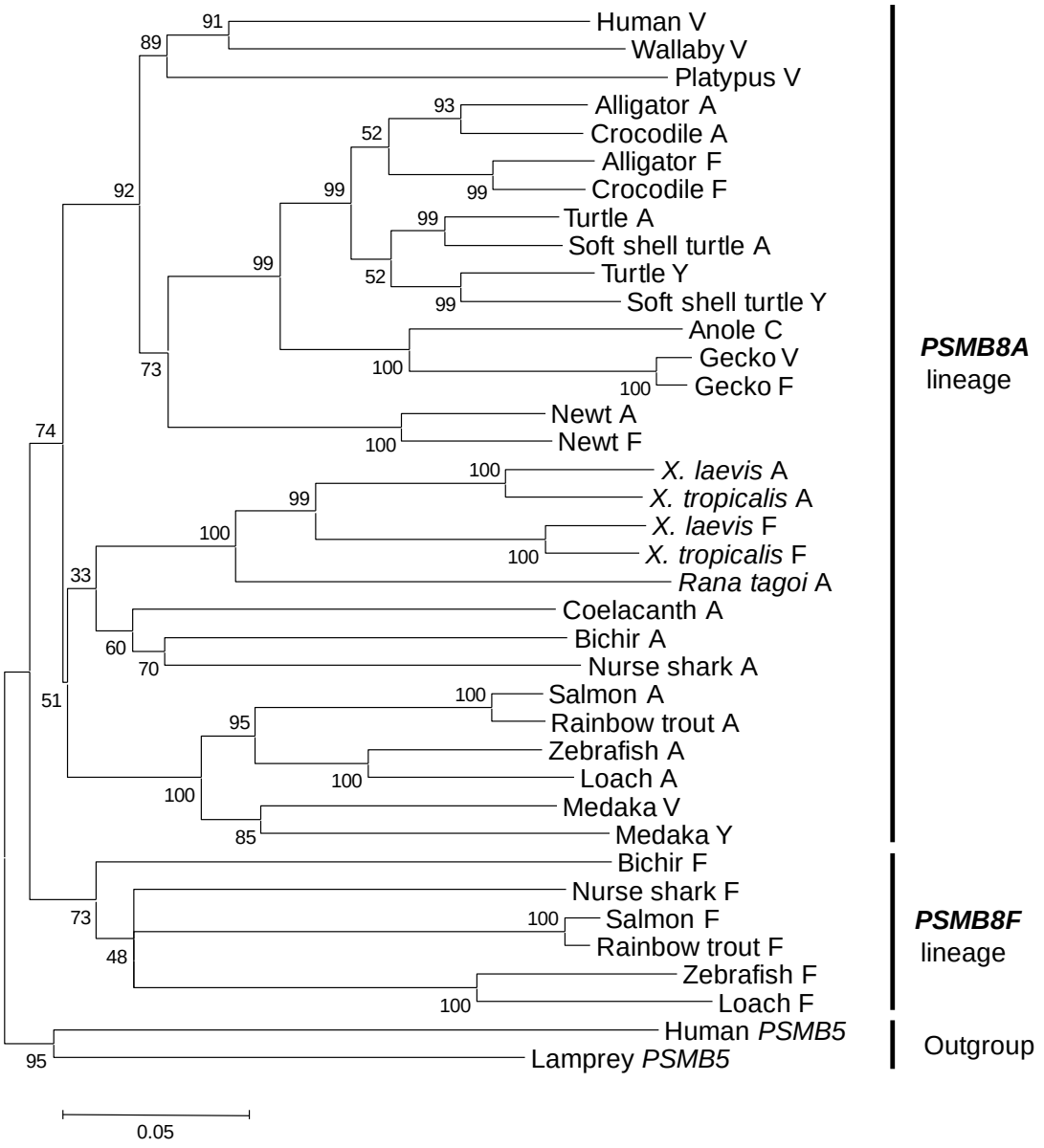


Figure 8. NJ phylogenetic tree based on the nucleotide sequences corresponding to the PSMB8 mature protein sequences with additional *Rana tagoi* and coelacanth sequence.

A neighbor-joining (NJ) tree was constructed based on the p-distance model mounted in MEGA5 (Tamura et al. 2011). The human and sea lamprey *PSMB5* sequences were used as an outgroup. Bootstrap percentages based on 500 replicates, are shown on each branch. The 31st residue of the mature protein is shown after the common name of the species. The scale bar represents 0.05 substitutions per nucleotide site.

Additional EMBL accession number of the nucleotide sequences used here is West Indian Ocean coelacanth *PSMB8* (*Latimeria chalumnae*, EMBL No.: ENSLACG00000014047).

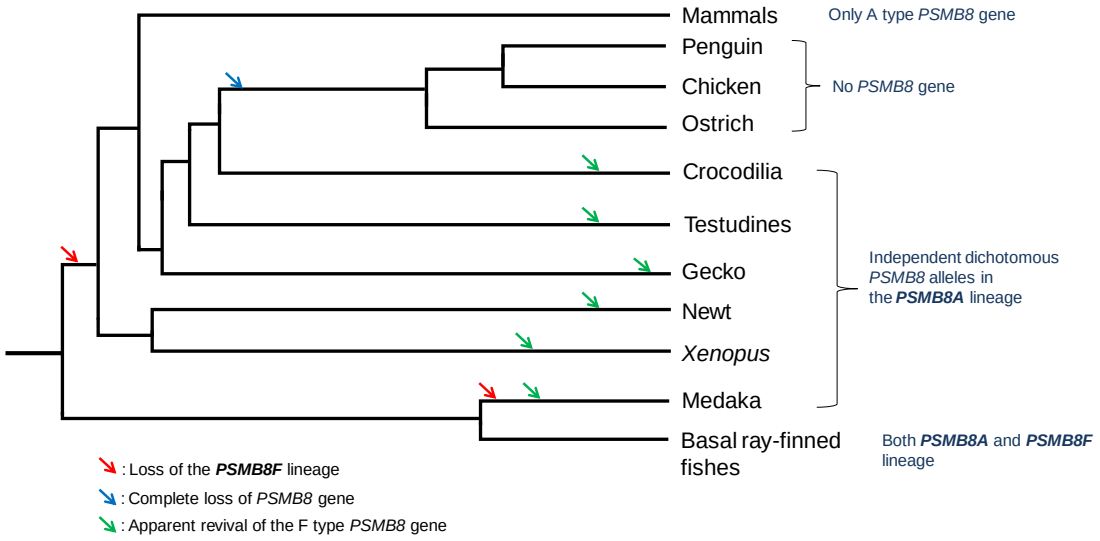


Figure 9. The evolutionary scenario of allelic dimorphism of *PSMB8* gene in jawed vertebrates based on phylogenetic analysis

The different colored arrow sign shows the important evolutionary events: the red arrow shows the loss of ***PSMB8F*** lineage in medaka and tetapods, the green arrow shows apparent revival of the F type *PSMB8* gene, and the blue arrow shows the complete loss of the *PSMB8* gene in birds. The F type *PSMB8* gene was apparently regenerated from ***PSMB8A*** lineage in medaka, *Xenopus*, newt, Crocodilia, Testudines and gecko, independently. The length of tree branches does not reflect the real divergence times.

Figure 10. *PSMB8* genotyping in Japanese gecko.

The *PSMB8* gene region spanning from exon 2 to exon 3, showing size difference between the A and F types (A type, 0.9 kb; F type, 1.6 kb) was amplified by genomic PCR using species-specific primers. Of 43 typed samples, the results of 15 samples were shown here. The genotypes of wild individuals were as follows: A type/A type, 12 individuals; F type/A type, 20 individuals; and F type/F type, 11 individuals.

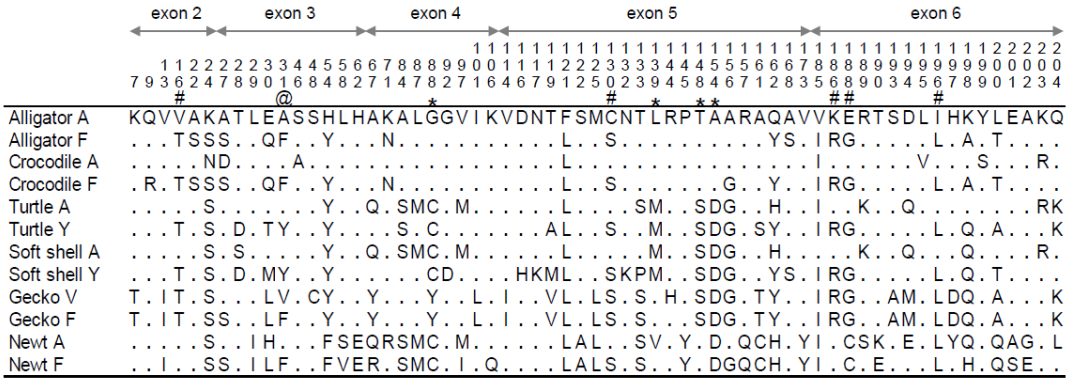
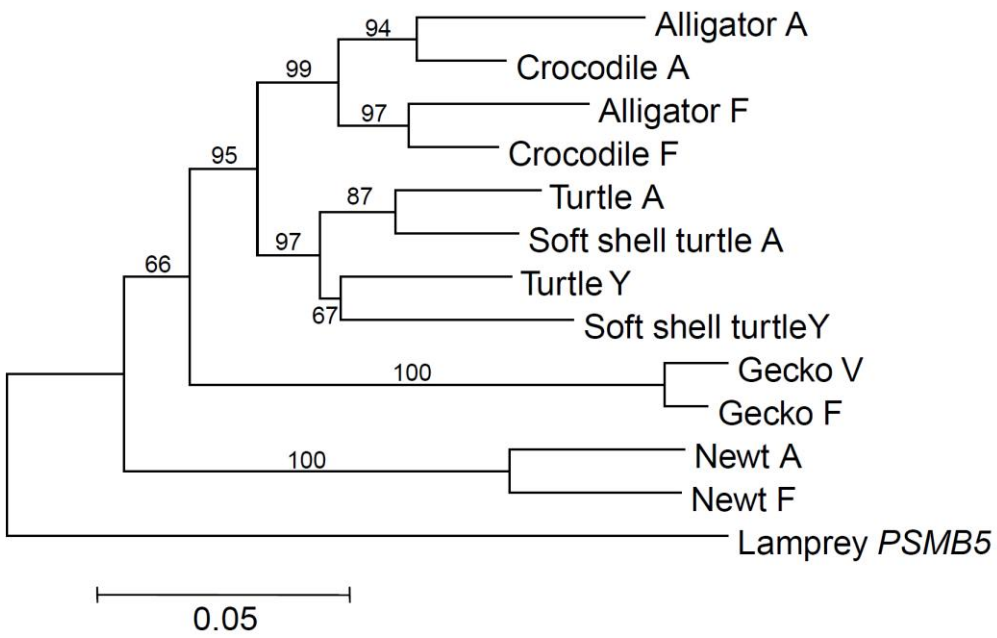


Figure 11. Comparison of the amino acid sequences of the mature proteins of PSMB8 of reptiles and newt.

The alignment of the amino acid sequences of the entire mature protein of PSMB8 is shown in Fig. 5. Sixty-three of the 204 positions showing amino acid substitutions are depicted here. Dots indicate the identity of the residues with the uppermost sequence. The 31st residue of the mature protein is indicated by @. The residues of taxon-specific substitutions and type-specific substitutions in Crocodilia and Testudines are indicated by * and #, respectively. The ranges of the exons are indicated by the double headed arrows above the alignment. An alphabet after the animal name is a single letter code for the 31st amino acid residue. Soft shell indicates soft shell turtle.

(a)



(b)

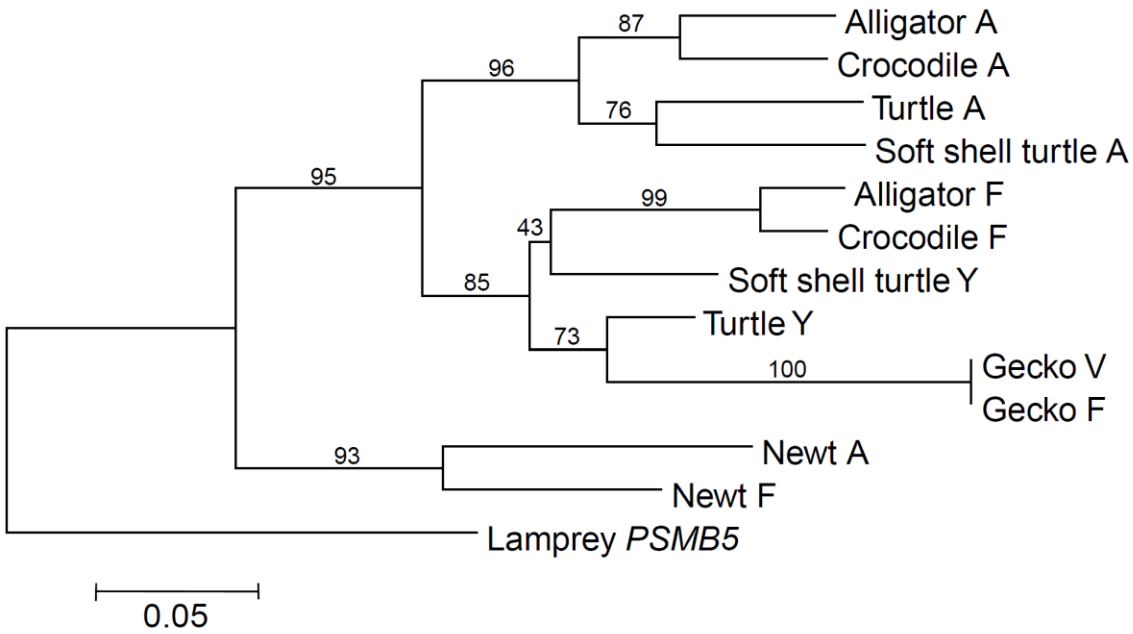
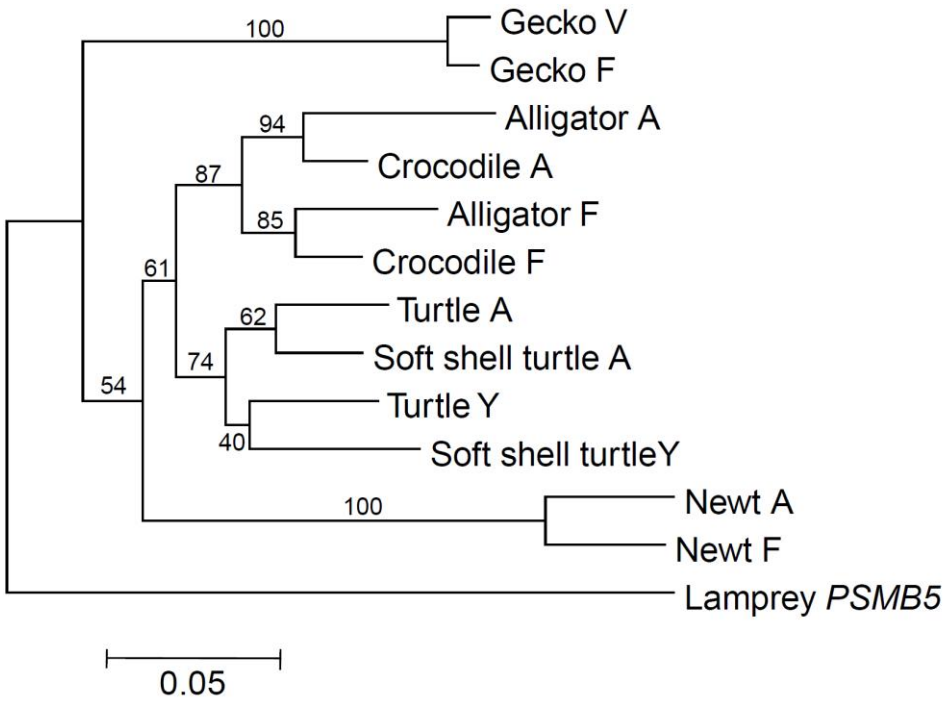


Figure 12. NJ phylogenetic trees based on the nucleotide sequences of different regions of the *PSMB8* gene.

Two neighbor-joining (NJ) trees were constructed based on the p-distance using MEGA5 (Tamura et al. 2011). The sea lamprey *PSMB5* sequence was used as an outgroup. Bootstrap percentages based on 500 replicates, are shown on each branch. The 31st residue of the mature protein is shown after the common name of the animal. (a) NJ tree based on the nucleotide sequences of exons 2 to 5. (b) NJ tree based on the nucleotide sequences of exon 6. The scale bar represents 0.05 substitutions per nucleotide site.

(a)



(b)

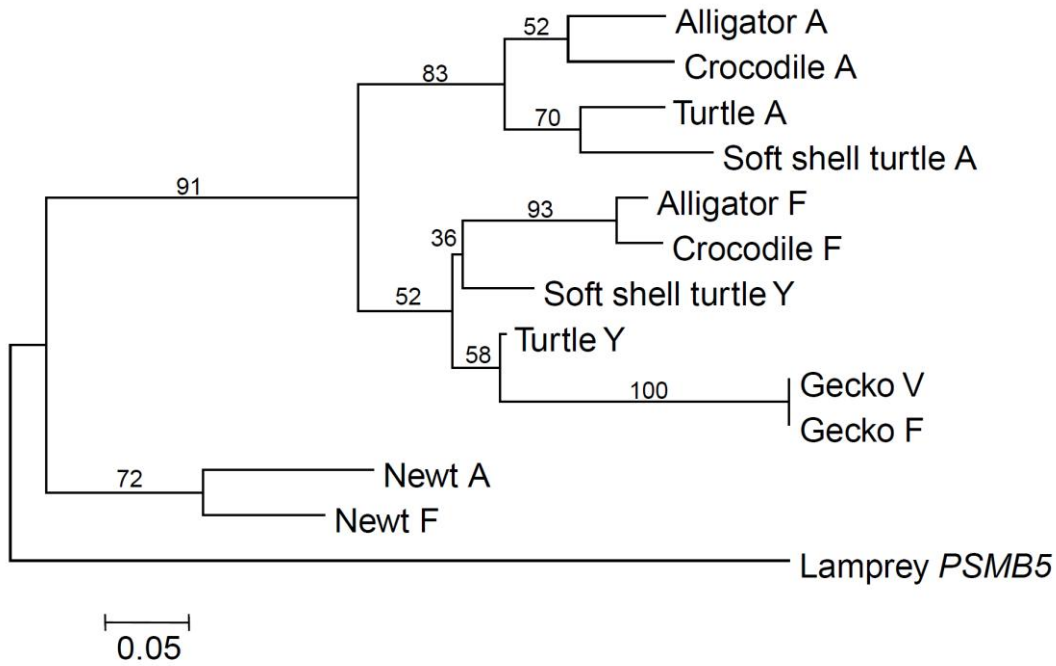
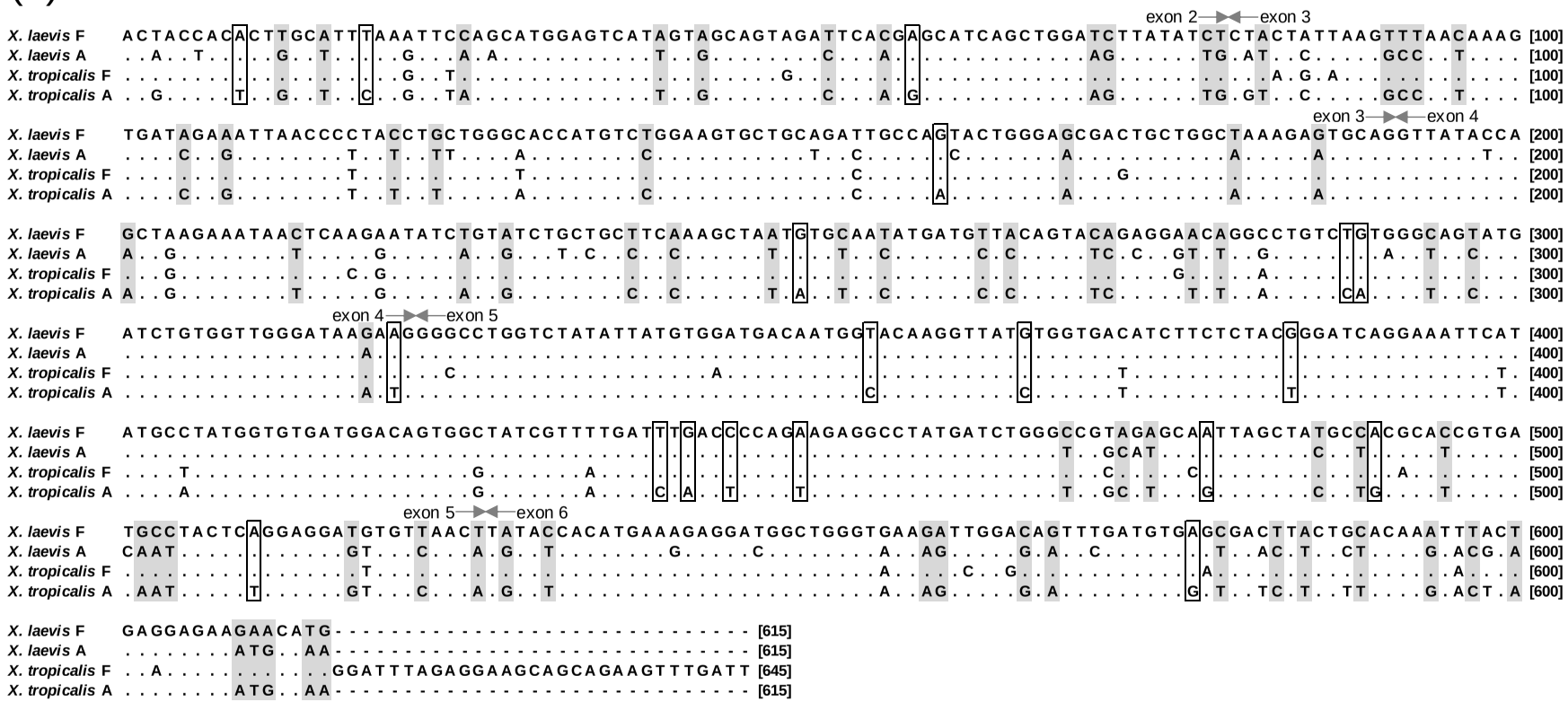


Figure 13. ML phylogenetic trees based on the nucleotide sequences of different regions of the *PSMB8* gene.

Two ML trees were constructed based on the Kimura's 2-parameter model mounted in MEGA5 (Tamura et al. 2011). The sea lamprey *PSMB5* sequence was used as an

outgroup. Bootstrap percentages based on 500 replicates, are shown on each branch. The 31st residue of the mature protein is shown after the common name of the species. (a) ML tree based on the nucleotide sequences of exons 2 to 5. (b) ML tree based on the nucleotide sequences of exon 6. The scale bar represents 0.05 substitutions per nucleotide site.

(a)


exon 2
exon 3
exon 3
exon 4
exon 4
exon 5
exon 5
exon 6
X. laevis F
X. laevis A
X. tropicalis F
X. tropicalis A

(b)

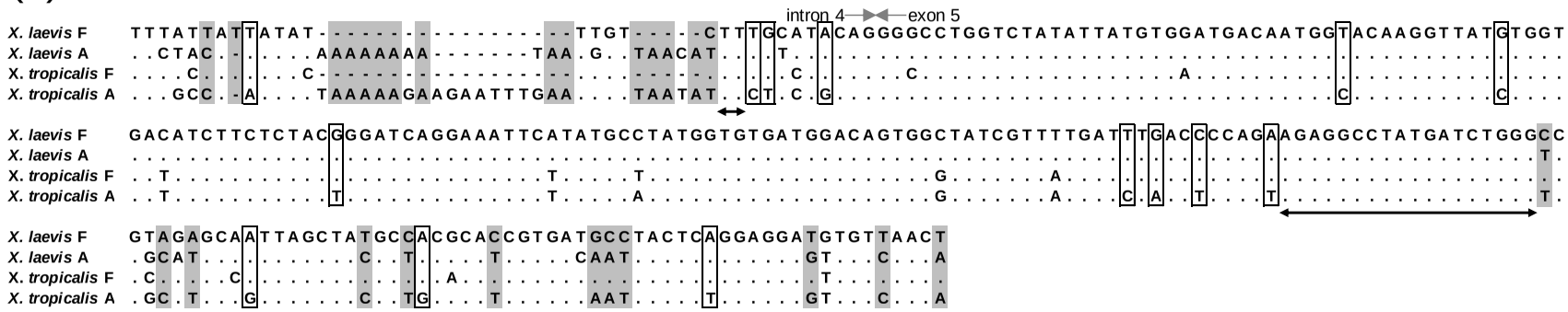

intron 4
exon 5
X. laevis F
X. laevis A
X. tropicalis F
X. tropicalis A

Figure 14

Fig. 14 Alignment of the nucleotide sequences of *PSMB8* of *X. laevis* and *X. tropicalis*

The DNA sequences of the *PSMB8* of *X. laevis* and *X. tropicalis* were aligned using MUSCLE (Edgar 2004). Names of sequences are shown by the species name followed by the single letter code for the 31st amino acid residue. Identity to the uppermost sequence is shown by dots, and gaps are shown by dashes. The locations of type-specific substitutions are shaded, and the locations of the *X. tropicalis* A specific substitutions are framed. (a) The nucleotide sequences corresponding to the PSMB8 mature protein are aligned. The ranges of the exons are indicated by the gray arrows above the alignment. (b) The nucleotide sequences around the boundary between intron 4 and exon 5 of the *PSMB8* gene are aligned. The ranges of the intron 4 and exon 5 are indicated by the gray arrows above the alignment and the possible boundaries of the homogenized region in *X. laevis* are indicated by black double-head arrows below the alignment. Xenbase/DDBJ accession numbers of the genomic sequences used here are as follows: *X. laevis* (*Xenopus laevis*) *PSMB8*, XB-GENE-479068 and AB830615; *X. tropicalis* (*Xenopus tropicalis*) *PSMB8*, XB-GENE-479064 and AB830614.

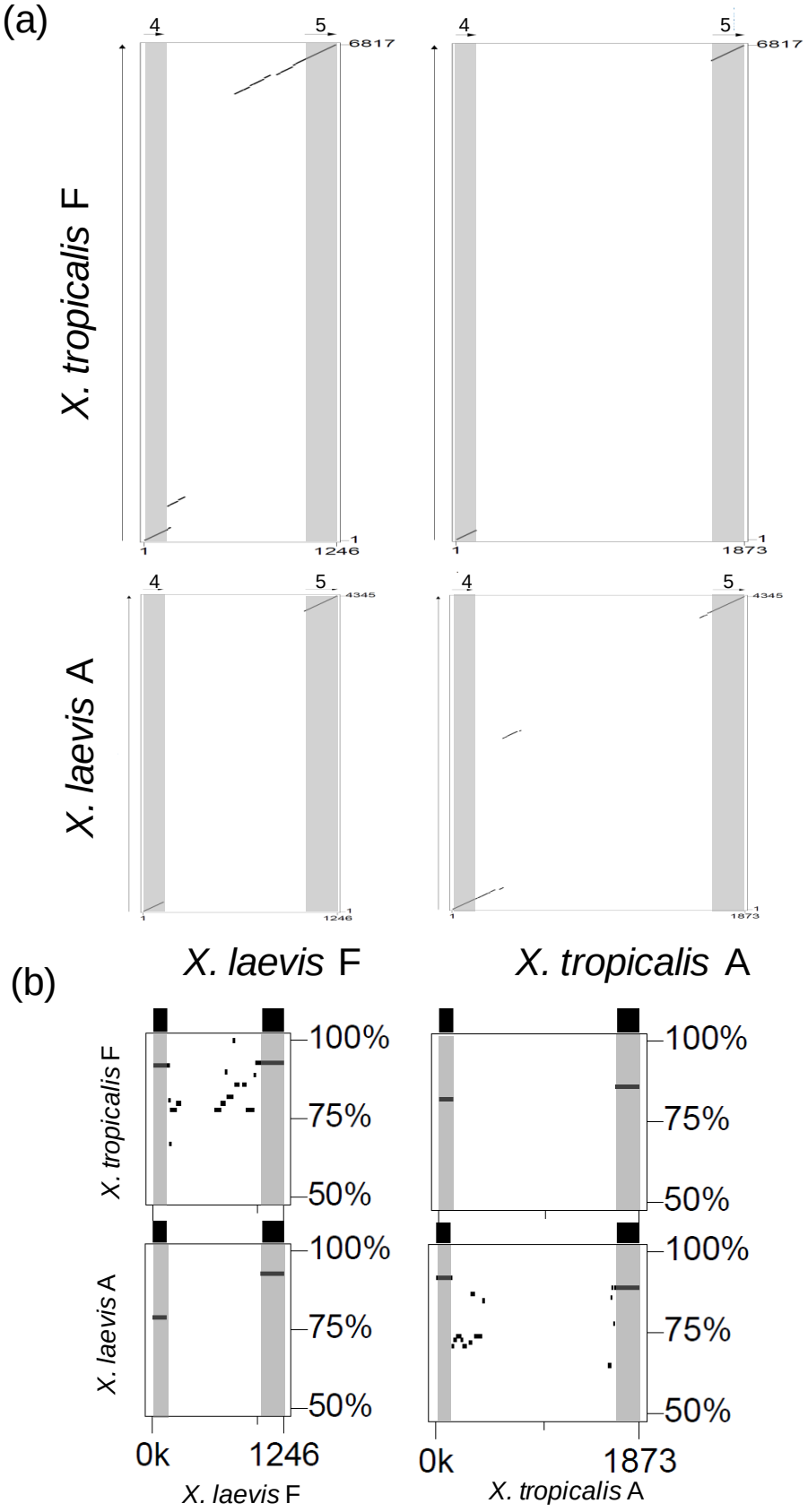


Figure 15. Dot plot and percent identity plot comparison between two *PSMB8* dichotomous alleles of *X. laevis* and *X. tropicalis*.

The sequences from exon 4 to exon 5 of the *PSMB8* dichotomous alleles in *X. laevis* were compared with that in *X. tropicalis*. (a) The horizontal axis represents the sequences of F type of *X. laevis* and A type of *X. tropicalis*. The vertical axis represents A type of *X. laevis* and F type of *X. tropicalis*. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the horizontal axis are shaded in the plot. (b) The horizontal axis represents the sequence of F type of *X. laevis* and A type of *X. tropicalis*, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type of *X. laevis* and A type of *X. tropicalis* is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

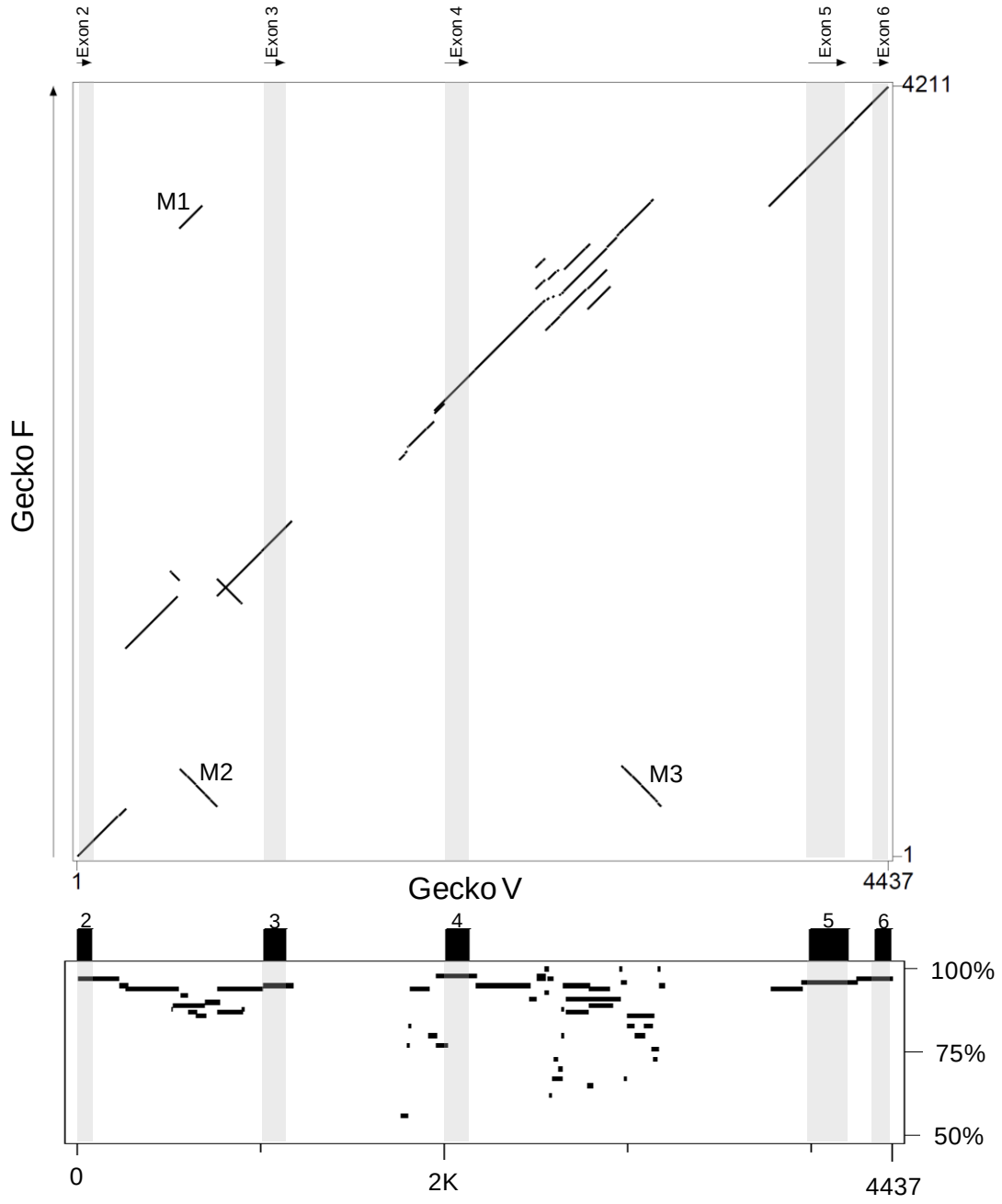


Figure 16. Dot plot and percent identity plot between *PSMB8* dichotomous alleles of gecko.

The nucleotide sequence of the A type (4437 bp) was compared with that of the F type (4211 bp, intron 3 is not completed). (a) The horizontal and vertical axes represent the sequences of A and F types in gecko, respectively. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the A type are shaded in

the plot. (b) The horizontal axis represents the sequence of A type, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. M1, M2 and M3: microsatellite-like sequence. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

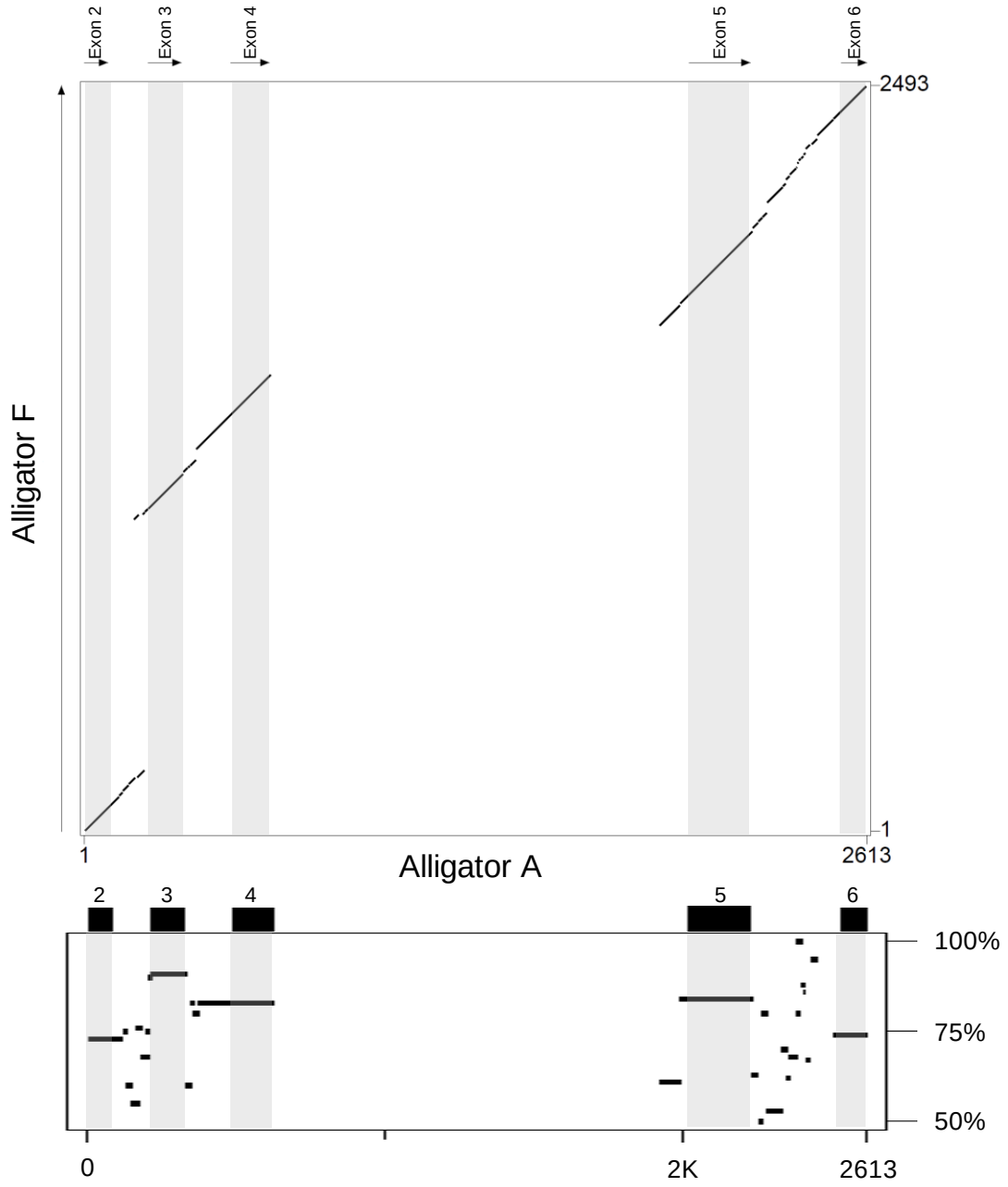


Figure 17. Dot plot and percent identity plot between *PSMB8* dichotomous alleles of alligator.

The nucleotide sequence of the A type (2613 bp) was compared with that of the F type (2493 bp). (a) The horizontal and vertical axes represent the sequences of A and F types in alligator, respectively. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the A type are shaded in the plot. (b) The horizontal axis represents the sequence of A type, the vertical axis is the percent

nucleotide identity. Each gap-free segment aligned with F type is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

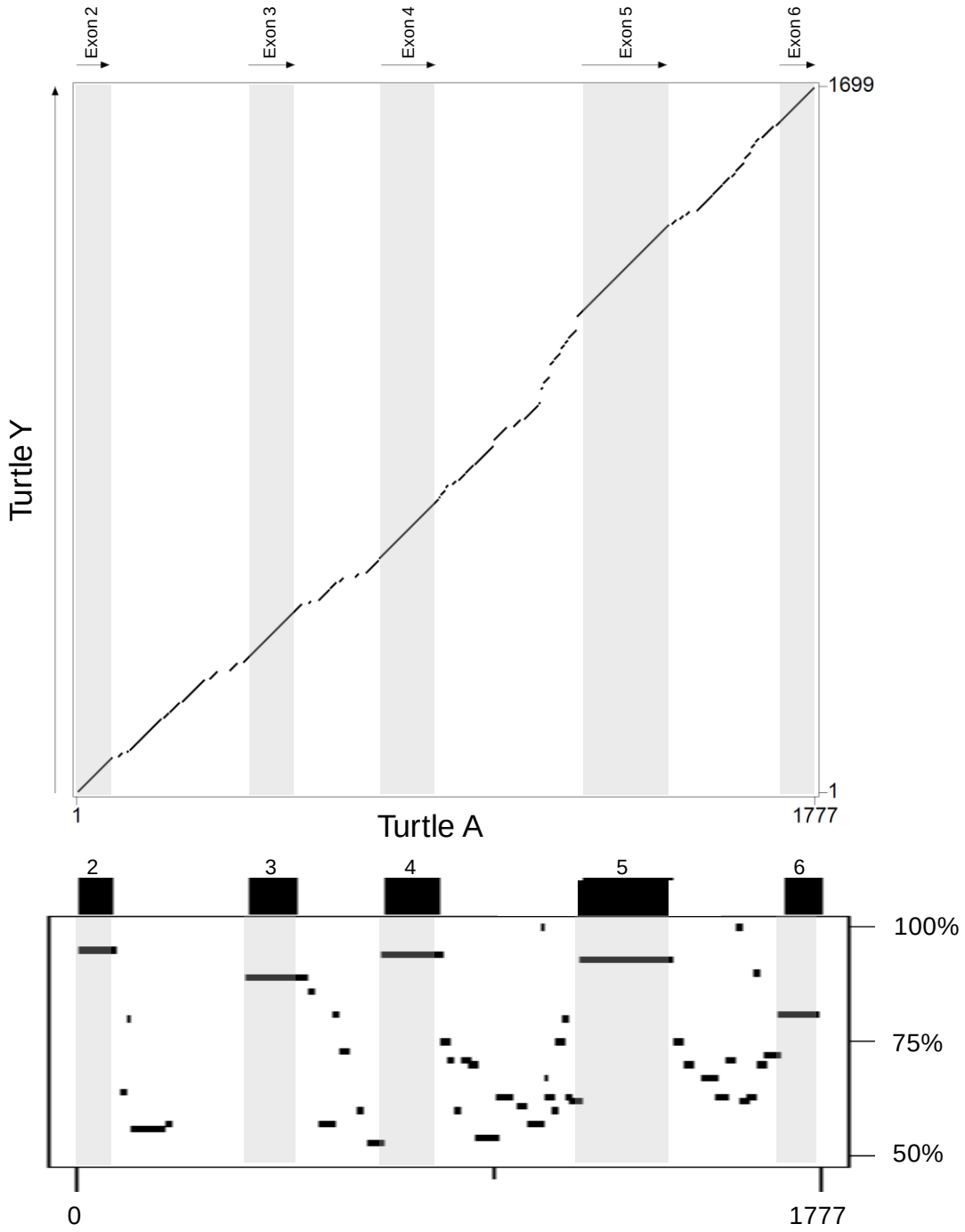
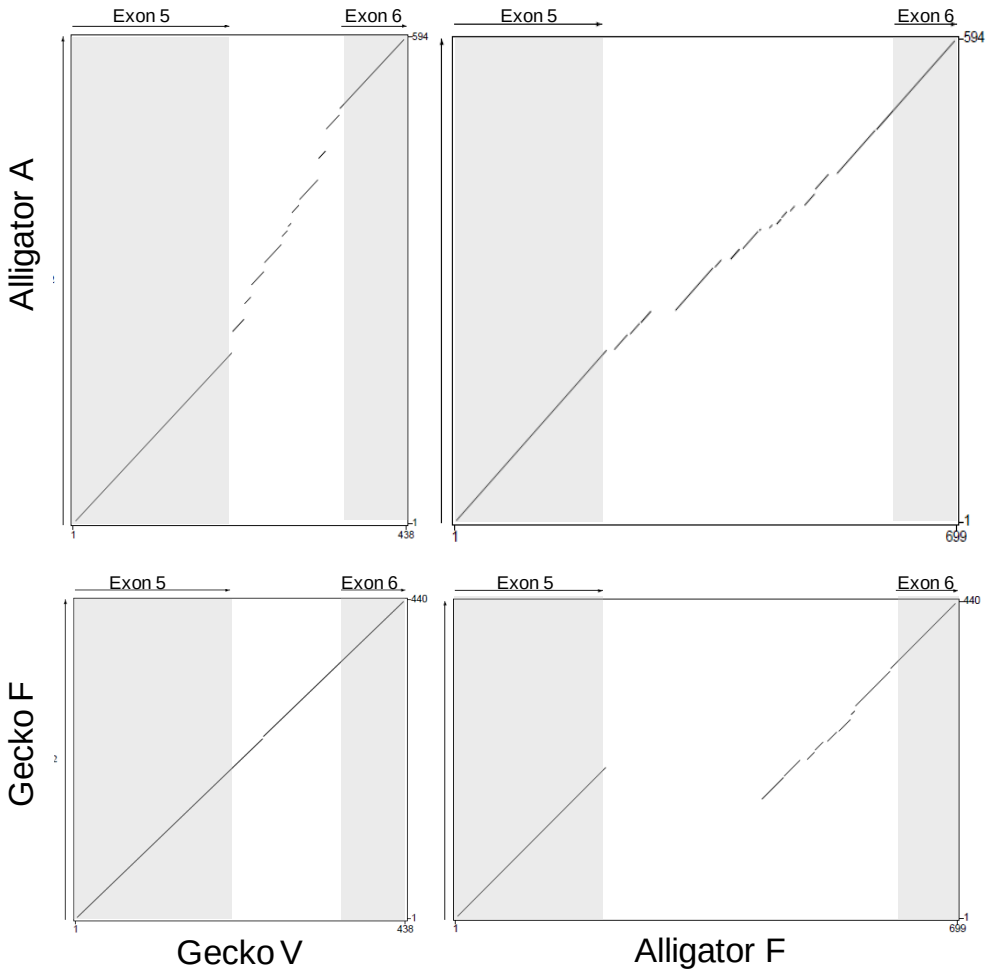


Figure 18. Dot plot and percent identity plot between *PSMB8* dichotomous alleles of turtle.

The nucleotide sequence of the A type (1777 bp) was compared with that of the F type (1699 bp). (a) The horizontal and vertical axes represent the sequences of A and F types in turtle, respectively. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the A type are shaded in the plot. (b) The

horizontal axis represents the sequence of A type, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

(a)



(b)

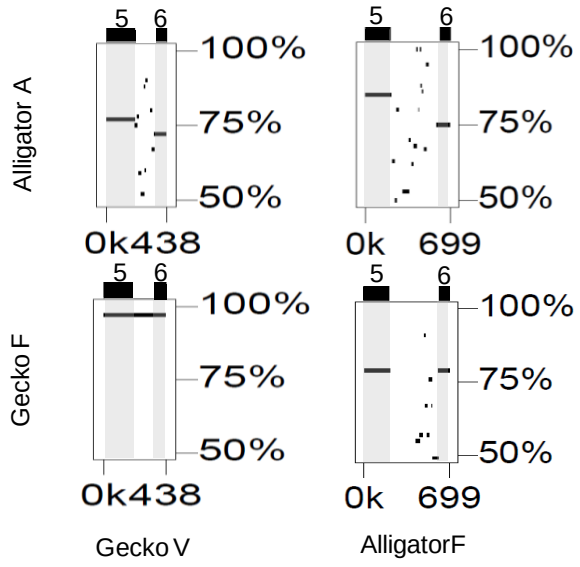


Figure 19. Dot plot and percent identity plot comparison between two *PSMB8* dichotomous alleles of gecko and alligator.

The sequences from exon 5 to exon 6 of the *PSMB8* dichotomous alleles in gecko were compared with that in alligator. (a) The horizontal axis represents the sequences of A type in gecko and F type in alligator. The vertical axis represents F type in gecko and A type in alligator. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the horizontal axis are shaded in the plot. (b) The horizontal axis represents the sequence of A type of gecko and F type of alligator, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type in gecko and A type in alligator is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

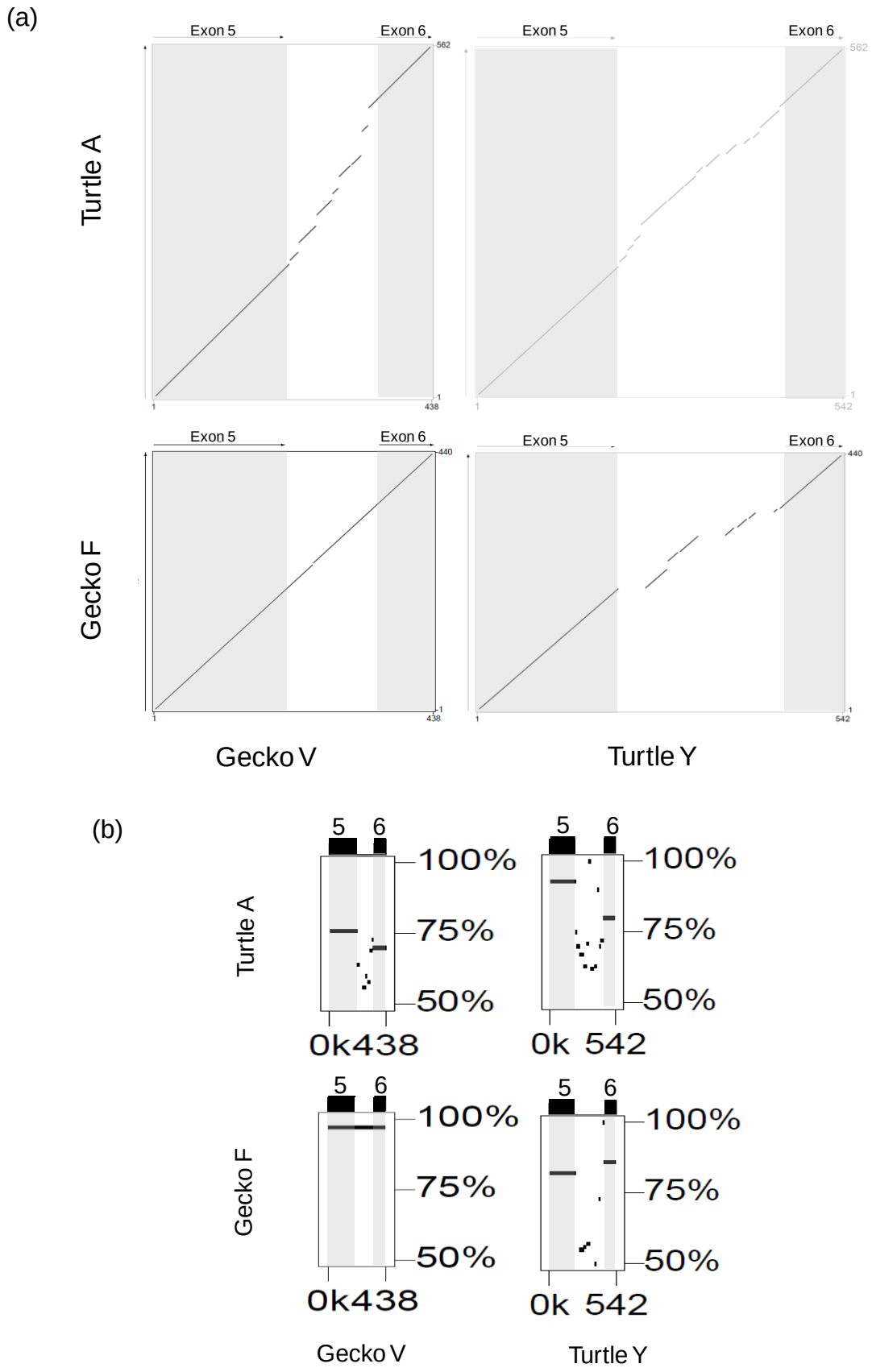


Figure 20. Dot plot and percent identity plot comparison between two *PSMB8* dichotomous alleles of gecko and turtle.

The sequences from exon 5 to exon 6 of the *PSMB8* dichotomous alleles in gecko were compared with that in turtle. (a) The horizontal axis represents the sequences of A type in gecko and F type in turtle. The vertical axis represents F type in gecko and A type in turtle. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the horizontal axis are shaded in the plot. (b) The horizontal axis represents the sequence of A type of gecko and F type of turtle, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type in gecko and A type in turtle is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

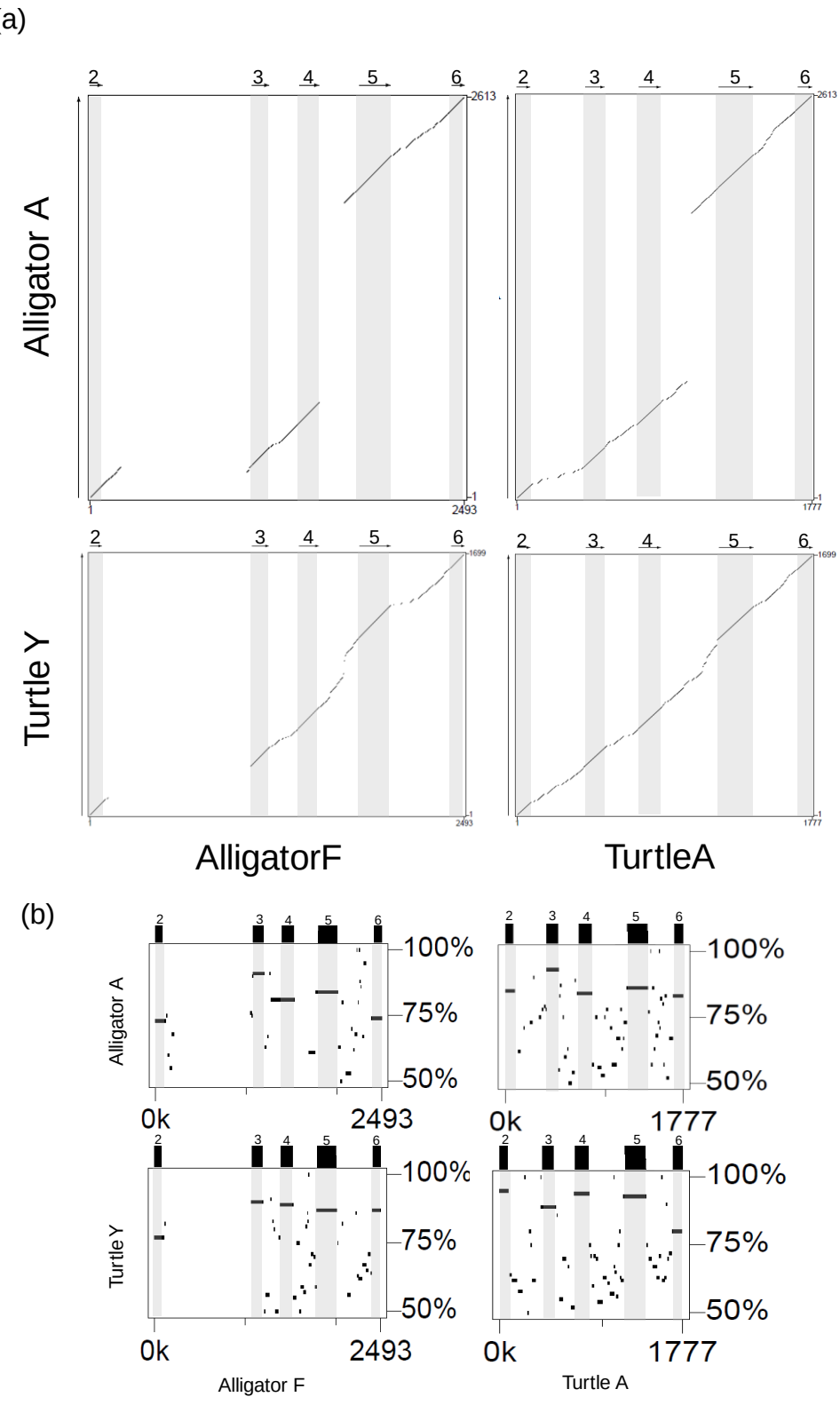


Figure 21. Dot plot and percent identity plot comparison between two *PSMB8* dichotomous alleles of alligator and turtle.

The sequence the *PSMB8* dichotomous alleles in alligator were compared with that in

turtle. (a) The horizontal axis represents the sequences of A type in alligator and F type in turtle. The vertical axis represents F type in alligator and A type in turtle. Each aligning gap-free segment with more than 50% identity is plotted. The exon positions of the horizontal axis are shaded in the plot. (b) The horizontal axis represents the sequence of A type of alligator and F type in turtle, the vertical axis is the percent nucleotide identity. Each gap-free segment aligned with F type in alligator and A type in turtle is shown by horizontal lines and plotted according to their percent identities. The black box represents exon position shaded in the plots. An alphabet after the animal name is a single letter code for the 31st amino acid residue.

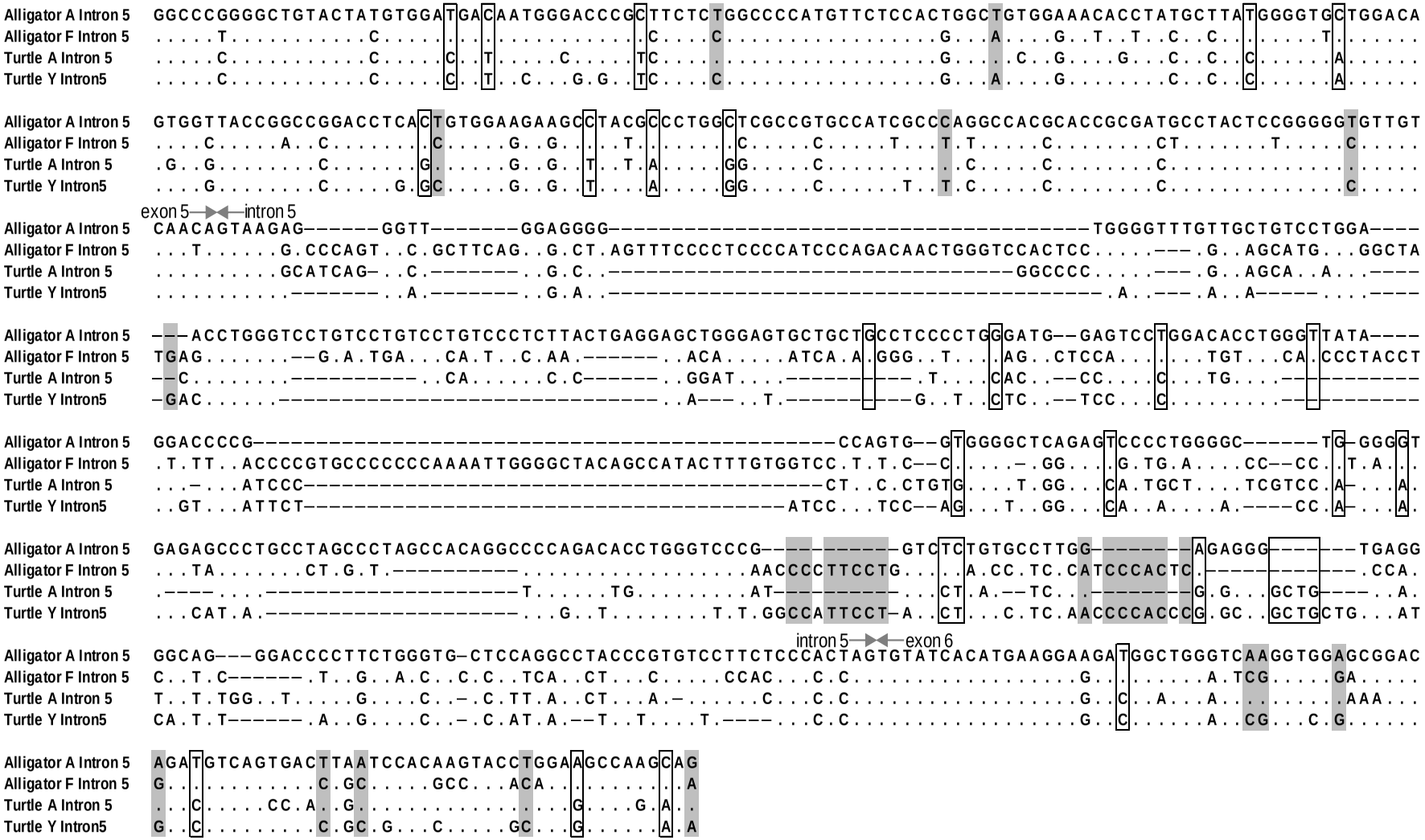

Alligator A Intron 5
Alligator F Intron 5
Turtle A Intron 5
Turtle Y Intron5
exon 5
intron 5
intron 5
exon 6

Figure 22

Figure 22. Alignment of the *PSMB8* intron 5 sequences of alligator and turtle.

The DNA sequences of the *PSMB8* of alligator and turtle were aligned using MUSCLE (Edgar 2004). Names of sequences are shown by the animal name followed by the single letter code for the 31st amino acid residue. Identity to the uppermost sequence is shown by dots, and gaps are shown by dashes. The locations of type-specific substitutions are shaded, and the locations of the taxon-specific substitutions are framed. The nucleotide sequences around the boundary between exon 5 and exon 6 of the *PSMB8* gene are aligned. The ranges of the exon 5, intron 5 and exon 6 are indicated by the gray arrows above the alignment.

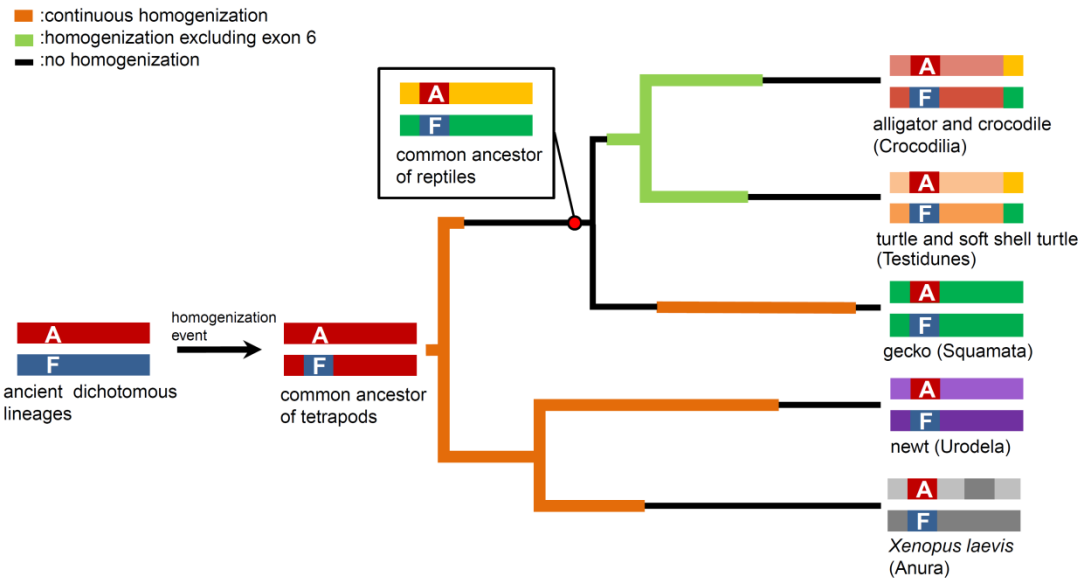


Figure 23. The homogenization-based evolutionary scenario of dichotomous *PSMB8* alleles in ectothermic tetrapods.

Each part of the *PSMB8* sequences of the extant 10 alleles shown at the right is colored so that the closest sequences are in the same color with different brightness. The original dichotomous lineages are shown in red and blue, and the 31st residues believed to have conserved the original sequences throughout evolution are shown in these two colors. The colors of branches of the tree indicate the following situations: black, no sequence homogenization; orange, continuous homogenization; green, continuous homogenization excluding exon 6. The red point indicates the last common ancestor of reptiles which had sufficiently diverged A and F alleles. The difference of the color brightness between A and F alleles in each clade indicated the level of the divergence after the last sequence homogenization. The type-specific exon 6 in reptile and homogenized region in *X. laevis* are also indicated.

Acknowledgement

I would like to express my deepest appreciation to Dr. Masaru Nonaka for his patient guidance, continual encouragement and critical suggestions during the whole course of my study.

I would also like to thank Dr. Mayumi Nonaka for her support and valuable advices**,** especially at my experimental part of this thesis. I wish to thank Dr. Masanori Taira for kindly providing *Xenopus tropicalis* DNA samples. I also need to thank Dr. Masaki Takechi for kindly providing liver sample of Australia lungfish. I am cordial grateful to Dr. Naoko T. Fujito and for kindly instruction of experimental skills and important comments for my project. I would like to thank Dr. Hidemi Bannai for useful suggestions for my academic presentation.

I must thank Dr. Hisayoshi Nozaki, Dr. Naruya Saitou, Dr. Fumio Tajima, and Dr. Rei Ueshima for their valuable comments and suggestions for this research. Thanks are also to all the colleague of out laboratory for daily encouragement. My family and friends deserve special thanks for supporting me in various ways. This work was supported in part by a Grant-in-Aid for Scientific Research (B), MEXT, Japan (to M.N.).

References

Agarwal AK, Xing C, DeMartino GN, Mizrachi D, Hernandez MD, Sousa AB, Martinez de Villarreal L, dos Santos HG, Garg A (2010) *PSMB8* encoding the β5i proteasome subunit is mutated in joint contractures, muscle atrophy, microcytic anemia, and panniculitis-induced lipodystrophy syndrome. Am J Hum Genet 87:866–872

Akiyama K, Yokota K, Kagawa S, Shimbara N, Tamura T, Akioka H, Nothwang HG, Noda C, Tanaka K, Ichihara A (1994) cDNA cloning and interferon gamma down-regulation of proteasomal subunits X and Y. Science 265:1231–1234

Arima K, Kinoshita A, Mishima H, et al (2011) Proteasome assembly defect due to a proteasome subunit beta type 8 (PSMB8) mutation causes the autoinflammatory disorder, Nakajo-Nishimura syndrome. Proc Natl Acad Sci USA 108: 14914–14919

Blanchard N, Shastri N (2008) Coping with loss of perfection in the MHC class I peptide repertoire. Curr Opin Immunol 20:82–88

Bly JE, Clem WL. (1992) Temperature and teleost immune function. Fish. Shellfish Immunol 2:159–171

Dalloul RA, Long JA, Zimin AV, et al (2010) Multiplatform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly

and analysis. PLoS Biol 8:e1000475

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Fehling HJ, Swat W, Laplace C, Kühn R, Rajewsky K, Müller U, von Boehmer H (1994) MHC class I expression in mice lacking the proteasome subunit LMP-7. Science 265:1234–1237

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Früh K, Gossen M, Wang K, Bujard H, Peterson PA, Yang Y (1994) Displacement of housekeeping proteasome subunits by MHC encoded LMPs: a newly discovered mechanism for modulating the multicatalytic proteinase complex. EMBO J 13:3236–3244

Fujito NT, Nonaka M (2012) Highly divergent dimorphic alleles of the proteasome subunit beta type-8 (*PSMB8*) gene of the bichir *Polypterus senegalus*: implication for evolution of the PSMB8 gene of jawed vertebrates. Immunogenetics 64:447–453

Gambón Deza F, Sánchez Espinel C, Magadán Mompó S (2009) The immunoglobulin heavy chain locus in the reptile *Anolis carolinensis*. Mol Immunol 46(8–9):1679–1687

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

Hao J, Li YW, Xie MQ, Li AX (2012) Molecular cloning, recombinant expression and antibacterial activity analysis of hepcidin from Simensis crocodile (*Crocodylus siamensis*). Comp Biochem Physiol B Biochem Mol Biol 163(3-4):309–315

Hillier LW, Miller W, Birney E et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

Kasahara M (1997) New insights into the genomic organization and origin of the major histocompatibility complex: role of chromosomal (genome) duplication in the emergence of the adaptive immune system. Hereditas 127:59–65

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kitamura A, Maekawa Y, Uehara H et al (2011) A mutation in the immunoproteasome subunit PSMB8 causes autoinflammation and lipodystrophy in humans. J Clin Invest 121:4150–4160

Li X, Zhu B, Chen N, Hu H, Chen J, Zhang X, Li J, Fang W. (2011) Molecular

characterization and functional analysis of MyD88 in Chinese soft-shelled turtle *Trionyx sinensis*. Fish Shellfish Immunol 30(1):33–38

Miura F, Tsukamoto K, Mehta RB, Naruse K, Magtoon W, Nonaka M (2010) Transspecies dimorphic allelic lineages of the proteasome subunit beta-type 8 gene (*PSMB8*) in the teleost genus *Oryzias*. Proc Natl Acad Sci USA 107:21599–21604

Muñoz FJ, Galván A, Lerma M, De la Fuente M. (2000) Seasonal changes in peripheral blood leukocyte functions of the turtle *Mauremys caspica* and their relationship with corticosterone, 17-β-estradiol and testosterone serum levels. Vet Immunol Immunopathol 77(1-2):27–42

Muñoz, FJ, and De la Fuente M. (2004) Seasonal changes in lymphoid distribution of the turtle *Mauremys caspica*. Copeia 1: 178–183.

Nathan JA, Spinnenhirn V, Schmidtke G, Basler M, Groettrup M, Goldberg AL.(2013) Immuno- and constitutive proteasomes do not differ in their abilities to degrade ubiquitinated proteins. Cell 152(5):1184–1194

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford: Oxford University Press

Nonaka M, Yamada-Namikawa C, Flajnik MF, Du Pasquier L (2000) Trans-species polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus, *Xenopus*. Immunogenetics 51:186–192

Ohta Y, McKinney EC, Criscitiello MF, Flajnik MF (2002) Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark *Ginglymostoma cirratum*: evidence for a stable class I region and MHC haplotype lineages. J Immunol 168:771–781

Ohta Y, Powis SJ, Lohr RL, Nonaka M, Du Pasquier L, Flajnik MF (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. Eur J Immunol 33: 3017–3027

Rock KL, Goldberg AL (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. Annu Rev Immunol 17:739–779

Saitou N, Nei M (1987) The neighbor-joining method—a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. (2000) PipMaker-a web server for aligning two genomic DNA sequences. Genome Res 10(4):577–586.

Seifert U, Bialy LP, Ebstein F, Bech-Otschir D, Voigt A, Schröter F, Prozorovski T, Lange N, Steffen J, Rieger M, Kuckelkorn U, Aktas O, Kloetzel PM, Krüger E. (2010) Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. Cell 42(4):613–624

Shen XX, Liang D, Wen JZ, Zhang P (2011) Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles. Mol Biol Evol 28:3237–3252

Slack KE, Jones CM, Ando T, Harrison GL, Fordyce RE, Arnason U, Penny D (2006) Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol Biol Evol 23(6):1144–1155

Sutoh Y, Kondo M, Ohta Y, Ota T, Tomaru U, Flajnik MF, Kasahara M (2012) Comparative genomic analysis of the proteasome β5t subunit gene: implications for the origin and evolution of thymoproteasomes. Immunogenetics 64(1):49–58

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 28:2731–2739

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

Tsukamoto K, Hayashi S, Matsuo MY, Nonaka MI, Kondo M, Shima A, Asakawa S, Shimizu N, Nonaka M (2005) Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. Immunogenetics 57:420–431

Tsukamoto K, Miura F, Fujito NT, Yoshizaki G, Nonaka M (2012) Long-lived dichotomous lineages of the proteasome subunit beta type 8 (*PSMB8*) gene surviving more than 500 million years as alleles or paralogs. Mol Biol Evol 29:3071–3079

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 A resolution. Structure 10:609–618

Warren WC, Clayton DF, Ellegren H, et al (2010) The genome of a songbird. Nature 464:757–762

Yamanoue Y, Miya M, Doi H, Mabuchi K, Sakai H, Nishida M (2011) Multiple invasions into freshwater by pufferfishes (teleostei: Tetraodontidae): a mitogenomic perspective. PLoS One 6:e17410

Zhao H, Gan TX, Liu XD, Jin Y, Lee WH, Shen JH, Zhang Y (2008) Identification and characterization of novel reptile cathelicidins from elapid snakes. Peptides 29(10):1685–1691