Molecular Evolution of Antigen-Processing Genes in Salamanders: Do They Coevolve with *MHC* Class I Genes?

Gemma Palomar[1,*], Katarzyna Dudek[1], Ben Wielstra[2,3], Elizabeth L. Jockusch[4], Michal Vinkler[5], Jan W. Arntzen[3], Gentile F. Ficetola[6,7], Masatoshi Matsunami[8], Bruce Waldman[9,10], Martin Těšický[5], Piotr Zieliński[1], and Wiesław Babik[1,*]

[1]Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland
[2]Institute of Biology Leiden, Leiden University, The Netherlands
[3]Naturalis Biodiversity Center, Leiden, The Netherlands
[4]Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA
[5]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic
[6]Department of Environmental Sciences and Policy, University of Milano, Italy
[7]Laboratoire d'Ecologie Alpine (LECA), CNRS, Université Grenoble Alpes and Université Savoie Mont Blanc, Grenoble, France
[8]Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho, Japan
[9]Department of Integrative Biology, Oklahoma State University, Stillwater, Oklahoma, USA
[10]School of Biological Sciences, Seoul National University, South Korea

*Corresponding authors: E-mails: gemma.palomar@yahoo.es; wieslaw.babik@uj.edu.pl.

Accepted: 8 December 2020

Abstract

Proteins encoded by antigen-processing genes (APGs) prepare antigens for presentation by the major histocompatibility complex class I (*MHC I*) molecules. Coevolution between APGs and *MHC I* genes has been proposed as the ancestral gnathostome condition. The hypothesis predicts a single highly expressed *MHC I* gene and tight linkage between APGs and *MHC I*. In addition, APGs should evolve under positive selection, a consequence of the adaptive evolution in *MHC I*. The presence of multiple highly expressed *MHC I* genes in some teleosts, birds, and urodeles appears incompatible with the coevolution hypothesis. Here, we use urodele amphibians to test two key expectations derived from the coevolution hypothesis: 1) the linkage between APGs and *MHC I* was studied in *Lissotriton* newts and 2) the evidence for adaptive evolution in APGs was assessed using 42 urodele species comprising 21 genera from seven families. We demonstrated that five APGs (*PSMB8*, *PSMB9*, *TAP1*, *TAP2*, and *TAPBP*) are tightly linked (<0.5 cM) to *MHC I*. Although all APGs showed some codons under episodic positive selection, we did not find a pervasive signal of positive selection expected under the coevolution hypothesis. Gene duplications, putative gene losses, and divergent allelic lineages detected in some APGs demonstrate considerable evolutionary dynamics of APGs in salamanders. Overall, our results indicate that if coevolution between APGs and *MHC I* occurred in urodeles, it would be more complex than envisaged in the original formulation of the hypothesis.

Key words: antigen-processing genes, coevolution, MHC, molecular evolution, PSMB lineages, salamanders.

Introduction

The adaptive immune response is a major evolutionary innovation of vertebrates (Müller et al. 2018). Understanding its evolution has fascinated and challenged evolutionary biologists and immunologists for decades (Flajnik 2018; Flajnik and Kasahara 2010; Kaufman 2018). During the adaptive immune response in jawed vertebrates, pathogen proteins are processed into antigens that are presented on the cell surface to allow the recognition and initiation of a highly specific, targeted protective response, as well as the formation of

Significance

Coevolution between two key components of adaptive immunity, antigen-processing genes (APGs) and major histocompatibility complex class I (*MHC I*) genes, may be widespread among nonmammalian vertebrates. We used an ancient tetrapod group, salamanders, to test two expectations stemming from the coevolution hypothesis. We confirmed the tight genetic linkage between APGs and *MHC I*. However, we did not find support for pervasive adaptive evolution of APGs across the salamander phylogeny. Thus, if APGs and *MHC I* indeed coevolved in salamanders, the process would be more complex than envisaged in the original formulation of the hypothesis: Salamanders may have evolved mechanisms to reconcile coevolution with *MHC I* gene duplications, increasing the efficiency of adaptive immunity, and improving protection against pathogens.

immunological memory. Based on the antigen processing compartment, the antigen-presenting molecule, and the recognition cell, different antigen presentation pathways can be distinguished (Blum et al. 2013; Murphy and Weaver 2016). In particular, the direct presentation or endogenous pathway is triggered in most cell types by intracellular pathogens. Evidence from placental mammals suggests that antigen processing in this pathway starts in the immunoproteasome, a large proteolytic complex, in which three immunoproteasome-specific catalytic subunits PSMB8 (LMP7), PSMB9 (LMP2), and PSMB10 (MECL-1) degrade pathogen proteins to short peptides with hydrophobic carboxy-terminal residues, which are suitable ligands for major histocompatibility complex class I (MHC I) proteins (reviewed in Murata et al. 2018). These peptide antigens are translocated from the cytosol to the lumen of the endoplasmic reticulum (ER) through a channel formed by the TAP1–TAP2 (transporter associated with antigen processing 1 and 2) heterodimer (Blees et al. 2017). Within the ER, the peptide-loading complex, especially TAPBP (TAP-binding protein), positions an MHC I molecule near the TAP complex and acts as a bridge between these. This enhances TAP stability and facilitates peptide translocation. TAPBP also stabilizes empty MHC molecules and optimizes MHC loading with peptides. Once an MHC I molecule has bound a peptide, the molecule travels to the cell surface to present the antigen to cytotoxic CD8+ T cells (Murphy and Weaver 2016; Paulsson 2004).

Antigen processing proteins and MHC I interact either directly or indirectly to initiate the adaptive immune response, which sets the stage for coevolution between genes that encode them. Coevolution between antigen-processing genes (APGs) and *MHC I* gene has been thoroughly studied in chicken, inferred in the frog *Xenopus* and rat and even suggested as the ancestral gnathostome condition (Joly et al. 1998; Kaufman 1999, 2015; Ohta et al. 2006). The strength of evidence for coevolution varies among the taxa and each has peculiarities indicating that the details of the process may differ. Nonetheless, in its essence, the coevolution hypothesis posits that the properties of antigenic peptides processed by proteins encoded by APG alleles match binding properties of the *MHC I* allele co-occurring on the same haplotype. Such coevolutionary fine-tuning would increase the efficiency of antigen presentation but would also lead to the emergence of specialist (i.e., processing or presenting a restricted spectrum of antigens) APGs and *MHC I* alleles. Experimental data from chicken support this last prediction (van Hateren et al. 2013; Walker et al. 2011). In several sequenced genomes of nonmammalian vertebrates, most APGs (i.e., *PSMB8*, *PSMB9*, *TAP1*, *TAP2*, and *TAPBP*) are located very close to the *MHC I*. Tight linkage would keep particular combinations of APG and *MHC I* alleles together long enough for coevolutionary fine-tuning to occur and would also reduce the chance of generating low-fitness allele combinations (reviewed in Kaufman 2015; Ohta and Flajnik 2015). Chicken and the frog *Xenopus* have a single highly expressed *MHC I* gene and highly polymorphic APGs, with particular interlocus combinations of alleles segregating as stable haplotypes (Flajnik et al. 1999; Kaufman 2015). A single highly expressed *MHC I* gene has been proposed as a consequence of coevolution (Kaufman 2015) and constitutes an important prediction stemming from the hypothesis. Such a mode of coevolution may have profound consequences for the efficiency of adaptive immunity and fighting pathogen assault, because it would impose constraints on *MHC I* variation by selecting against gene duplication. The breakup of tight linkage between APGs and *MHC I* that occurred in mammals would have led to the emergence of monomorphic generalist APGs that provide peptides to any *MHC I* allele, which in turn might have brought about evolution of the multigene *MHC I* family and better protection against pathogens (Kaufman 2015).

APG–*MHC I* coevolution as the ancestral gnathostome condition (Kaufman 1999; Ohta and Flajnik 2015) is difficult to reconcile with the pattern found in several teleost fishes, birds, and urodele amphibians (e.g., Drews and Westerdahl 2019). For example, although at least some APGs (*TAP1*) are highly polymorphic in urodeles (Fijarczyk et al. 2016, 2018), the presence of multiple highly expressed *MHC I* genes (Fijarczyk et al. 2018) should preclude coevolution under the original formulation of the hypothesis. From 6 to 21, *MHC I* loci are expressed in the axolotl, *Ambystoma mexicanum* (Sammut et al. 1999), and from two to at least five, probably

more, are highly expressed (revealed by transcriptome sequencing) and polymorphic in the *Lissotriton vulgaris* complex (Fijarczyk et al. 2018). Although it remains unclear whether all these genes represent classical *MHC I* genes, urodeles nevertheless present an intriguing pattern of *MHC I* and APG variation. High *MHC* polymorphism and the unambiguous signal of adaptive evolution in several species (Babik et al. 2009; Bos and DeWoody 2005; Fijarczyk et al. 2018) testify to the functional importance of *MHC* polymorphism in urodeles, despite previous suggestions linking a weak immune response of the axolotl with a low *MHC* polymorphism (Tournefier et al. 1998). Urodeles may have found a way to achieve duplication and expansion of *MHC I*, which could provide better protection against pathogens, while maintaining coevolution with APGs. Thus, research on this group might reveal novel mechanisms relaxing the selective constraints that coevolution imposes on adaptive immunity.

Under the coevolution hypothesis, APGs should be affected by adaptive evolution driving *MHC I* variation. Novel *MHC* variants are positively selected, which leads to their establishment in populations, while various forms of balancing selection maintain the polymorphism over extended periods (Radwan et al. 2020). Together, these processes generate a signal of adaptive evolution at phylogenetic scale detectable with standard tests for positive selection (Yang 2007). Such a signal has indeed been detected in *MHC* genes of most vertebrate species analyzed so far (Radwan et al. 2020), including urodeles (e.g., Fijarczyk et al. 2018). Similarly, even though the coevolution is an intraspecific process, adaptive evolution at functionally relevant APG codons should be detectable at phylogenetic scales if these genes coevolve with *MHC I*. Although this expectation has not been tested in nonmammalian vertebrates as molecular evolution of APGs has been studied in detail only in mammals (Forni et al. 2014), the pattern observed in some APGs is likely a result of adaptive evolution and may be related to coevolution with *MHC I*. Divergent allelic lineages have been described in several jawed vertebrates for *PSMB8*, *TAP1*, and *TAP2* (Huang et al. 2013; Kandil et al. 1996; McConnell et al. 2016; Miura et al. 2010; Namikawa et al. 1995; Nonaka et al. 2000; Ohta et al. 2003), and, in some cases, these are strongly associated with *MHC I* lineages, implying coevolution (Joly et al. 1998; Walker et al. 2011). The two lineages of *PSMB8* found in some fishes, amphibians, and reptiles could imply different specificities, thereby contributing to an expanded MHC I antigen recognition repertoire and increasing fitness of heterozygous individuals. This lineage dimorphism might be under a strong overdominance-type of balancing selection (Tsukamoto et al. 2012; but see Yamaguchi and Dijkstra 2019). In fact, *PSMB8* F type has apparently been regenerated from the *PSMB8* A type at least five times independently during tetrapod evolution (Huang et al. 2013).

In this study, we tested in the urodele amphibians two of the crucial expectations derived from the coevolution hypothesis. First, we checked whether APGs and *MHC I* are indeed tightly linked, by direct estimation of the recombination rate in a large newt pedigree. Second, we checked whether APGs show pervasive adaptive evolution across urodele phylogeny, by examining the signal of positive selection. Furthermore, we studied the possible functional role of codons under positive selection within the APG proteins, inferring their potential for interaction with other proteins (measured as surface accessibility) and their electrostatic surface charge. We also tested for the presence of divergent allelic lineages in APGs, similar to those described previously in other ectotherms.

Materials and Methods

In addition to the APGs (i.e., *PSMB8*, *PSMB9*, *TAP1*, *TAP2*, and *TAPBP*), we analyzed several non-APGs: genes within the *MHC* region, tightly linked to *MHC I*, but not involved in antigen processing or presentation (i.e., *BRD2*, *DAXX*, *KIFC1*, *RGL2*, and *RXRBA*). The non-APGs were included as a control to check whether patterns recovered for APGs could be a simple consequence of their location within the *MHC* region.

Linkage Analysis of APGs, Non-APGs, and *MHC* Genes

To verify linkage between APGs, *MHC*, and non-APGs and to estimate the recombination rate within the region, we used genomic DNA from a large mapping (recombinant) newt population. The mapping population consisted of lab-produced F2 generation hybrids between *Lissotriton montandoni* and *L. vulgaris*, derived from two pairs of interspecific crosses (generation P). APGs, non-APGs, *MHC I* and *II*, as well as two additional markers (*LGR4* and *RABGAP1*) were genotyped by sequencing with molecular inversion probes (MIPs, supplementary table S1, Supplementary Material online) using the procedure of Niedzicka et al. (2016). The markers *LGR4* and *RABGAP1* were mapped previously in the proximity of *MHC* (Fijarczyk et al. 2018) and were used here to orient the genes of interest along the centromere–telomere axis. Because *MHC* alleles and haplotypes segregating in the mapping population were already known (Dudek et al. 2019; Fijarczyk et al. 2018), we genotyped *MHC* using allele-specific MIPs (supplementary table S1, Supplementary Material online). Cases of ambiguous genotyping were resolved using Illumina sequencing of amplicons as described previously (Dudek et al. 2019; Fijarczyk et al. 2018). Cri-map 2.507 (Green et al. 1990) was used to estimate recombination distances between genes, order them, and identify recombinants. Details of genotyping procedure and analysis of recombination are in supplementary methods, Supplementary Material online.

Transcriptomic Data

Transcriptome assemblies from 42 species of the Urodela representing 21 genera and seven out of nine extant families

were obtained from various sources (table 1 and fig. 1). Transcriptomes of adult *Hydromantes italicus*, *Hyd. strinatii*, *Hynobius leechi*, and *Hyn. retardatus* (only larval transcriptomes have been available for this species so far, and *MHC I* expression is limited in amphibian larval stages, Salter-Cid et al. 1998) were newly generated for this study from two individuals per species, except for the *Hyd. italicus* transcriptome obtained from a single individual. RNA was extracted from tail tips stored in RNAlater; libraries were prepared with Illumina TruSeq stranded mRNA kits and sequenced on an Illumina platform, yielding 2 × 100 bp reads, except for *Hyn. leechi*, whose transcriptome was obtained using a proprietary BGI technology. Raw reads were cleaned with Trimmomatic (Bolger et al. 2014) and assembled de novo with Trinity (Grabherr et al. 2011) using settings recommended by the authors.

Recovering Sequences of Target Genes

To identify coding sequences of our genes of interest (APGs as well as non-APGs), transcriptome assemblies of all the focal species were blasted against the reference sequences identified previously in the transcriptome of *L. montandoni/vulgaris* (Stuglik and Babik 2016, http://newtbase.eko.uj.edu.pl/, last accessed September 20, 2020). In rare cases, fragments of some target genes were not recovered from a transcriptome assembly. To fill these gaps, we mapped raw RNAseq reads to the reference from a closely related species and used the mapped reads to recover the sequence of the missing fragments. If substantial regions of unknown sequence remained, the gene was removed from subsequent analyses for this species. When no contig was recovered for a gene from the transcriptome assembly, the gene may have been lost in a given species, but we inferred a putative gene loss only if additional criteria were met. As genes may be missing from assemblies because of low expression level or poor quality assembly, we checked whether the gene was absent from assemblies of related species and whether its expression was comparable to that of other APGs in other species. To confirm duplications, when two transcriptome contigs were found for a gene, we visually inspected read mappings to transcripts to check whether more than two alleles were present within an individual. When more than one transcript per gene was present in the assembly, one of the sequences was picked at random for the analysis of positive selection in this gene.

Sequences of all the APGs were examined to detect the presence of divergent lineages. Phylogenetic analysis of genes with confirmed divergent lineages, *PSMB8* and *PSMB9*, was performed using all transcriptome contigs for each species and, in some cases, additional sequences recovered from read mapping, which differed by more than several bp. Maximum likelihood trees were constructed under the General Reversible Time model of evolution in MEGA7 (Kumar et al. 2016), using *Xenopus laevis* as outgroup. Node support was assessed with nonparametric bootstrapping with 1,000 replicates.

Analysis of Positive Selection

Recombinant sequences were identified using Genetic Algorithm for Recombination Detection (GARD, Kosakovsky Pond et al. 2006). Only the break points (i.e., points that define boundaries between segments of the alignment with no evidence for ancestral recombination) supported by both the topological incongruence test (at a Bonferroni-corrected *P* value = 0.01) and the comparison of single vs. multiblock maximum likelihood models were considered. For genes with such well-supported recombination break points, analyses of selection were performed separately for each nonrecombining block.

In protein-coding sequences, nonsynonymous nucleotide substitutions change the amino acid sequence, whereas synonymous substitutions do not. The ratio (ω) of the rate of nonsynonymous changes per nonsynonymous site (d*N*) to the rate of synonymous changes per synonymous site (d*S*) is used to quantify the mode and strength of selection under the assumption that selection affects mainly nonsynonymous changes. When ω exceeds unity, positive selection that promotes changes in the protein sequence is inferred, whereas ω lower than unity is expected if purifying selection opposes changes in the protein sequence (Yang 2019). The overall d*N* and d*S* as well as ω for each of our genes of interest were calculated using MEGA7 (Kumar et al. 2016). To compare the strength of purifying selection acting on APGs versus non-APGs, we carried out a two-tailed Mann–Whitney *U* test on the values of ω (in genes with partitions we used weighted means for ω, taking into account the length of each partition).

To test for positive selection, the M7 codon-based substitution model that assumes a variable, beta distributed, negative selective pressure ($0 \leq \omega \leq 1$), was compared with the M8 model that additionally assumes nonzero fraction of codons under positive selection ($\omega > 1$). Both models were also compared with the null model (M0) that assumes a single ω for all codons. The calculations were performed in codeml from PAML 4.9 (Yang 2007). A comprehensive phylogeny of the Urodela (Jetz and Pyron 2018), modified to include the most recent phylogeny of the Salamandridae (Rancilhac et al. 2021), was used as the phylogenetic tree in all the analyses except those of *PSMB8* and *PSMB9*. For these two genes, maximum likelihood trees under the General Time Reversible model of evolution were constructed with MEGA7 (Kumar et al. 2016) based on our sequence information, because gene trees departed substantially from the species phylogeny. Models were compared with the likelihood ratio test and the Akaike information criterion. When the M8 model was supported, the Bayes empirical Bayes method (Zhang et al. 2005) was used to identify the specific codons

Table 1

Summary of Transcriptome Data Used in Analyses of Positive Selection

Family	Genus	Species	*PSMB8*	*PSMB9*	*TAP1*	*TAP2*	*TAPBP*	*BRD2*	*DAXX*	*KIFC1*	*RGL2*	*RXRBA*	SRA accession	References
Ambystomatidae	*Ambystoma*	*laterale*											SRR5346170	McElroy et al. (2017)
Ambystomatidae	*Ambystoma*	*maculatum*											SRR5144808	Burns et al. (2017)
Ambystomatidae	*Ambystoma*	*mexicanum*											SRR2885869	Bryant et al. (2017)
Ambystomatidae	*Ambystoma*	*texanum*											SRR5346172	McElroy et al. (2017)
Ambystomatidae	*Ambystoma*	*tigrinum*											SRR5346171	McElroy et al. (2017)
Cryptobranchidae	*Andrias*	*davidianus*											SRR5260688	Huang et al. (2017)
Cryptobranchidae	*Cryptobranchus*	*alleganiensis*											NA	Provided by D. Weisrock
Hynobiidae	*Hynobius*	*chinensis*											SRR1042328, SRR7299432	Che et al. (2014)
Hynobiidae	*Hynobius*	*leechii*											Will be provided upon acceptance	This study
Hynobiidae	*Hynobius*	*retardatus*											Will be provided upon acceptance	This study
Plethodontidae	*Batrachoseps*	*gregarius*											NA	Provided by E. Jockusch
Plethodontidae	*Batrachoseps*	*nigriventris*											NA	Provided by E. Jockusch
Plethodontidae	*Desmognathus*	*fuscus*											SRR4253224	Madison-Villar et al. (2016)
Plethodontidae	*Hydromantes*	*italicus*											Will be provided upon acceptance	This study
Plethodontidae	*Hydromantes*	*strinatii*											Wwill be provided upon acceptance	This study
Plethodontidae	*Karsenia*	*koreana*											NA	Provided by T. Kwon
Proteidae	*Proteus*	*anguinus*											SRX2382497	Irisarri et al. (2017)
Salamandridae	*Calotriton*	*asper*											SRR5062019	Irisarri et al. (2017)
Salamandridae	*Chioglossa*	*lusitanica*											SRR11118089	Provided by J. W. Artzen
Salamandridae	*Cynops*	*cyanurus*											NA	Provided by D. Weisrock
Salamandridae	*Cynops*	*pyrrhogaster*											SRR2083850	Sousounis et al. (2015)
Salamandridae	*Ichthyosaura*	*alpestris*											SRX7755081	Rancilhac et al. (2021)
Salamandridae	*Lissotriton*	*maltzani*											SRR2535189, SRR2495449	Nourisson et al. (2017)
Salamandridae	*Lissotriton*	*boscai*											SRR2481123	Nourisson et al. (2017)
Salamandridae	*Lissotriton*	*helveticus*											SRR3303063, SRR7396737	Fijarczyk et al. (2016)
Salamandridae	*Lissotriton*	*montandoni*											PRJNA316531	Stuglik and Babik (2016)
Salamandridae	*Notophthalmus*	*viridescens*											SRR653294, SRR653288	Abdullayev et al. (2013)
Salamandridae	*Ommatotriton*	*ophryticus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Pleurodeles*	*waltl*											SRR6001111	Elewa et al. (2017)
Salamandridae	*Salamandra*	*salamandra*											SRR1693191, SRR5494542	Rodríguez et al. (2017)
Salamandridae	*Triturus*	*anatolicus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*carnifex*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*cristatus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*dobrogicus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*ivanbureschi*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*karelinii*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*macedonicus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*marmoratus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Triturus*	*pygmaeus*											PRJNA498336	Wielstra et al. (2019)
Salamandridae	*Tylototriton*	*wenxianensis*											SRR2989161	Farrer et al. (2017)
Sirenidae	*Siren*	*intermedia*											NA	Provided by D. Weisrock
Sirenidae	*Siren*	*lacertina*											SRR5062014	Irisarri et al. (2017)
		N	33	36	38	34	40	41	41	40	36	38		

NOTE.—Gray cells indicate gene/species included in the analyses, total count is at the end of each column.

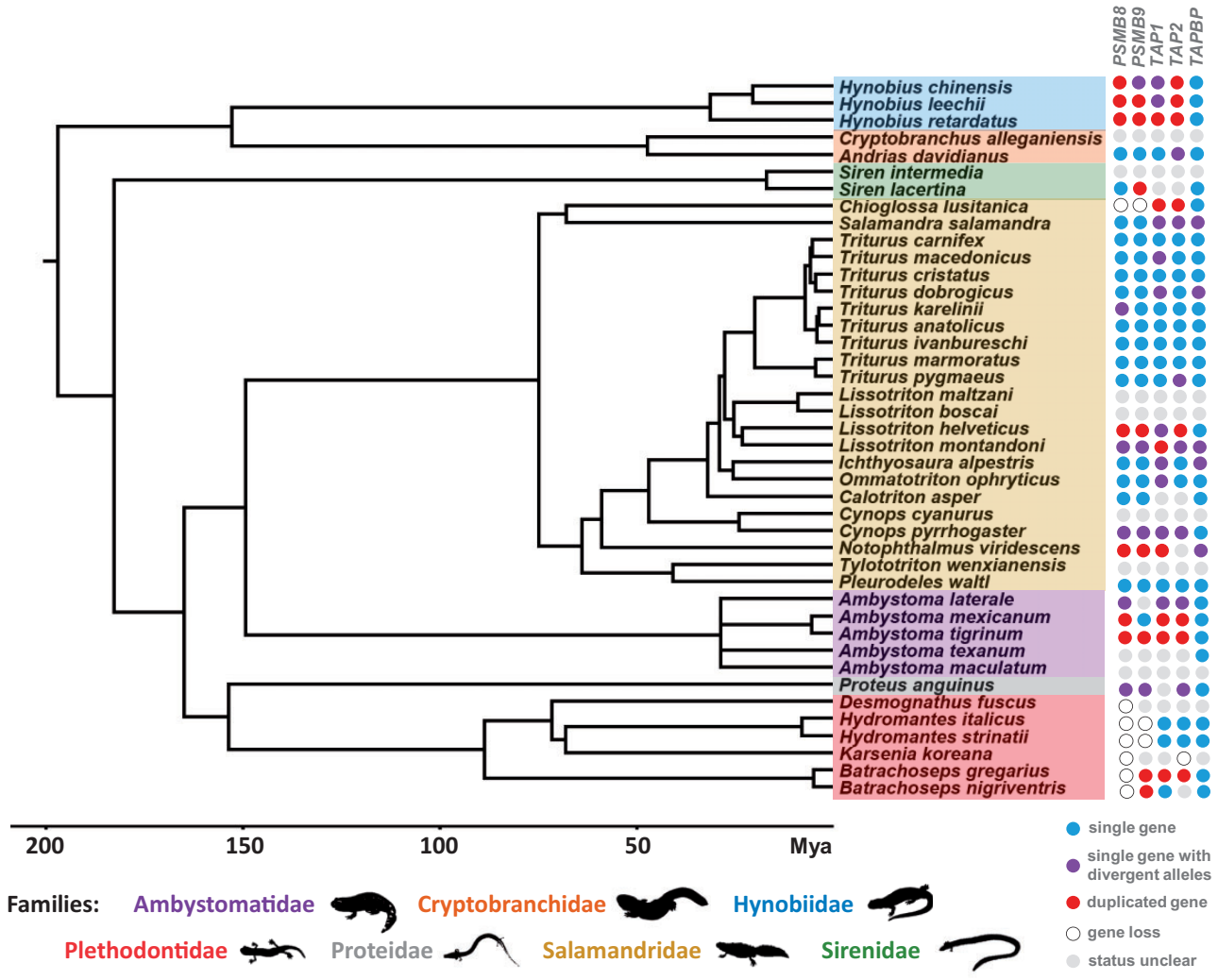

Fig. 1.—Phylogenetic tree of the species used in this study. Different colors define families. The right panel shows pattern of duplication of APGs. Blue circle indicates a single gene, whereas red circle designates a duplicated gene. Purple circle shows species with two different contigs for a gene but where three alleles were not detected in the read mapping. Empty circle indicates an apparent gene loss. Finally, gray circle indicates species in which only transcriptome assembly was available and there was no possibility to check number of alleles in the read mapping.

under positive selection. In addition to codeml analyses, tests of site-level pervasive (FUBAR, Murrell et al. 2013) and episodic (MEME, Murrell et al. 2012) positive selection were performed in Datamonkey (Weaver et al. 2018).

Protein Structure Analysis

The 3D structure of the proteins encoded by APGs in each species was modeled to locate the position of residues under positive selection and to gain insight into their possible functional significance. The structure of the protein encoded by each gene in each species was predicted using its protein sequence, the most similar protein cryo-EM structure (hereafter template) from the Protein DataBank (PDB, www.rcsb.org, last accessed June 05, 2019), and the homology modeling software Modeller (Webb and Sali 2016). For each protein sequence, we generated ten models and selected the one with the lowest Discrete Optimized Protein Energy score for further analysis. The quality of these modeled structures was checked in ModFOLD (Maghrabi and McGuffin 2017). PyMOL (Schrödinger 2019) was used to align these predicted 3D structures to the template, and its plug-in APBS to predict the electrostatic surface charge of each molecule. Because amino acid similarity between human and urodele TAPBP is $<35\%$, we decided to be conservative and did not try to infer models for this gene based on the human template (PDB accession number 6ENY).

To shed light on the potential functional relevance of the positively selected sites, we estimated the surface receptiveness of each residue. For this, we calculated the relative solvent accessibility (RSA) from the residue's solvent accessibility

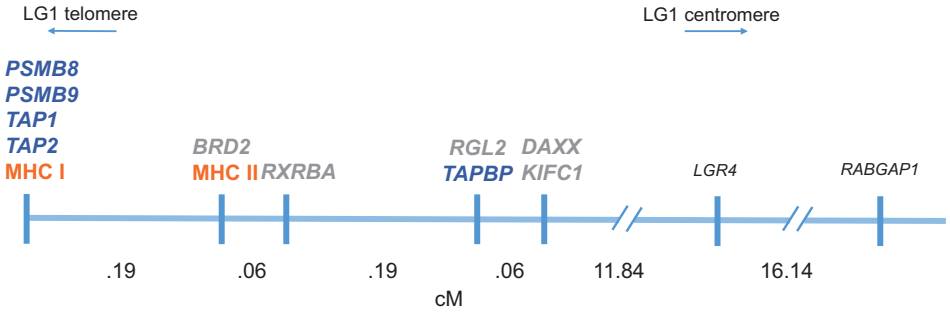

Fig. 2.—*MHC* genomic region in *Lissotriton* newts. Recombination distance is expressed in centimorgans (cM). APGs are in dark blue, non-APGs in gray, and *MHC* genes in orange. *LGR4* and *RABGAP1* are additional markers outside the *MHC* region, used to orient the genes of interest along the centromere–telomere axis.

surface area (ASA) computed from the PDB models of five distantly related species: *Hyd. italicus* (Plethodontidae), *Hyn. chinensis* (Hynobiidae), *Proteus anguinus* (Proteidae), *Siren intermedia* (Sirenidae), and *Triturus dobrogicus* (Salamandridae), using the webserver xssp (http://www.cmbi.ru.nl/xssp/, last accessed August 29, 2019; based on Kabsch and Sander 1983). Then, we divided RSA by the corresponding maximum possible ASA for a given amino acid (Tien et al. 2013). We considered the residue sufficiently exposed on the protein surface to allow external interactions when the ratio was >0.2.

PSMB8 and *PSMB9* Population-Level Resequencing

To clarify whether the divergent *PSMB8* and *PSMB9* lineages represented divergent alleles of the same gene or duplicated genes, we designed lineage-specific MIPs (supplementary table S1, Supplementary Material online) to resequence *PSMB8* and *PSMB9* in two populations (10–21 individuals per population) of each of seven species of the genus *Triturus* (supplementary table S2, Supplementary Material online). This genus was selected because numerous samples from natural populations were available for several species (e.g., Wielstra, Burke, Butlin, and Arntzen 2017; Wielstra, Burke, Bultin, Avci, et al. 2017). The number of reads mapped to each of the two references of each *PSMB* gene (i.e., the consensus sequence of a lineage obtained from the transcriptome data of all *Triturus* species) was counted to determine the lineage/s of each individual. In the case of polymorphic populations, we used Genepop (Rousset 2008) to test whether genotype frequencies followed Hardy–Weinberg expectations, which would provide support for the presence of a single locus.

Results

Linkage Analysis

APGs, non-APGs, and *MHC* are all tightly linked in *Lissotriton* newts. In total, eight recombinants were detected among 766 *L. montandoni* × *L. vulgaris* F2 individuals (i.e., >1,500 meioses) that formed the mapping population (supplementary table S3, Supplementary Material online). The linkage map of the *MHC* region, together with recombination distances, is given in figure 2. No recombination was detected between four APGs (*PSMB8*/*PSMB9*/*TAP1*/*TAP2*) and *MHC I*, and the map distance between this block and the most distant APG (*TAPBP*) was estimated at 0.448 cM. All five non-APGs are either embedded in this region (*BRD2*, *RXRBA*) or tightly linked to it (*DAXX*, *KIFC1*, and *RGL2*) (fig. 2). Further details are provided in supplementary results, Supplementary Material online.

Duplication and Loss of APGs

Analyzing transcriptome assemblies, we found evidence for duplication as well as putative losses of some APGs in some species (fig. 1). In most cases, gene duplications appeared to have occurred recently, because they were not shared with closely related genera, with the possible exception of *PSMB8* and *PSMB9* (see below). The families Ambystomatidae and Hynobiidae showed frequent duplications in all APGs except *TAPBP* (fig. 1). Both *TAPBP* and non-APGs were only rarely duplicated possibly due to their location further from the remaining APGs and *MHC I* genes. Almost all cases of putative gene loss were observed in plethodontid salamanders: *PSMB8* was not found in all four genera examined, in addition, *PSMB9* was not found in *Hydromantes*, and *TAP2* was not found in *Karsenia*. The loss of *PSMB8* in plethodontids is unlikely to result from its low expression or poor transcriptome quality because *PSMB9*, a functionally related gene, was recovered at high coverage, although it has lower expression than *PSMB8* in most investigated vertebrate taxa (Petryszak et al. 2016), including the remainder of urodele species that had both genes transcribed. *PSMB8* and *PSMB9* were not recovered in *Chioglossa lusitanica*, but limited transcriptome data were available for this species, so it remains unclear whether these genes are indeed missing (table 1).

Positive Selection and Protein Structure Analyses

Depending on the gene, from 33 to 41 species were employed in the analysis of positive selection (table 1). Only

the first 2,319 bp of *DAXX* were used because the quality of alignment of the remaining part was dubious. Recombination break points were identified in *DAXX*, *KIFC1*, and *TAP2*. Alignments of these genes were divided into nonrecombining blocks, and each block was analyzed separately (table 2). *TAP2* sequences were divided into three parts with break points at alignment positions 543 and 1175. *KIFC1* and *DAXX* sequences were split into two fragments with the break point at alignment positions 1656 and 2040, respectively. The overall ratio of nonsynonymous to synonymous substitutions (ω) was significantly higher for APGs ($\omega = 0.20$) than for non-APGs ($\omega = 0.10$) (Mann–Whitney $U = 2$, $P < 0.05$). However, we did not find a consistent signal of positive selection in APGs: The M8 model, allowing positive selection, was supported only for *TAP1* and *TAP2*, whereas among non-APGs support for the M8 model was found in *BRD2* and *RGL2*. The Bayes empirical Bayes method identified two codons under positive selection in *TAP1* and *RGL2* and four in *TAP2* (table 2). FUBAR identified three codons under positive selection in *PSBM8* and *BRD2* and four in *TAP2*. More codons under episodic positive selection were localized in the MEME analysis: six in *PSMB8*, one in *PSMB9*, seven in *TAP1*, eight in *TAP2*, three in *TAPBP*, four in *BRD2*, ten in *DAXX*, four in *KIFC1*, and three in *RGL2* (table 2). However, only seven codons were identified by more than one method: three in *PSMB8* (codons 26, 104, and 190), three in *TAP2* (codons 67, 144, and 470), and one in *BRD2* (codon 8). The majority of substitutions in codons identified as positively selected were physicochemically nonconservative (supplementary table S4, Supplementary Material online), and the residues were surface accessible (supplementary table S5, Supplementary Material online), which allows them to interact with other molecules.

TAP1 and *TAP2* protein structures were modeled based on the Cryo-EM structure of the human TAP ATP-Binding Cassette Transporter (PDB accession number 5U1D). This model covered residues 144–714 of the TAP1 protein alignment and 148–702 of the TAP2 protein alignment (supplementary video S1, Supplementary Material online). PSMB8 and PSMB9 protein structures were modeled based on chain K and chain N, respectively, of the Cryo-EM structure of the mouse 20S immunoproteasome (PDB accession number 3UNF). The chain K model covered residues 74–274 of the PSMB8 protein alignment and chain N model covered positions 19–217 of the PSMB9 protein alignment (supplementary video S2, Supplementary Material online). Considerable differences between urodele species in surface charge (e.g., residue 295 and 470 TAP2) and volume/shape (e.g., 470 TAP2 and 104 PSMB8) were observed at the residues under positive selection (supplementary figs. S1–S3, Supplementary Material online).

PSMB8 and *PSMB9* Lineages

At deeper levels, the maximum likelihood tree based on *PSMB8* nucleotide sequences reflected the species phylogeny; every family was recovered as monophyletic and the relationships among them were as expected. However, relationships of *PSMB8* genes within three of the families (i.e., the Hynobiidae, Ambystomatidae, and Salamandridae) differed substantially from the relationships among species (fig. 3*a*). All three of these families, and also *P. anguinus*, contained divergent lineages that correspond to the two divergent lineages (A and F) defined by codon 104 of the alignment (codon 97 in *L. montandoni* sequence and codon 31 in the tetrapod alignment of Huang et al. 2013). The amino acid that defines the clades varies between Alanine (A) and Valine (V) for the A lineage and Phenylalanine (F) and Tyrosine (Y) for the F lineage (supplementary fig. S4, Supplementary Material online). Within the Salamandridae, the deepest divergence in *PSMB8* separated the two subfamilies (i.e., the Salamandrinae and Pleurodelinae) and subfamilies were split into two clades according to the A and F lineages. Furthermore, in *P. anguinus*, the Hynobiidae, Ambystomatidae, and Salamandrinae, both the A and F lineages were detected in most species sampled. In the Pleurodelinae, this was the case at the genus level but usually only one lineage was found at species level. We detected more than two alleles, that is, gene duplication, in at least one individual of the following species: *Ambystoma mexicanum*, *A. tigrinum*, *Hyn. leechii*, *Hyn. chinensis*, *Hyn. retardatus*, *L. helveticus*, and *Notophthalmus viridescens* (fig. 1).

Interestingly, a similar pattern of divergent lineages appeared in the phylogenetic tree of *PSMB9* in the Salamandridae (fig. 3*b*): two clades are present in each subfamily. However, the protein sequence divergence between the clades of the Pleurodelinae is higher and widely distributed along the sequence, whereas in the Salamandrinae only two amino acids at the beginning of the protein diverged between the clades (supplementary fig. S5, Supplementary Material online). In the former subfamily, the clades could be defined by amino acid 49 of the alignment (corresponding to 48 in *L. montandoni* sequence and 31 in the alignment of Ferrington and Gregerson 2012) which was either Methionine (M) or Phenylalanine (F). This amino acid corresponds, in function and position, to the amino acid that defines the lineages in *PSMB8* (i.e., amino acid 104). These amino acids participate in defining the cleaving specificity of the immunoproteasome (Ferrington and Gregerson 2012). *PSMB9* was not found in *Hyd. italicus*, *Hyd. strinatii*, or *Chioglossa lusitanica*, although it was duplicated in at least one individual of *A. tigrinum*, *B. gregarius*, *B. nigriventris*, *Hyn. leechii*, *Hyn. retardatus*, *L. helveticus*, *N. viridescens*, and *Siren lacertina* (fig. 1).

PSMB8 and *PSMB9* Population Resequencing

Population-scale, genomic DNA-based resequencing revealed a single *PSMB8* and *PSMB9* lineage in five of the seven investigated *Triturus* species, in each case, it was the lineage detected earlier in the transcriptome of the species. *PSMB8*

Table 2.

Analysis of positive selection.

Gene	*TAP1****	*TAP2*^{**,NS,*}	*TAPBP*	*PSMB8*	*PSMB9*	*KIFC1*	*RXRBA*	*RGL2***	*BRD2**	*DAXX*
Length (bp)	2175	543, 633, 1026	1401	828	651	1656, 327	1359	2367	2472	2040, 279
?AIC (M8-M7)	-23.82	-20.69, 2.75, -2.92	3.3	3.98	4	4, 4	0.78	-12.73	-5.14	4, 4
dN/dS=ω	0.11/0.583=0.19	0.226/0.406=0.56 0.095/0.507=0.19 0.117/0.609=0.19	0.156/0.664=0.23	0.083/0.491=0.17	0.085/0.596=0.14	0.137/1.059=0.13 0.020/0.651=0.03	0.012/0.642=0.02	0.103/0.703=0.15	0.041/0.837=0.05	0.152/1.031=0.15 0.109/1.042=0.10

Analysis	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME	CODEML	FUBAR	MEME
Codons	-	-	118*			31*	-	-	277*	-	26*	26*	-	-	45*	-	-	6*	-	-	-	-	71**	-	8*	8**	-	-	24*
			132**	66**					**364***		**104**	**104***						37*					153*		65*				25*
			145*		67	67*			454*			161**						94*				229*				588**			58*
			292*			122*					190	190**						224*				233*				628*			106**
			468**	132*								223*											429*			812*			215**
	473*			144*	144	144*						234**																	224**
			664*			**284****																							397*
			700*			295**																							416*
	708*					**323***																							454**
					470*	**470***																							731**
					476																								
				599*																									

Significant support for the M8 model of positive selection (likelihood-ratio test) is indicated with asterisks following the gene symbol. For genes with the evidence of recombination(*TAP2*, *KIFC1*, *DAXX*) tests were performed for each non-recombining block separately. The ratio of nonsynonymous (dN) to synonymous (dS) substitutions was calculated in MEGA using the Nei-Gojobori method with Jukes-Cantor correction for multiple substitutions. For each gene codons identified as positively selected by three methods are indicated. Numbering of codons reflects their position in aligment. In bold codons of potential functional relevance as identified in previous research (see text for details). No asterisk: posterior probability (PP) > 0.9;

$^{*}p < 0.05$ level, or PP > 0.95;

$^{**}p < 0.01$, or PP > 0.99.

lineage F and *PSMB9* lineage F occurred in *T. cristatus*, *T. macedonicus*, *T. marmoratus*, and *T. pygmaeus*, whereas *PSMB8* lineage A and *PSMB9* lineage M were detected in *T. ivanbureschi*. Both *PSMB8* and *PSMB9* lineages were detected in *T. dobrogicus* and *T. karelinii*. In *T. dobrogicus*, individuals with both *PSMB8* and *PSMB9* lineages, as well as *PSMB8* A–*PSMB9* M individuals were present in both populations. One population, Senta (Serbia), departed from Hardy–Weinberg equilibrium with an excess of heterozygotes (individuals possessing both lineages). Together with the presence of individuals possessing three alleles, this indicates that gene duplication has occurred in *T. dobrogicus*. In *T. karelinii*, the two sampled populations exhibited different patterns. In Alushta (Crimea, Ukraine), all individuals were *PSMB8* A–*PSMB9* M, a combination revealed by transcriptome sequencing of this species, whereas in Chiantba Lake (Georgia), all individuals exhibited both lineages for both genes. This excess of heterozygotes indicated a duplication. We were not able to check whether these duplications included the whole gene or whether both loci were transcribed because transcriptomes were obtained from individuals with just one lineage per gene. Additionally, our resequencing revealed an apparently nontranscribed (judging from comparison with transcriptome data) sequence of *PSMB8*, possibly a pseudogene, covering exon 1 in *T. cristatus*, *T. macedonicus*, *T. karelinii*, and *T. dobrogicus.*

Discussion

In this study, we tested in urodele amphibians two key predictions derived from the APGs–*MHC I* coevolution hypothesis. We found tight linkage between APGs and *MHC I*, a condition considered necessary for coevolution to operate. However, we did not find pervasive adaptive evolution in APGs across the urodele phylogeny, which would be expected under coevolution hypothesis as a consequence of adaptive evolution in *MHC I*. Nonetheless, gene duplications, gene losses, and divergent allelic lineages detected here testify to a considerable evolutionary dynamics of APGs in the Urodela, compared with other genes encoded in the same region.

Linkage between APGs and MHC I

By analyzing recombination in a large experimental mapping population, we verified that APGs are tightly linked with both *MHC* classes and located closer to class I than to class II in *Lissotriton* newts. The segment containing all APGs and both *MHC* classes was shorter than 0.5 cM. Because no recombinants were observed among products of >1,500 meioses, APGs, except *TAPBP*, are exceptionally tightly linked to *MHC I*.

Adopting a high-resolution linkage approach, we estimated the recombination distances among all genes of interest, which is vital in the absence of a complete physical map of the region in any urodele. The *MHC* region remains fragmented even in the best currently available urodele genome—the chromosomal scale *Ambystoma mexicanum* assembly (Smith et al. 2019). Although the ca. 11 Mb segment of the *A. mexicanum* genome spanning from *MHC I* to *KIFC1* contains all our non-APGs, *TAPBP*, and *MHC* class II gene, it lacks most *MHC I* genes, *PSMB8*, *PSMB9*, *TAP1*, and *TAP2*, which are scattered over multiple unplaced scaffolds. From our data, we inferred that at least *PSMB8*, *TAP1*, and *TAP2* are duplicated in *A. mexicanum*, which may explain the fragmentation of the assembly. Nonetheless, the *A. mexicanum* assembly confirms the tight linkage among at least some of our genes of interest, suggesting this as an ancestral urodele condition, while showing its large physical size, exceeding 10 Mb, and apparent genomic complexity. To sum up, the tight linkage between APGs and *MHC I* confirmed in *Lissotriton*, and likely to occur also in other genera, meets a condition deemed required for coevolution between APGs and *MHC I* (Kaufman 2015).

Coevolution between APGs and MHC I *in Salamanders*

Despite the tight linkage between APG and *MHC I* in *Lissotriton*, two other patterns detected in salamanders appear at odds with the coevolution hypothesis as currently formulated. First, contrary to the expectation of a single classical *MHC I* gene, multiple polymorphic, highly expressed, apparently classical *MHC I* genes are present in urodele species studied so far (Fijarczyk et al. 2018; Sammut et al. 1999). Also our rough, transcriptome-based assesement, which almost certainly underestimated the number of genes, points to multiple *MHC I* genes in most examined species (supplementary table S7, Supplementary Material online). Second, although pervasive adaptive evolution of APGs was expected, we detected only a weak signal of positive selection, restricted to only some APGs. Still, it cannot be ruled out that coevolution does occur in salamanders, but the process would have to be more complicated than previously thought. Most importantly, mechanisms allowing *MHC I* gene duplication without disrupting coevolved interactions would have to operate. Such mechanisms could be favored by selection as they would remove the constraints in flexibility imposed by having just one highly expressed *MHC I* gene without losing the high efficiency of immune response that coevolution provides. The end result would be adaptive immune response combining the benefits of multiple *MHC I* loci and coevolved combinations of APG–*MHC I* alleles (Kaufman 1999). Whether this is the case in salamanders remains an open question. A similar situation might occur in the rat (*Rattus norvegicus*), where more than one gene is expressed, at least at the mRNA level (Walter 2020), but evidence of coevolution has been demonstrated (Joly et al. 1998).

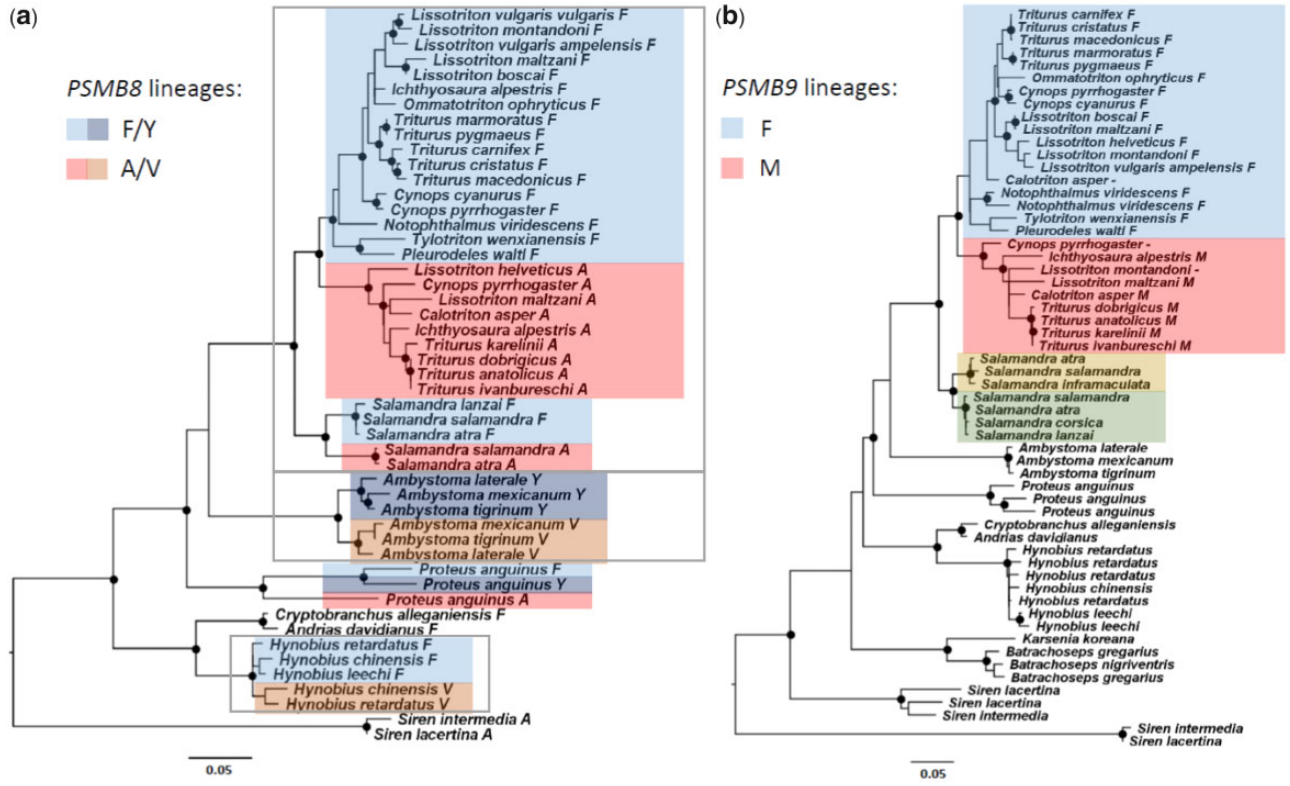

Fig. 3.—Maximum likelihood phylogenetic trees of *PSMB8* and *PSMB9*. Lineages are marked with different colors. Nodes with circles were supported by >80% bootstrap replicates. Potential lineages in the Salamandrinae *PSMB9* were marked in yellow and green. Sequences from four species of *Salamandra* from Rodríguez et al. (2017), which were not used in other analyses, were added to the trees. In *PSMB8*, gray rectangles defined the three families where relationships between species differed substantially from phylogeny, (*a*) PSMB8 lineages and (*b*) PSMB9 lineages.

The slightly higher ω values in APGs compared with non-APGs suggest that the former are less constrained, leaving some space for adaptive evolution to occur, but differences between the two categories of genes are small. APGs themselves are a heterogeneous category, with strong functional links between *PSMB8* and *PSMB9* on the one hand and between *TAP1* and *TAP2* on the other hand. It is, thus, possible that the signal of positive selection detected in *TAP1* and *TAP2* stems from coevolution with *MHC I*, although the two *PSMB8* (and in some taxa *PSMB9*) lineages may still reflect coevolution, but without signal of positive selection detectable with standard tests. In addition, purifying selection appears more pervasive in *PSMB8*, *PSMB9*, and *TAPBP* than in *TAP1* and *TAP2*, at least in humans (Forni et al. 2014). If this also applies to urodeles, it might affect the detectability of their positive selection signal (Anisimova et al. 2001). Further tests of the coevolution hypothesis should examine patterns of co-occurrence of APGs and *MHC I* alleles within individuals as well as confirm the expression of several MHC I molecules at the cell surface. Finally, it remains unclear which evolutionary mechanisms are behind relatively common and predominantly recent duplications of APGs and whether they are related to *MHC I* duplications.

Targets of Positive Selection within APGs

The residues corresponding to codons identified as positively selected might be functionally relevant, due to their interaction with other proteins or their effect on the specificity of antigen processing. The majority of these positions showed enough surface accessibility to interact with other molecules, and some exhibited variation in charge or volume that could be associated with different specificities of such interactions. In fact, two codons identified here as positively selected had previously been recognized for their crucial functional role in the protein. The amino acid at position 284 of our TAP2 alignment (human L266) has been implicated in determining substrate specificity of the protein: changes in this position can alter the epitope repertoire (Lehnert and Tampé 2017). The amino acid at position 104 of the PSMB8 alignment (position 97 in *L. montandoni* and 31 in other tetrapods, Huang et al. 2013) characterizes the two divergent PSMB8 lineages. The two lineages seem to have different specificities contributing to an expanded *MHC I* antigen recognition repertoire and increasing the fitness of heterozygous individuals (Huang et al. 2013). Another three positively selected sites may also be relevant for the functionality of the protein. The amino acid at position 364 of the TAPBP alignment is in the area forming

hydrogen bonds with MHC I (human H334 and H335 Fisette et al. 2016). The codon 323 of the TAP2 alignment is within the biochemically identified substrate-binding region (human 305 within the 301–389 region, Oldham et al. 2016). Finally, the amino acid at position 470 of TAP2 is one of the residues that differentiate the two TAP2 lineages in the rat (rat N452, Ohta et al. 2003) demonstrating its importance in the functionality of the protein. The observed concordance between the signal of selection and the functional significance of these positions points to their importance in the evolutionary fine-tuning of the adaptive immune response. Interestingly, neither of the two *TAP1* codons under positive selection found previously in *Lissotriton* species (Fijarczyk et al. 2018) were confirmed in our study, similar to what was found in mammals and human populations (Forni et al. 2014). This discrepancy might reflect the different kinds of data used for the positive selection analysis and point to differences in selective pressures at different evolutionary scales: the current study used sequences from multiple divergent species, whereas Fijarczyk et al. (2018) used intraspecific polymorphism data. Therefore, there is a pressing need to complement the available data on intraspecific variation in species where coevolution has been inferred, such as chicken and *Xenopus* frogs, by studies of molecular evolution of APGs at phylogenetic scales.

PSMB8 and *PSMB9* Lineages

We found two divergent lineages (i.e., A and F) of *PSMB8* in the Urodela. The two highly supported clades that did not reflect the species phylogeny were present in four of the six families that possess this gene. This mirrors the pattern described by Huang et al. (2013) for several ectothermic vertebrates, in which two ancestral divergent lineages have been maintained for a long time as trans-species polymorphisms, with gene conversion leading to partial sequence homogenization between the lineages. We found a similar pattern in the phylogeny of the Salamandridae for *PSMB9*—close inspection of the sequences allowed us to characterize two divergent lineages in *PSMB9* as well. Similar evolutionary mechanisms might act on both genes because of the tight linkage between them and the interaction of their proteins in the immunoproteasome. Distinct lineages of *PSMB9* have also been described in zebrafish (McConnell et al. 2016). Population data from *Triturus* confirmed the nonrandom association of *PSMB8* and *PSMB9* lineages, *PSMB8* A with *PSMB9* M and *PSMB8* F with *PSMB9* F. This might reflect higher efficiency of the immunoproteasome catalytic subunits encoded by the respective haplotypes, in generating ligands for MHC I proteins encoded on the same haplotype, as suggested by the coevolution hypothesis (Kaufman 2015).

In some taxa, such as sharks, the divergent *PSMB8* lineages are encoded by different genes, whereas in other fish and tetrapods, including the newt *Cynops pyrrhogaster*, these *PSMB8* lineages are alleles of a single locus (Huang et al. 2013; Tsukamoto et al. 2012). However, without detailed genomic-level analysis, it is difficult to rule out their pseudoallele status (i.e., different paralogs lost from different haplotypes with the remaining genes behaving as alleles of a single locus). In our *Lissotriton* mapping population, polymorphisms in *PSMB8* and *PSMB9* segregated as alleles, but only a single lineage was present. In *Triturus* newts, a radiation of closely related species (Wielstra et al. 2019), we obtained data from transcriptome sequencing of all species and data from population-level resequencing from genomic DNA of seven species. In five of the later, we found evidence for a single lineage of both *PSMB8* and *PSMB9*, whereas in *T. dobrogicus* and *T. karelinii*, two lineages were detected. Both *T. dobrogicus* and one *T. karelinii* populations were polymorphic. The gene duplications in these species could not be confirmed in the transcriptomes and the general pattern in *Triturus* seems to be more compatible with a single locus. However, we cannot entirely rule out the possibility that the lineages represent two paralogous genes, with one of them independently lost in some *Triturus* species. Whatever the status of divergent *PSMB8* and *PSMB9* lineages, their distribution across the urodele phylogeny testifies to the existence of selective mechanisms that maintain this polymorphism, possibly in low frequency, for considerable periods of time. The process could be similar to the one described for sticklebacks, where selection sorts standing genetic variation extremely rapidly during adaptation to novel conditions (Laurentino et al. 2020). On the other hand, the relaxation of selection or its change from balancing to directional could also result in the loss of one lineage in certain species.

The apparent loss of *PSMB8* and, in some cases, also *PSMB9* in plethodontids inferred in this study might indicate the presence of an alternative way of cleaving peptides in the endogenous antigen presentation pathway of the adaptive immune response. Indeed, birds lack immunoproteasome and associated genes, including *PSMB8* and *PSMB9*, and it has been assumed that they use constitutive proteasome to cleave peptides (Kasahara and Flajnik 2019).

Conclusions

In conclusion, our study demonstrates, for the first time in the Urodela, the tight linkage between APGs and *MHC I*, which is considered a necessary condition for their coevolution. However, we did not find the pervasive signal of adaptive evolution in APGs, expected under the coevolution hypothesis as a consequence of adaptive evolution of *MHC I*. The APGs nonetheless evolve dynamically, with frequent gene conversion and duplication in several families and gene losses in plethodontids. The lack of a widespread signal of adaptive evolution in APGs and the presence of multiple highly expressed *MHC I* genes indicate that, if coevolution between the two indeed occurs, its mechanism must be flexible enough to allow duplication of *MHC* genes, divergent

lineages of *PSMB* genes or even loss of some APGs. Further insights into the presence and nature of coevolutionary processes in the urodele *MHC* might be obtained by exploring a correlation between genetic variation of APGs and *MHC I* in a comparative framework. There is also a need for studies looking at molecular evolution of APGs in taxonomic groups such as galliform birds, in which coevolution between APGs and *MHC I* was detected using intraspecific data.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The study was funded by the Polish National Science (Grant No. UMO-2016/23/B/NZ8/00738). We thank Magda Migalska for her insightful comments, Mateusz Buczek for his help in script modifications, Samantha Garza, Jong Yoon Jeon, Eunsun Lee, Mi-Sook Min, Iñigo Martinez Solano, Rodrigo Megía Palma, and Kinya Nishimura for their assistance in field work, and Dirk-Jan de Koning, Marian Novotný, and Christos Feidakis for their guidance with some data analysis. Taejoon Kwon, Miguel Vences, and David Weisrock kindly provided unpublished transcriptome sequences and Cera Fisher generated the *Batrachoseps* transcriptome data. We also thank University of Connecticut Research Foundation for support. M.M. was supported by a KAKENHI Grant-in-Aid for Young Scientists (B) (No. JP16K18613) from the Japan Society for the Promotion of Science. M.T. was supported by the Institutional Research Support (Grant No. 260571/2020). B.Wa. was funded by the Ministry of Education (Grant No. 2015R1D1A01057282) and the Ministry of Science, ICT, and Future Planning (Grant No. 2018R1A2B6006833) of the Republic of Korea. M.V. was supported by a grant from Charles University (Grant No. PRIMUS/17/SCI/12). We are indebted with Ministero dell'Ambiente e della Tutela del Territorio e del Mare (prot 56097/T-A31 2017) and Comunidad de Madrid (Ref: 10/150875.9/20 and Parques Regionales Ref: 054.20) for the authorizations for collecting the samples.

Author Contributions

W.B. and G.P. conceived and designed the study. K.D. performed laboratory procedures. G.P. compiled the data and performed analyses of positive selection, protein structure, and intraspecific variation. W.B. carried out the linkage analysis. M.T. and M.V. and P.Z. assisted in computational analyses. E.J., G.F.F., M.M., B.W., B.Wa., J.W.A., and G.P. contributed samples or transcriptomes. G.P. and W.B. led the writing of the manuscript. All authors contributed to the last version and gave final approval for publication.

Data Availability

This article includes supplementary materials, results, figures, tables, and alignments. Transcriptomic data are available on NCBI (https://www.ncbi.nlm.nih.gov): BioProject ID PRJNA681711.

Literature Cited

Abdullayev I, Kirkham M, Björklund ÅK, Simon A, Sandberg R. 2013. A reference transcriptome and inferred proteome for the salamander *Notophthalmus viridescens*. Exp Cell Res. 319(8):1187–1197.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol. 18(8):1585–1592.

Babik W, et al. 2009. Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. Mol Ecol. 18(5):769–781.

Blees A, et al. 2017. Structure of the human MHC-I peptide-loading complex. Nature 551(7681):525–528.

Blum JS, Wearsch PA, Cresswell P. 2013. Pathways of antigen processing. Annu Rev Immunol. 31(1):443–473.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Bos DH, DeWoody JA. 2005. Molecular characterization of major histocompatibility complex class II alleles in wild tiger salamanders (*Ambystoma tigrinum*). Immunogenetics 57(10):775–781.

Bryant DM, et al. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 18(3):762–776.

Burns JA, Zhang H, Hill E, Kim E, Kerney R. 2017. Transcriptome analysis illuminates the nature of the intracellular interaction in a vertebrate-algal symbiosis. Elife 6:e22054.

Che R, Sun Y, Wang R, Xu T. 2014. Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers. PLoS One 9(1):e87940.

Drews A, Westerdahl H. 2019. Not all birds have a single dominantly expressed *MHC*-I gene: transcription suggests that siskins have many highly expressed *MHC*-I genes. Sci Rep. 9:1–11.

Dudek K, Gaczorek T, Zieliński P, Babik W. 2019. Massive introgression of *MHC* genes in newt hybrid zones. Mol Ecol. 28(21):4798–4810.

Elewa A, et al. 2017. Reading and editing the Pleurodeles waltl genome reveals novel features of tetrapod regeneration. Nat Commun. 8:2286.

Farrer RA, et al. 2017. Genomic innovations linked to infection strategies across emerging pathogenic chytrid fungi. Nat Commun. 8:14742.

Ferrington DA, Gregerson DS. 2012. Immunoproteasomes: structure, function, and antigen presentation. Prog Mol Biol Trans Sci. 109:75–112.

Fijarczyk A, Dudek K, Babik W. 2016. Selective landscapes in newt immune genes inferred from patterns of nucleotide variation. Genome Biol Evol. 8(11):3417–3432.

Fijarczyk A, Dudek K, Niedzicka M, Babik W. 2018. Balancing selection and introgression of newt immune-response genes. Proc R Soc B. 285(1884):20180819.

Fisette O, Wingbermühle S, Tampé R, Schäfer LV. 2016. Molecular mechanism of peptide editing in the tapasin–MHC I complex. Sci Rep. 6:19085.

Flajnik MF. 2018. A cold-blooded view of adaptive immunity. Nat Rev Immunol. 18(7):438–453.

Flajnik MF, Kasahara M. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet. 11(1):47–59.

Flajnik MF, et al. 1999. Two ancient allelic lineages at the single classical class I locus in the *Xenopus MHC*. J Immunol. 163:3826–3833.
Forni D, et al. 2014. An evolutionary analysis of antigen processing and presentation across different timescales reveals pervasive selection. PLoS Genet. 10(3):e1004189.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.
Green P, Falls K, Crooks S. 1990. Documentation for CRI-MAP, version 2.4 (3/26/90). St Louis (MO): Washington University School of Medicine.
Huang C-H, Tanaka Y, Fujito NT, Nonaka M. 2013. Dimorphisms of the proteasome subunit beta type 8 gene (*PSMB8*) of ectothermic tetrapods originated in multiple independent evolutionary events. Immunogenetics 65(11):811–821.
Huang Y, Xiong JL, Gao XC, Sun XH. 2017. Transcriptome analysis of the Chinese giant salamander (*Andrias davidianus*) using RNA-sequencing. Genomics Data 14:126–131.
Irisarri I, et al. 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol. 1(9):1370–1378.
Jetz W, Pyron RA. 2018. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat Ecol Evol. 2(5):850–858.
Joly E, et al. 1998. Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. Curr Biol. 8(3):169–180.
Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637.
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(11):4245–4253.
Kasahara M, Flajnik MF. 2019. Origin and evolution of the specialized forms of proteasomes involved in antigen presentation. Immunogenetics 71(3):251–261.
Kaufman J. 1999. Co-evolving genes in *MHC* haplotypes: the "rule" for nonmammalian vertebrates? Immunogenetics 50(3–4):228–236.
Kaufman J. 2015. Co-evolution with chicken class I genes. Immunol Rev. 267(1):56–71.
Kaufman J. 2018. Unfinished business: evolution of the *MHC* and the adaptive immune system of jawed vertebrates. Annu Rev Immunol. 36(1):383–409.
Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22(24):3096–3098.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.
Laurentino TG, et al. 2020. Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. Nat Commun. 11:1–9.
Lehnert E, Tampé R. 2017. Structure and dynamics of antigenic peptides in complex with TAP. Front Immunol. 8:10.
Madison-Villar M, Sun C, Lau NC, Settles ML, Mueller RL. 2016. Small RNAs from a big genome: the piRNA pathway and transposable elements in the salamander species *Desmognathus fuscus*. J Mol Evol. 83(3–4):126–136.
Maghrabi AH, McGuffin LJ. 2017. ModFOLD6: an accuate web server for the global and local quality estimation of 3D protein models. Nucleic Acid Res. 45(W1):W416–W421.
McConnell SC, et al. 2016. Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc Natl Acad Sci USA. 113(34):E5014–E5023.
McElroy KE, et al. 2017. Genome expression balance in a triploid trihybrid vertebrate. Genome Biol Evol. 9(4):968–980.
Miura F, et al. 2010. Transspecies dimorphic allelic lineages of the proteasome subunit *β*-type 8 gene (*PSMB8*) in the teleost genus Oryzias. Proc Natl Acad Sci USA. 107(50):21599–21604.
Müller V, De Boer RJ, Bonhoeffer S, Szathmáry E. 2018. An evolutionary perspective on the systems of adaptive immunity. Biol Rev. 93(1):505–528.
Murata S, Takahama Y, Kasahara M, Tanaka K. 2018. The immunoproteasome and thymoproteasome: functions, evolution and human disease. Nat Immunol. 19(9):923–931.
Murphy K, Weaver C. 2016. Janeway's immunobiology. New York: Garland Science.
Murrell B, et al. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 30(5):1196–1205.
Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8(7):e1002764.
Namikawa C, et al. 1995. Isolation of *Xenopus LMP-7* homologues. Striking allelic diversity and linkage to *MHC*. J Immunol. 155(4):1964–1971.
Niedzicka M, Fijarczyk A, Dudek K, Stuglik M, Babik W. 2016. Molecular inversion probes for targeted resequencing in non-model organisms. Sci Rep. 6:24051.
Nonaka M, Yamada-Namikawa C, Flajnik MF, Du Pasquier L. 2000. Transspecies polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus, *Xenopus*. Immunogenetics 51(3):186–192.
Nourisson C, Muñoz-Merida A, Carneiro M, Sequeira F. 2017. De novo transcriptome assembly and polymorphism detection in two highly divergent evolutionary units of Bosca's newt (*Lissotriton boscai*) endemic to the Iberian Peninsula. Mol Ecol Resour. 17(3):546–549.
Ohta Y, Flajnik MF. 2015. Coevolution of *MHC* genes (*LMP/TAP*/class Ia, *NKT*-class Ib, *NKp30-B7H6*): lessons from cold-blooded vertebrates. Immunol Rev. 267(1):6–15.
Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF. 2006. Ancestral organization of the *MHC* revealed in the amphibian *Xenopus*. J Immunol. 176(6):3674–3685.
Ohta Y, et al. 2003. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (*TAP*) gene in *Xenopus*: further evidence for co-evolution among *MHC* class I region genes. Eur J Immunol. 33(11):3017–3027.
Oldham ML, et al. 2016. A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. Nature 529(7587):537–540.
Paulsson KM. 2004. Evolutionary and functional perspectives of the major histocompatibility complex class I antigen-processing machinery. CMLS: Cell Mol Life Sci. 61(19–20):2446–2460.
Petryszak R, et al. 2016. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res. 44(D1):D746–D752.
Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. 2020. Advances in the evolutionary understanding of *MHC* polymorphism. Trends Genet. 36(4):298–311.
Rancilhac L, et al. Forthcoming 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. Mol Phylogenet Evol. 155:106967.
Rodríguez A, et al. 2017. Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. Mol Phylogenet Evol. 115:16–26.
Rousset F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res. 8(1):103–106.
Salter-Cid L, Nonaka M, Flajnik MF. 1998. Expression of *MHC* class Ia and class Ib during ontogeny: high expression in epithelia and coregulation of class Ia and *lmp7* genes. J Immunol. 160(6):2853–2861.

Sammut B, et al. 1999. Axolotl *MHC* architecture and polymorphism. Eur J Immunol. 29(9):2897–2907.
Schrödinger LLC. 2019. The PyMOL Molecular Graphics System, V2. 0.0. New York (NY): Schrödinger, LLC.
Smith JJ, et al. 2019. A chromosome-scale assembly of the axolotl genome. Genome Res. 29(2):317–324.
Sousounis K, et al. 2015. A robust transcriptional program in newts undergoing multiple events of lens regeneration throughout their lifespan. Elife 4:e09594.
Stuglik MT, Babik W. 2016. Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. Ecol Evol. 6(13):4513–4525.
Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. 2013. Maximum allowed solvent accessibilites of residues in proteins. PLoS One 8(11):e80635.
Tournefier A, et al. 1998. Structure of *MHC* class I and class II cDNAs and possible immunodeficiency linked to class II expression in the Mexican axolotl. Immunol Rev. 166(1):259–277.
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(10):3071–3079.
van Hateren A, et al. 2013. A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. J Biol Chem. 288(45): 32797–32808.
Walker BA, et al. 2011. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci USA. 108(20):8396–8401.
Walter L. 2020. Nomenclature report on the major histocompatibility complex genes and alleles of the laboratory rat (*Rattus norvegicus*). Immunogenetics 72(1–2):5–8.
Weaver S, et al. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 35(3):773–777.
Webb B, Sali A. 2016. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics. 54(1):5.6.1–5.6.37.
Wielstra B, Burke T, Butlin R, Arntzen J. 2017. A signature of dynamic biogeography: enclaves indicate past species replacement. Proc R Soc B. 284(1868):20172014.
Wielstra B, et al. 2017. A genomic footprint of hybrid zone movement in crested newts. Evol Lett. 1(2):93–101.
Wielstra B, McCartney-Melstad E, Arntzen J, Butlin RK, Shaffer HB. 2019. Phylogenomics of the adaptive radiation of Triturus newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle. Mol Phylogenet Evol. 133:120–127.
Yamaguchi T, Dijkstra JM. 2019. Major histocompatibility complex (*MHC*) genes and disease resistance in fish. Cells 8:378.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Yang Z. 2019. Adaptive molecular evolution. In: Balding DJ, Moltke I, Marioni J, editors. Handbook of statistical genomics. Oxford: Wiley. p. 369–396.
Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22(12):2472–2479.

Associate editor: Das Sabyasachi