Transspecies dimorphic allelic lineages of the proteasome subunit β-type 8 gene (*PSMB8*) in the teleost genus *Oryzias*

Fumi Miura[a,1], Kentaro Tsukamoto[a,1,2], Ratnesh Bhai Mehta[a], Kiyoshi Naruse[b], Wichian Magtoon[c], and Masaru Nonaka[a,3]

[a]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan; [b]Laboratory of Bioresources, National Institute for Basic Biology, Aichi 444-8585, Japan; and [c]Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand

Edited* by Masatoshi Nei, Pennsylvania State University, University Park, PA, and approved November 5, 2010 (received for review August 31, 2010)

The proteasome subunit β-type 8 (*PSMB8*) gene in the jawed vertebrate MHC genomic region encodes a catalytic subunit of the immunoproteasome involved in the generation of peptides to be presented by the MHC class I molecules. A teleost, the medaka (*Oryzias latipes*), has highly diverged dimorphic allelic lineages of the *PSMB8* gene with only about 80% amino acid identity, termed "*PSMB8d*" and "*PSMB8N*," which have been retained by most wild populations analyzed. To elucidate the evolutionary origin of these two allelic lineages, seven species of the genus *Oryzias* were analyzed for their *PSMB8* allelic sequences using a large number of individuals from wild populations. All the *PSMB8* alleles of these species were classified into one of these two allelic lineages based on their nucleotide sequences of exons and introns, indicating that the *Oryzias PSMB8* gene has a truly dichotomous allelic lineage. Retention of both allelic lineages was confirmed except for one species. The *PSMB8d* lineage showed a higher frequency than the *PSMB8N* lineage in all seven species. The two allelic lineages showed curious substitutions at the 31st and 53rd residues of the mature peptide, probably involved in formation of the S1 pocket, suggesting that these allelic lineages show a functional difference in cleavage specificity. These results indicate that the *PSMB8* dimorphism was established before speciation within the genus *Oryzias* and has been maintained for more than 30–60 million years under a strict and asymmetric balancing selection through several speciation events.

long-term balancing selection | antigen processing | transspecies polymorphism

Transspecies polymorphism (TSP) is the passage of allelic lineages from ancestral to descendent species (1), and the underlying selective mechanism is referred to as balancing selection caused by "overdominant selection," "frequency-dependent selection," or "selection that varies in time and space" (2). The vertebrate MHC class I and II genes with a large number of alleles and wide allelic differences provide a classical example of TSP (3, 4), most probably arising by overdominant selection (5). TSP of the MHC genes sometimes persists for very long periods. Certain HLA-DRB alleles are reported to have persisted for 50–60 million years (6). Another well-characterized TSP has been reported for the rabbit Ig heavy-chain variable-region genes; its persistence has been estimated at 50 million years based on molecular phylogenetic analysis (7, 8). Except for these genes encoding molecules showing high degree of binding specificity, information for long-term balancing selection resulting in TSP is limited (9–11). Moreover, all TSP reported thus far involve only a few species, and the possible presence of a long-lasting TSP surviving through many speciation events is still to be clarified by systematic phylogenetic analysis.

The MHC class I molecules deliver the peptides derived from cytosolic proteins to the cell surface for recognition by cytotoxic T cells. The proteins are degraded proteolytically into short peptides by proteasomes whose catalytic core, or 20S 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 (12, 13). Of these subunits, only three β subunits—PSMB5, PSMB6, and PSMB7—have proteolytic activity with chymotrypsin-like, caspase-like, and trypsin-like specificity, respectively (14). Immunoproteasomes are formed by replacing PSMB5, PSMB6, and PSMB7 with the IFN-γ–inducible β subunits, PSMB8, PSMB9, and PSMB10, respectively (15, 16). These subunit substitutions enhance the chymotrypsin-like activity of the immunoproteasomes, which are responsible for generation of peptides with a hydrophobic residue at the C terminus suitable for binding to MHC class I molecules. In particular, PSMB8, with its chymotrypsin-like activity, is a critical component for supplying MHC class I-binding peptides, because *PSMB8*-knockout mice show reduced expression of MHC class I molecules on the cell surface (17).

The Japanese population of medaka, *Oryzias latipes*, is divided into two subpopulations, the Northern Population (NP) and the Southern Population (SP), which diverged 5–18 Mya (18, 19). Using BAC clones, we determined the complete nucleotide sequences of the approximately 400-kb MHC class I region of two inbred strains, HNI (derived from the NP) and Hd-rR (derived from the SP) (20, 21). Although the order and transcriptional orientation of the 22 identified genes were conserved perfectly between these two strains, the nucleotide sequences showed an extremely high degree of divergence at the ∼100 kb segment in the middle of the medaka MHC class I region harboring the two MHC class IA genes, *Orla-UAA* and *-UBA*, and the two immunoproteasome β subunit genes, *PSMB8* and *PSMB10*. The sequence divergence was especially conspicuous with the *PSMB8* gene, which showed only 81.5% deduced amino acid sequence identity between these two inbred strains. The average nucleotide divergence between these two inbred strains was estimated to be 3.4% based on the whole-genome sequences (22), and the two sides of this highly diverged segment of the MHC class I region showed similar levels of nucleotide divergence from this average (21, 23). Extensive analysis of allelic polymorphism of the *PSMB8* gene using wild individuals from nine localities representing both

Author contributions: F.M., K.T., and M.N. designed research; F.M., K.T., and R.B.M. performed research; F.M., K.T., K.N., and W.M. analyzed data; and F.M., K.T., and M.N. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AB449218–AB449245 and AB551017–AB551031).

[1]F.M. and K.T. contributed equally to this work.

[2]Present address: Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan.

[3]To whom correspondence should be addressed. E-mail: mnonaka@biol.s.u-tokyo.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012881107/-/DCSupplemental.

the NP and SP clarified that all *PSMB8* alleles were clearly classified into the dichotomous allelic lineages, *PSMB8N* or *PSMB8d* (24). Both *PSMB8d* and *PSMB8N* lineages were retained in NP and SP although the allelic frequency of the *PSMB8d* lineage (73–100%) was much higher than that of the *PSMB8N* lineage (0–27%) in all the populations we analyzed. These findings suggest that the dichotomous allelic lineages were established before the divergence between NP and SP and have been maintained in each group for 5–18 million years (18, 19).

The genus *Oryzias* contains about 20 species distributed in southeastern Asia. These species most likely were generated by allopatric speciation, and the few hybridization experiments among them performed thus far showed the presence of reproductive isolation resulting in abortive embryonic development in the case of *O. latipes* x *O. javanicus* (25) or sterile F1 males in the case of *O. latipes* x *O. celebensis* (26) and *O. latipes* x *O. curvinotus* (27). The *Oryzias* species are divided into three species groups, the *latipes*, *javanicus*, and *celebensis* groups, based on mitochondrial DNA sequences (18, 19), corresponding to the biarmed, monoarmed, and fused chromosome groups of karyological grouping. The divergence time among these three species groups was estimated as 29–32 Mya, based on the molecular clock assumption using the mitochondrial 12S and 16S rRNA sequences (18), or as 58–65 Mya, based on a Bayesian relaxed molecular clock analysis of whole-mitogenome sequences (19). On the other hand, the formation of the Makassar Strait, the possible vicariant event between the *celebensis* group and the other two groups, occurred in the Eocene (34–56 Mya) (28). All these results suggest that the speciation within genus *Oryzias* started 30–60 Mya.

In this study, to elucidate the origin and evolution of these two allelic lineages of the *PSMB8* gene found in *O. latipes*, we analyzed the genetic polymorphism of *PSMB8* in this genus using wild populations of seven species from the *latipes* (*O. curvinotus*), *javanicus* (*O. javanicus*, *O. minutillus*, and *O. dancena*), and *celebensis* (*O. celebensis*, *O. marmoratus*, and *O. matanensis*) groups.

Results

Identification of the PSMB8 Alleles in Wild Populations of Seven *Oryzias* Species. To investigate the allelic diversity of the *PSMB8* gene in wild populations, seven *Oryzias* species [*O. minutillus* (n = 244 individuals), *O. javanicus* (n = 178), *O. curvinotus* (n = 69), *O. celebensis* (n = 190), *O. matanensis* (n = 106), *O. marmoratus* (n = 106), and *O. dancena* (n = 150)] were analyzed (Fig. 1). First, the *PSMB8* gene fragment from exons 2–3 was amplified by genomic PCR. Each individual gave either single or double bands upon agarose gel electrophoresis, probably representing the homozygous or heterozygous state, respectively. Two to eight bands with distinct sizes were detected in each species. All bands amplified from all individuals were sequenced directly, and the lineage was determined based on the 31st amino acid residue of the mature peptide encoded by exon 3 (24). One homozygous individual for each distinctive band was selected for another round of PCR amplification from exon 1 to exon 6 (the last exon), and nucleotide sequences were determined by direct sequencing. The entire coding sequences for the mature peptide were elucidated by determining the 3′-most 69 bp not covered by this PCR amplification using 3′ RACE.

The deduced complete amino acid sequences of the mature peptide of the *PSMB8d* and *PSMB8N* allelic lineages identified from these eight *Oryzias* species, including *O. latipes*, were aligned, together with dimorphic *PSMB8* alleles of *Xenopus* (29), two paralogous *PSMB8* genes of sharks (30, 31), and human *PSMB8* (Fig. S1). These sequences were aligned perfectly without any insertion/deletion except for a few residues at the C terminus. Fig. 2 shows amino acid substitutions found among the *PSMB8* alleles of eight *Oryzias* species in the mature peptide region. Based on the 31st amino acid residue of the mature peptide and degree of amino acid sequence identity to the HNI

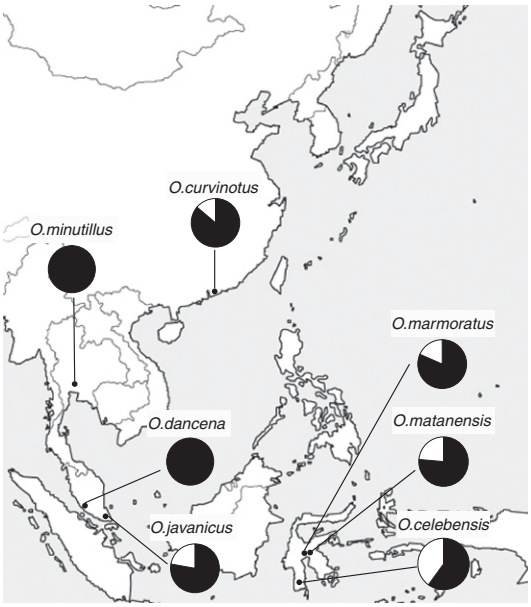

Fig. 1. Map showing collection sites of wild populations of *Oryzias* species and allelic frequencies of the *PSMB8d* and *PSMB8N* lineages in wild populations. Dots indicate the collection site of each wild population of *Oryzias* species. Allelic frequencies of the *PSMB8d* and *PSMB8N* lineages in each wild population are shown in black and white, respectively, in the circle diagrams. Actual frequencies of the *d* allele (number of *d* alleles/number of total alleles) for each species are *O. curvinotus*, 0.862 (119/138); *O. celebensis*, 0.600 (228/380); *O. matanensis*, 0.764 (162/212); *O. marmoratus*, 0.816 (173/212); *O. dancena*, 1.00 (300/300); *O. minutillus*, 1.00 (488/488); and *O. javanicus*, 0.778 (277/356).

(*PSMB8N*) and Hd-rR (*PSMB8d*) alleles of *O. latipes* (Fig. 2), these *Oryzias PSMB8* alleles clearly were classified into either the *PSMB8d* or the *PSMB8N* lineages. Both lineages were identified from wild populations of *O. javanicus*, *O. curvinotus*, *O. celebensis*, *O. matanensis*, and *O. marmoratus*. In contrast, only the *PSMB8d* lineage was detected from the wild populations of *O. minutillus* and *O. dancena*. However, *O. dancena* has the *PSMB8N* lineage at the species level, because a *PSMB8N* lineage gene coding for tyrosine at the 31st position (Tyr31) has been isolated from a BAC library constructed using individuals from the closed colony kept at the University of Tokyo (32). This *O. dancena PSMB8N* sequence was added to Fig. 2 and Fig. S1 and was used in the following analyses. There are two sublineages, *PSMB8d(V)* and *PSMB8d(A)*, in the*PSMB8d* lineage of medaka, having valine and alanine, respectively, at the 31st amino acid position, respectively (24). These two sublineages were present in all the wild populations of *Oryzias* species we analyzed. Names for the PSMB8 alleles of the *Oryzias* species were designated as follows: the species name; lineage name *d* or *N*; sublineage name *V* or *A* in parenthesis in the case of the *d* lineage, and the number if there are multiple alleles in that lineage or sublineage of each species. In the coding region, the *PSMB8d(V)* alleles of these species showed 92.7–98.5% nucleotide identity and 96.6–99.6% amino acid identity to the Hd-rR allele [*O. latid(V)* in Fig. 2 and Fig. S1], and the *PSMB8d(A)* alleles of these species showed 93.0–95.6% nucleotide identity and 94.0–97.5% amino acid identity to the Hd-rR allele. The *PSMB8N* alleles of these species showed 90.1–94.8% nucleotide identity and 93.6–96.1% amino acid identity to the HNI allele (*O. latiN* in Fig. 2 and Fig. S1). Because the nucleotide and amino acid identities between the coding regions of the Hd-rR and HNI *PSMB8* genes were only 80.3% and 81.5%, respectively (21), there was no ambiguity in allocating the analyzed sequences into the *PSMB8d* or *PSMB8N* lineage.

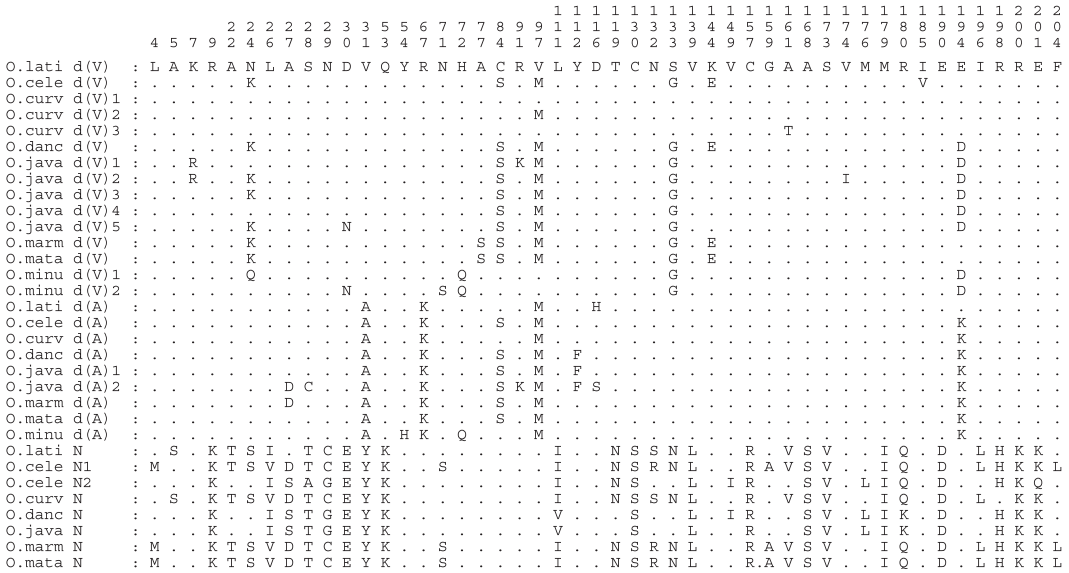

Fig. 2. Comparison of the amino acid sequences of the mature peptides of *PSMB8*. *PSMB8* mature peptides of *Oryzias* species and two allelic lineages of *O. latipes* (NCBI accession nos. AB183488 and BA000027) were aligned with ClustalX 2.0. All *Oryzias PSMB8* sequences were determined in this study, except for two *O. latipes PSMB8N* and *PSMB8d(V)* sequences and the *O. dancena PSMB8N* sequence (NCBI accession no. FJ481084). Of the 204 positions of the mature peptides, only the 48 positions where amino acid substitutions were observed are shown. Dots indicate identity with the residues in the uppermost sequence.

Phylogenetic Analysis of the *Oryzias PSMB8* Alleles. To clarify the phylogenetic relationship of the *PSMB8* alleles from these *Oryzias* species, phylogenetic analysis was performed using the nucleotide sequences of the mature peptide region. When the phylogenetic trees were constructed using the neighbor-joining (NJ) method (33) with the PSMB8 sequences of spotted green pufferfish, human, and mouse as an outgroup, *Oryzias PSMB8* alleles were divided into the *PSMB8d* and *PSMB8N* clades supported by 100% bootstrap percentage (Fig. 3). Both clades contained the *PSMB8* alleles from the *latipes*, *javanicus*, and *celebensis* groups, indicating that the divergence between these two *PSMB8* lineages occurred before separation of the *Oryzias* species. The *PSMB8d(V)* and *PSMB8d(A)* sublineages did not form their respective subclades and intermingled with each other. On the other hand, the *PSMB8N* clade was divided into two subclades supported by high bootstrap percentages of 96% and 93%. However, one subclade contained only the *javanicus* and *celebensis* groups, and the other contained only the *latipes* and *celebensis* groups. When the amino acid sequences of the mature peptides were used to construct a NJ tree, the dichotomous *PSMB8d* and *PSMB8N* clades were reproduced with high bootstrap percentages of 100% and 99%, respectively (Fig. S2). However, the topology within the *PSMB8d* clade showed a significant divergence from that of the NJ tree based on nucleotide sequences. It is especially noteworthy that the *PSMB8d (A)* sublineage formed a clade, although with a low bootstrap percentage of 65%.

Polymorphism of Deduced Amino Acid Sequences Between Two *PSMB8* Lineages. The *Oryzias PSMB8d* and *PSMB8N* lineages were clearly discriminated by 10 diagnostic substitutions: $R^{8}K$, $Q^{53}K$, $C^{130}S$, $V^{139}L$, $C^{157}R$, $A^{168}S$, $S^{173}V$, $M^{179}I$, $E^{190}D$, and $R^{200}K$ (Fig. 2). In addition, these two lineages show substitutions in five other positions, although two different amino acids are found in one of the lineages: $N^{29}C/G$, $D/N^{30}E$, $V/A^{31}Y$, $R^{180}Q/K$, and $E^{201}Q/K$ (Fig. 2). Among these diagnostic substitutions, $V/A^{31}Y$ and $Q^{53}K$ are especially interesting because these residues are likely to be involved in the formation of the S1 pocket that determines cleaving specificity (13). The other residues involved in S1 pocket shaping—A^{20}, I^{35}, M^{45}, and A^{49}—were conserved between these two lineages. *Xenopus* dimorphic alleles and two shark paralogous copies of the *PSMB8* gene also show a similar substitution at the 31st position (Fig. S1) (29, 30). On the other hand, it is not clear whether there is functional diversification between *PSMB8d(V)* and *PSMB8d(A)*, because sublineage-specific substitutions were recognized only at two positions, $V^{31}A$ and $R^{67}K$ (Fig. 2). The valine/alanine substitution at the 31st residue is a relatively conservative one, and no direct functional importance can be assigned to the 67th position. The two sublineages of the *PSMB8N* lineage show seven sublineage-specific substitutions, $A^{22}T$, $N^{24}S$, $G^{29}C$, $N^{133}S$, $V^{161}A$, $M^{176}L$, and $L^{196}I$ (Fig.2). However, the functional significance of these substitutions, if any, is still to be clarified.

Comparison of Nucleotide Sequences of Exons and Introns Among the *PSMB8* Alleles. The sizes of exons 2–6 were conserved completely in all the analyzed *PSMB8* genes of these eight *Oryzias* species. In contrast, intronic sequences showed interspecific as well as intraspecific variation in length. The nucleotide sequences of these *PSMB8* alleles were compared with those of the *PSMB8d(V)* and *PSMB8N* of *O. latipes* by dot-plot analysis (Fig. S3). One *PSMB8d (V)* allele, one *PSMB8d(A)* allele, and one *PSMB8N* allele were selected from *O. curvinotus*, representing the *latipes* group, from *O. celebensis*, representing the *celebensis* group, and from *O. javanicus*, representing the *javanicus* group. The only exception was the *PSMB8N* of *O. javanicus*, for which PCR amplification was unsuccessful, probably because of the presence of long intron(s). As shown in Fig. S3, diagonal lines were detected in the dot plots for all exons in any pair of comparisons. In contrast, diagonal lines were detected only in certain pairs of comparisons for the intronic regions. In the comparison with *PSMB8N* of *O. latipes*, only the *PSMB8N* alleles of *O. curvinotus* and *O. celebensis* showed diagonal lines in these regions. On the other hand, in the comparison with the *PSMB8d(V)* of *O. latipes*, only the *PSMB8d(V)* alleles showed clear diagonal lines in intronic regions, and the *PSMB8d(A)* alleles showed only short intronic lines. These results indicate that the intronic sequences of the *PSMB8d* and *PSMB8N* lineages have almost no sequence similarity and that there has been considerable intronic sequence diversification even between the *PSMB8d(V)* and *PSMB8d(A)* sublineages. On the other hand,

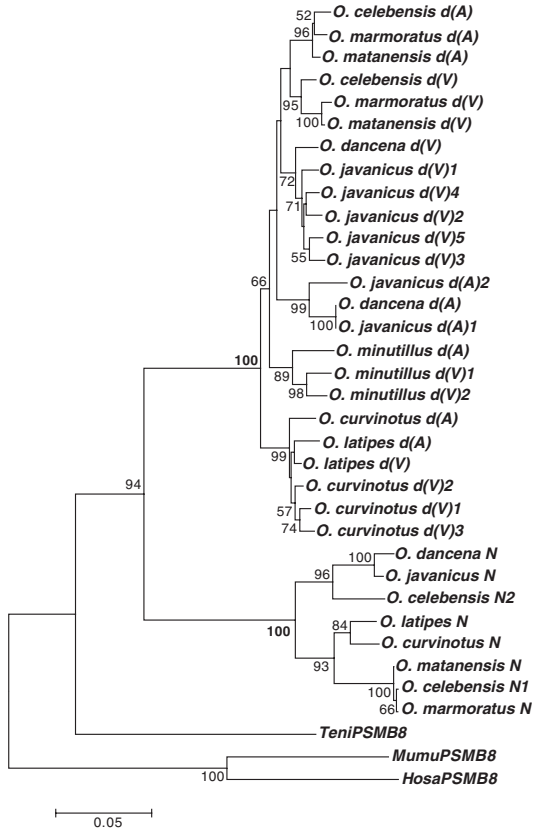

Fig. 3. Phylogenetic tree of *Oryzias PSMB8* alleles. The nucleotide sequences of mature peptides of 612 residues were aligned by ClustalX 2.0(37), and the phylogenetic tree was constructed by the NJ method (33). The numbers on each branch represent bootstrap probabilities (>50%) based on 1,000 bootstrap trials. The sequences of TeniPSMB8 (*Tetraodon nigroviridis*, NCBI accession no. CR697191), HosaPSMB8 (*Homo sapiens*, NCBI CR541661), and MumuPSMB8 (*Mus musculus*, NCBI BC013785) were used as an outgroup.

intronic sequences have been more or less conserved within the *PSMB8N* and *PSMB8d* lineages, excluding the remote possibility that the observed sequence similarity in the protein-coding region among the *PSMB8* sequences of various species belonging to the same lineage has been caused by convergent evolution rather than by common ancestry.

Frequency of Two *PSMB8* Allelic Lineages in Wild *Oryzias* Populations. The allelic frequencies of the *PSMB8d* and *PSMB8N* lineages in wild populations of each species are shown by circle diagrams in Fig. 1. The allelic frequency of the *PSMB8d* lineage (black) was higher than that of the *PSMB8N* lineage (white) in all analyzed species, ranging from 0.6 to 1.0. The results of the genotyping of each individual of these *Oryzias* species are summarized in Table S1. The observed numbers do not deviate significantly from the values expected from allelic frequency. The *PSMB8N* lineage was not identified in wild populations of *O. minutillus* and *O. dancena*. However, because the frequency of the *PSMB8N* lineage was less than 0.01 in certain populations of *O. latipes* (24), it still is unclear whether the *PSMB8N* lineage is missing or is present at an extremely low frequency in these populations.

Discussion

We found that dichotomous allelic lineages of the single-copy *PSMB8* gene (20, 21), *PSMB8d* and *PSMB8N*, are present in most of the *Oryzias* species we analyzed. Lineage-specific sequences were identified not only in the protein-coding regions (Fig. 2) but also in introns (Fig. S3), indicating that these lineages reflect real lineages established by ancient diversification rather than false lineages formed by convergent evolution. Because these two allelic lineages were found among *Oryzias* species belonging to all three species groups—*O. latipes* and *O. curvinotus* of the *latipes* group, *O. dancena* and *O. javanicus* of the *javanicus* group, and *O. celebensis*, *O. marmoratus*, and *O. matanensis* of the *celebensis* group—they should have been established before the start of speciation in the genus *Oryzias*. Thus, it is likely that these dimorphic lineages of the *Oryzias PSMB8* gene survived through the *Oryzias* speciation process and were transferred from species to species.

Long-term balancing selection maintaining multiple alleles through multiple speciation events has been reported with the MHC and other host-defense genes in vertebrates (3, 4, 7–10). The longest persistence time estimated thus far based on molecular data is 50–60 million years for the human HLA-DRB and rabbit Ig heavy-chain variable-region genes (6–8). The phylogenetic analysis of the *Oryzias PSMB8* genes (Fig. 3) indicated that the *PSMB8d* and *PSMB8N* lineages diverged long before the speciation of the *Oryzias* species, which is believed to have started 30–60 Mya (18, 19). To estimate roughly the time of divergence between these two *PSMB8* lineages, we constructed a linearized tree (Fig. S4) based on the NJ tree of Fig. S2. To make the estimation of the persistence time comparable with that of the rabbit Ig heavy-chain variable-region genes (8), the divergence time of 100 Mya between human and mouse (34) was used as a standard. The rate of amino acid substitution for the *PSMB8* gene was estimated to be $0.42 \times 10^{-9} \cdot \text{site}^{-1} \cdot \text{y}^{-1}$, and the divergence time between the two allelic lineages of the *PSMB8* gene was calculated to be 178 Mya, much more ancient than the origin of the rabbit Ig heavy-chain variable-region gene polymorphism. Based on this rate of amino acid substitution, the divergence time between spotted green pufferfish and *Oryzias* was calculated to be about 240 Mya. This value does not show great discrepancy with the recent estimation of 191 Mya based on mitochondria DNA sequences (35). This result suggests that age estimation based on this linearized tree is not unrealistic. However, the evolutionary rate of the *PSMB8N* lineage apparently is faster than that of the *PSMB8d* lineage (Fig.3 and Fig. S2), limiting the reliability of age estimation based on this linearized tree. Thus, we conclude that the two *Oryzias PSMB8* lineages diverged well before the start of *Oryzias* speciation 30–60 Mya.

Like the *Oryzias PSMB8* gene, the *Xenopus PSMB8* gene showed dichotomous allelic lineages that were transferred from ancestral to descendant species for more than 80 million years (36). However, transspecies dimorphism of the *Xenopus PSMB8* gene was inferred indirectly from Southern blotting analysis; our current analysis of the *Oryzias PSMB8* genes provides conclusive evidence for transspecies dimorphism of *PSMB8* based on the nucleotide and amino acid sequences. Interestingly, phylogenetic analysis has indicated that the origin of the *PSMB8* dimorphism of *Xenopus* and *Oryzias* species was independent (21), suggesting that the *PSMB8* dimorphism was not maintained throughout the evolution of jawed vertebrates. However, once established, the dimorphism was transferred from species to species for more than 30–60 or 80 million years. In addition to the dichotomous, highly diverged allelic lineages in *Xenopus* and *Oryzias*, two paralogous *PSMB8* genes with a similar level of sequence diversity have been reported in sharks (30). Again, the phylogenetic tree analysis of these two types of the *PSMB8* genes in these species did not show any orthologous relationship, indicating that they were generated by independent evolutionary events (21). However, the two diverged *PSMB8* types of each species, whether alleles or paralogous genes, show curious similarities in amino acid substitution pattern at certain positions. Thus, the amino acid residues at the 31st position of the mature peptide are either alanine or phenylalanine in the shark and *Xenopus* and

either valine/alanine or tryptophan in *Oryzias*. Because this position in bovine *PSMB5* is involved in the S1 pocket formation (13) and is occupied by phenylalanine or tyrosine with bulky and neutral side chains or alanine or valine with smaller side chains, it is conceivable that the two types of the *PSMB8* gene in these animals show similar differences in cleaving specificity. Because there is no orthologous relationship, the similar amino acid substitutions at the 31st position in the two types of *PSMB8* gene of each animal lineage might have been formed by convergent evolution, suggesting the presence of a strong selective pressure to induce functional diversification of the *PSMB8* gene within each species group.

In all the wild populations we analyzed here, *PSMB8d* was the major allelic lineage with a gene frequency of 0.6–1.0. This predominance also is true for wild populations of *O. latipes* in Japan and Korea, where the frequency of the *PSMB8N* lineage was extremely low (<0.01) in some populations (24). Thus, it still is not clear whether the *PSMB8N* lineage is absent from the wild populations of *O. minutillus* and *O. dancena* we analyzed or is present at an extremely low frequency. Recently, the nucleotide sequence of the MHC class I region of *O. dancena* was determined, and a BAC clone encompassing the *PSMB8N* lineage was identified (32), indicating that two allelic lineages were maintained in at least some wild populations of *O. dancena*. Conservation of this biased dimorphism of the *PSMB8* gene in *Oryzias* species for at least 30–60 million years suggests the presence of asymmetric balancing selection. One possible explanation for this asymmetric balancing selection may be that the *PSMB8d* lineage is more efficient in processing most pathogen-derived proteins but that there are certain pathogen-derived proteins for which PSMB8N is more effective than PSMB8d. Overdominant selection with the differential relative fitness for *PSMB8d* and *PSMB8N* could explain this asymmetric balancing selection. However, the molecular basis for the differential relative fitness is still to be clarified through biochemical analysis addressed at different cleaving specificities of these two PSMB8 lineages.

Materials and Methods

Fishes. The wild individuals of *Oryzias* species were 244 specimens of *O. minutillus* from Patum Thani and Chai Nat, Thailand; 178 specimens of *O. javanicus* from Singapore; 69 specimens of *O. curvinotus* from Hong Kong, China; 190 specimens of *O. celebensis* from Malino and Pattunnuang, Indonesia; 106 specimens of *O. matanensis* from Lake Matano, Indonesia; 106 specimens of *O. marmoratus* from Lake Towuti, Indonesia; and 150 specimens of *O. dancena* from Linggi, Malaysia. All specimens were fixed in100% ethanol after being collected from the field and were kept at 4 °C until DNA and/or RNA extraction.

Genomic DNA Extraction. Genomic DNA was extracted from the caudal fin 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).

PCR Amplification and Sequencing of the *Oryzias PSMB8* Alleles. The *PSMB8* gene was amplified using genomic DNA as template with two primer sets. First, the *PSMB8* fragment from the second to third coding exons was amplified with a pair of primers that were designed on the conserved *PSMB8* sequences among two allelic lineages of *O. latipes* (NCBI accession nos. AB183488 and BA000027), fugu (*Takifugu ruburipes*, NCBI accession no. CAC13117, and *Tetraodon nigroviridis*, NCBI accession no. CAG11683), and zebrafish (*Danio rerio*, NCBI accession no. BC066288). The forward primer (Oryzias PSMB8.E2F) was 5′-CATGGAGTCATHGTNGCNGTNGA-3′ at the second coding exon, and the reverse primer (Oryzias PSMB8.E3R) was 5′-AGTCTGCNGCRCTNCCNGACAT-3′ at the third coding exon of *PSMB8*. The PCR condition was denaturation at 98 °C for 30 s, 35 cycles of denaturation at 98 °C for 10 s, annealing and elongation at 56 °C for 2 min, and final elongation at 72 °C for 3 min with *LA*-Taq (Takara Bio Inc.). Sequencing reaction was performed with BigDye Terminator v3.1 Sequencing Standard kit (Applied Biosystems), and nucleotide sequences were determined by a 3100/3130*xl* Genetic Analyzer (Applied Biosystems).

Determination of the Full Coding Sequence of the PSMB8 Mature Peptide. To elucidate the entire coding sequence of the mature peptide, a region from exon 1 to exon 6 (the last exon) was amplified by the previously reported primers and PCR conditions (24). The PCR products were sequenced directly using homozygous individuals or after cloning into pCR 2.1-TOPO vector or pCR-XL-TOPO vector (Invitrogen).

Because the 3′-most 69 bp of the coding sequence are not covered by this PCR amplification, 3′ RACE was performed according to the SMART RACE cDNA Amplification Kit protocol (Clontech Laboratories). Total RNA was isolated from 100% ethanol-fixed internal organs of fish using ISOGEN (Nippon Gene) according to the manufacture's instructions and was reverse transcribed into cDNA by SuperScript II (Invitrogen). The cDNA was used for PCR amplification of *PSMB8N* and *PSMB8d* alleles using the primers specific for each type.

Sequence Alignment and Phylogenetic Analysis. The nucleotide sequences of PSMB8 mature peptides were aligned using ClustalX 2.0 (37). Based on the alignments, the phylogenetic trees were constructed using the NJ method (33) with the PSMB8 sequences of spotted green pufferfish, human, and mouse as an outgroup, and bootstrap possibilities were determined with 1,000 bootstrap replications. The evolutionary distances were computed using the Kimura two-parameter method (38). These evolutionary analyses were conducted in MEGA4 (39).

Sequence Comparison Among *PSMB8* Alleles. Nucleotide sequences of alleles were compared using the dot plot generated by PipMaker (http://bio.cse.psu.edu) (40). In the dot plot, gap-free segments with more than 50% identity between two sequences were plotted. Percent identities of protein-coding regions were computed using the homology search program of GENETYX-MAC version 11.1.0 (GENETYX Corp.).

ACKNOWLEDGMENTS. We thank Drs. Peter N. G. Kee Ling and Tan Heok Hui (Raffles Museum, National University of Singapore, Singapore) and Ms. Virginia L. F. Lee (Agriculture, Fisheries and Conservation Department, Cheung Sha Wan Government Offices, Hong Kong, China) for providing wild individuals of *O. javanicus* and *O. curvinotus*, respectively; Drs. Kazunori Yamahira (University of the Ryukyus, Okinawa, Japan) and Bambang Soeroto (Sam Ratulangi University, Manado, Indonesia) for help in collecting *O. celebensis* and providing the *O. marmoratus* and *O. matanensis*; and Ms. Pingnapa Koaichumpol, Ms. Wilawan Kamsri, Ms. Sinotai Smitthikunanon (Srinakharinwirot University, Bangkok, Thailand), Dr. Ahmad Ismail, and Ms. Shahrizad Yusef (University of Putra, Selangor Darul Ehsan, Malaysia) for their help in collecting *O. minutillus* and *O. dancena*, respectively. This work was supported by Grant-in-Aid for Scientific Research on Priority Area Comparative Genomics 20017009 from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to M.N.) and by a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists (to K.T.).

1. Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: From immunogenetics to evolutionary genetics. *Annu Rev Genet* 41:281–304.
2. Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. *Annu Rev Ecol Evol Syst* 37:67–93.
3. Figueroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature* 335:265–267.
4. Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268–271.
5. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170.
6. Satta Y, Mayer WE, Klein J (1996) HLA-DRB intron 1 sequences: Implications for the evolution of HLA-DRB genes and haplotypes. *Hum Immunol* 51:1–12.
7. Esteves PJ, et al. (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: An unusual case of transspecies polymorphism. *Immunogenetics* 57: 874–882.
8. Su C, Nei M (1999) Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. *Proc Natl Acad Sci USA* 96: 9710–9715.
9. Ferguson W, Dvora S, Gallo J, Orth A, Boissinot S (2008) Long-term balancing selection at the West Nile Virus resistance gene, Oas1b, maintains transspecific polymorphisms in the house mouse. *Mol Biol Evol* 25:1609–1618.
10. Newman RM, et al. (2006) Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. *Proc Natl Acad Sci USA* 103: 19134–19139.
11. Wheat CW, Haag CR, Marden JH, Hanski I, Frilander MJ (2010) Nucleotide polymorphism at a gene (Pgi) under balancing selection in a butterfly metapopulation. *Mol Biol Evol* 27:267–281.
12. Groll M, et al. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* 386:463–471.

13. Unno M, et al. (2002) The structure of the mammalian 20S proteasome at 2.75 A resolution. *Structure* 10:609–618.
14. 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.
15. Akiyama K, et al. (1994) cDNA cloning and interferon gamma down-regulation of proteasomal subunits X and Y. *Science* 265:1231–1234.
16. Früh K, et al. (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.
17. Fehling HJ, et al. (1994) MHC class I expression in mice lacking the proteasome subunit LMP-7. *Science* 265:1234–1237.
18. Takehana Y, Naruse K, Sakaizumi M (2005) Molecular phylogeny of the medaka fishes genus Oryzias (Beloniformes: Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. *Mol Phylogenet Evol* 36:417–428.
19. Setiamarga DH, et al. (2009) Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biol Lett* 5:812–816.
20. Matsuo MY, Asakawa S, Shimizu N, Kimura H, Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). *Immunogenetics* 53:930–940.
21. Tsukamoto K, et al. (2005) Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. *Immunogenetics* 57:420–431.
22. Kasahara M, et al. (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.
23. Nonaka MI, Nonaka M (2010) Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: Haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics* 62:319–332.
24. Tsukamoto K, et al. (2009) Dichotomous haplotypic lineages of the immunoproteasome subunit genes, PSMB8 and PSMB10, in the MHC class I region of a teleost medaka, *Oryzias latipes*. *Mol Biol Evol* 26:769–781.
25. Iwamatsu T, Kobayashi H, Yamashita M, Shibata Y, Yusa A (2003) Experimental hybridization among Oryzias species. II. Karyogamy and abnormality of chromosome separation in the cleavage of interspecific hybrids between *Oryzias latipes* and *O. javanicus*. *Zoolog Sci* 20:1381–1387.
26. Iwamatsu T, Uwa H, Inden A, Hirata K (1984) Experiments on interspecific hybridization between *Oryzias latipes* and *Oryzias celebensis*. *Zoolog Sci* 1:653–663.
27. Hamaguchi S, Sakaizumi M (1992) Sexually differentiated mechanisms of sterility in interspecific hybrids between *Oryzias latipes* and *O. curvinotus*. *J Exp Zool* 263: 323–329.
28. Guntoro A (1999) The formation of the Makassar Strait and the separation between SE Kalimantan and SW Sulawesi. *J Asian Earth Sci* 17:79–98.
29. Namikawa C, et al. (1995) Isolation of Xenopus LMP-7 homologues. Striking allelic diversity and linkage to MHC. *J Immunol* 155:1964–1971.
30. Kandil E, et al. (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.
31. 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.
32. Mehta RB, Nonaka MI, Nonaka M (2009) Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus Oryzias. *Immunogenetics* 61:385–399.
33. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
34. Novacek MJ (1992) Mammalian phylogeny: Shaking the tree. *Nature* 356:121–125.
35. Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M (2006) The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet Syst* 81: 29–39.
36. 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.
37. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
38. 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.
39. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
40. Schwartz S, et al. (2000) PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res* 10:577–586.