GBE

MHC Architecture in Amphibians—Ancestral Reconstruction, Gene Rearrangements, and Duplication Patterns

Ke He [1], Wiesław Babik [2], Mateusz Majda[2], and Piotr Minias [3,*]

[1]College of Animal Science and Technology, College of Veterinary Medicine, Key Laboratory of Applied Technology on Green-Eco-Healthy Animal Husbandry of Zhejiang Province, Zhejiang Agriculture and Forestry University, Hangzhou, China

[2]Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

[3]Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Poland

*Corresponding author: E-mail: pminias@op.pl.

Accepted: 02 May 2023

Abstract

The hypervariable major histocompatibility complex (MHC) is a crucial component of vertebrate adaptive immunity, but large-scale studies on MHC macroevolution in nonmodel vertebrates have long been constrained by methodological limitations. Here, we used rapidly accumulating genomic data to reconstruct macroevolution of the MHC region in amphibians. We retrieved contigs containing the MHC region from genome assemblies of 32 amphibian species and examined major structural rearrangements, duplication patterns, and gene structure across the amphibian phylogeny. Based on the few available caecilian and urodele genomes, we showed that the structure of ancestral MHC region in amphibians was probably relatively simple and compact, with a close physical linkage between MHC-I and MHC-II regions. This ancestral MHC architecture was generally conserved in anurans, although the evolution of class I subregion proceeded toward more extensive duplication and rapid expansion of gene copy number, providing evidence for dynamic evolutionary trajectories. Although, in anurans, we recorded tandems of duplicated MHC-I genes outside the core subregion, our phylogenetic analyses of MHC-I sequences provided little support for an expansion of nonclassical MHC-Ib genes across amphibian families. Finally, we found that intronic regions of amphibian classical MHC genes were much longer when compared with other tetrapod lineages (birds and mammals), which could partly be driven by the expansion of genome size. Our study reveals novel evolutionary patterns of the MHC region in amphibians and provides a comprehensive framework for further studies on the MHC macroevolution across vertebrates.

Key words: amphibians, copy number variation, genomic data, major histocompatibility complex, MHC architecture, macroevolution.

Significance

The progress in our understanding of the genomic structure of the major histocompatibility complex region in nonmodel vertebrates has long been limited by methodological constraints. Here, we have taken advantage of rapidly expanding genomic resources to provide a large-scale and comprehensive comparison of MHC architecture across the major evolutionary lineages of amphibians. Our results shed new light on the ancestral architecture, duplication patterns, and structural rearrangements within the amphibian MHC region, laying foundations for the understanding of its macroevolution.

Introduction

The major histocompatibility complex (MHC) is a critical component of the adaptive immune system, and it is considered the most polymorphic gene family within vertebrate genomes (Geraghty et al. 2002). In brief, classical MHC genes encode cell surface glycoproteins, which bind and present peptides either originating from pathogens present in the cytoplasm and nucleus (classical MHC class I genes, MHC-Ia) or peptides from intracellular vesicles, which are in contact with the extracellular space (classical MHC class II molecules, MHC-II) (Garrido and Algarra 2001, but see Joffre et al. 2012 for peptide cross-presentation). The peptides are presented to either cytotoxic CD8 or helper CD4 T cells (by MHC-Ia and MHC-II, respectively), and upon the recognition of nonself peptides, a cascade of immune responses is triggered (Murphy and Weaver 2016).

The evolutionary origin of the MHC gene family and the entire genomic MHC region has been the subject of considerable interest; it is, however, still poorly understood and intensively debated (Kaufman 2018; Ohta et al. 2019; Nakatani et al. 2021). Both MHC classes are found in all major groups of jawed vertebrates, while they are probably completely missing in jawless fish and nonvertebrate chordates (Flajnik 2018; Kaufman 2018). In certain groups of teleost fish, MHC-II genes and the associated pathways have been lost (Star et al. 2011; Malmstrøm et al. 2016; Swann et al. 2020).

Among tetrapods, genomic architecture of the MHC has been best resolved in mammals (Shiina et al. 2017; Abduriyim et al. 2019; Zhou et al. 2021). For example, in humans, the core MHC region is spread over 4-Mb fragment of chromosome 6 and contains classical MHC genes (MHC-Ia and MHC-II), MHC region antigen-processing genes encoding proteins involved in peptide processing and loading (PSMBs, TAPs, and TAPBP), and other MHC region genes (located in the MHC region, but not directly involved in antigen processing, loading, and presentation), which are clustered into three main subregions (MHC class I, class III, and class II) (Shiina et al. 2017). This general pattern of MHC organization is relatively conserved among eutherian mammals (de Sá et al. 2019; Li, Chen, et al. 2019; Plasil et al. 2022), although both MHC-I and MHC-II regions have been subject to some lineage-specific reorganizations. For example, DP (MHC-II) genes have been replaced by DI/DY genes in ruminants (Andersson et al. 1988; Li, Chen, et al. 2019), while DQ and DR genes have been deleted and expanded, respectively, in felines (Yuhki et al. 2003; Plasil et al. 2022). In contrast, a large chromosomal inversion has been inferred to happen in the ancestors of Cetruminantia, dividing MHC class II into two subregions, as shown in both ruminants (Li, Chen, et al. 2019) and cetaceans (de Sá et al. 2019). A recent analysis of monotreme genome assemblies provided insights into the putative ancestral MHC organization in mammals, revealing the presence of MHC-Ia genes within class II subregion and a close physical linkage between classical MHC genes and antigen-processing genes (Zhou et al. 2021). In the marsupial lineage, however, tight linkage between MHC-Ia and antigen-processing genes was not retained, due to a translocation believed to occur independently from a similar evolutionary event in eutherian mammals (Siddle et al. 2009; Krasnec et al. 2015). At the same time, marsupial MHC-II genes underwent an expansion and formed two clusters separated by class III region, as shown in the tammar wallaby *Macropus eugenii* (Siddle et al. 2011).

Macroevolutionary trajectories of MHC region architecture and gene copy numbers in birds were strikingly different when compared with mammals. A single-core region of fewer than 100 kb in length (chromosome 16) was reported in the chicken *Gallus gallus* (Kaufman et al. 1999), and MHC-IIA was found to be located roughly 5.6 cM away on the same chromosome (Salomonsen et al. 2003). This so-called minimal essential MHC is generally well conserved in Galliformes, for example, in the black grouse *Tetrao tetrix* (Wang et al. 2012, 2014). In general, the ancestral avian MHC architecture is compact in size and exhibits a tight linkage between MHC-I, MHC-IIA, and MHC region antigen-processing genes, showing some similarities with the ancestral MHC in marsupial mammals (He, Liang, et al. 2022). However, data from more derived avian groups (e.g., Passeriformes) indicate that the evolution of the entire MHC region in birds was highly dynamic in terms of structural organization and size (Westerdahl et al. 2022), although until recently, this kind of information was highly limited due to methodological constraints. Our recent study demonstrated that long-read third-generation sequencing (TGS) genomes can provide novel insights into the long-term macroevolutionary patterns in the MHC architecture, revealing many previously unknown gene translocation and rearrangement events within the avian MHC (He, Liang, et al. 2022).

Research on the MHC region in nonavian reptiles has been conducted in several species (e.g., Komodo dragon *Varanus komodoensis* (Reed and Settlage 2021), tuatara *Sphenodon punctatus* (Miller et al. 2015; Gemmell et al. 2020), but information on its architecture is incomplete. Recently, genomic analyses of the MHC in two *Anolis* lizards (*Anolis carolinensis* and *Anolis sagrei*) (Card et al. 2022) and the Chinese alligator *Alligator sinensis* (He, Zhu, et al. 2022) provided a detailed characterization of reptile MHC, which contains a single-core MHC region with linked class I and class II subregions. An inversion of the class II subregion has also been reported in reptiles (He, Zhu, et al. 2022).

Amphibians, due to their basal phylogenetic position within tetrapods, are a key taxon for understanding the evolution of MHC architecture in this group. In contrast

to bony fish, whose MHC underwent massive reorganization (e.g., broken linkage between MHC-I and MHC-II in teleosts; Kulski et al. 2002; Kaufman 2018), amphibians may have retained the ancestral and relatively simple organization of the MHC region (Ohta et al. 2006). An analysis of the MHC region in the *Xenopus tropicalis* showed that while the overall genomic architecture is similar to that in placental mammals, it retains such apparently ancestral characteristics as the tight linkage between the MHC-Ia, MHC-II, and genes that process antigens for MHC presentation (Ohta et al. 2006). However, the information on the genomic organization of MHC in amphibians is extremely limited—it was described only in two closely related anurans, *X. tropicalis* (Ohta et al. 2006) and *Xenopus laevis* (Session et al. 2016), and a salamander, the axolotl, *Ambystoma mexicanum* (Schloissnig et al. 2021). Not surprisingly, the MHC architecture in both *Xenopus* frogs is similar despite whole-genome duplication in *X. laevis* (Session et al. 2016). The MHC organization in the axolotl also shares similarities with both *Xenopus* anurans and placental mammals; despite its major expansion, as the entire axolotl MHC region encompasses approximately 100 Mb (Schloissnig et al. 2021). It is worth noting, however, that the location of MHC-I and MHC-II genes within the axolotl MHC genomic region has not been examined in detail.

High genomic complexity of the MHC organization is further exacerbated by the presence of nonclassical MHC class I genes (Adams and Luoma 2013), henceforth, referred to as MHC-Ib. In general, the classical MHC-Ia molecules present antigens to the CD8 T cells with diverse T-cell receptors (TCRs), whereas nonclassical MHC-Ib molecules usually show low polymorphism and expression limited to certain tissues. If MHC-Ib presents antigens, these are peptide or nonpeptide molecules, recognized by specialized, unconventional lymphocytes, rather than by the canonical cytotoxic T cells with diverse TCR repertoires (Mayassi et al. 2021). MHC-Ia molecules exhibit conserved amino acids in positions binding peptide ends (Kaufman et al. 1994). As MHC-Ib molecules often have amino acid substitutions in these positions, the number of conserved amino acids in positions binding peptide ends may be helpful in distinguishing between MHC-Ia and Ib. Only limited information about amphibian MHC-Ib is available (Sammut et al. 1999; Zhu et al. 2014; Palomar, Dudek, Migalska, et al. 2021), with one notable exception. *Xenopus* has a family of more than 20 MHC-Ib genes, which are located outside the MHC region, are highly divergent from the single MHC-Ia gene, but have been well conserved throughout the evolutionary history of the genus (Goyos et al. 2011; Edholm et al. 2014), and are essential for larval immunity (Edholm et al. 2013, 2018).

The progress in the understanding of the genomic organization of amphibian MHC has been hampered by the large size and complexity of their genomes (Gregory 2022), which makes the assembly challenging. Only recently, a combination of TGS technologies and chromosomal-level scaffolding was allowed to produce high-quality reference genomes for all amphibian orders (e.g., Li, Ren, et al. 2019; Rhie et al. 2021; Schloissnig et al. 2021), with the majority of the available assemblies from anurans. The few genomic studies describing amphibian MHC architecture complemented a broad spectrum of research that focused on genotyping specific MHC-I and MHC-II genes or their fragments. For example, targeted genotyping of exons coding for peptide-binding regions of MHC molecules provided support for variation in copy numbers of MHC-I and MHC-II genes across different amphibian taxa (Kiemnec-Tyburczy et al. 2012; Zhang et al. 2015). This is not surprising as the three extant amphibian orders diverged from each other more than 250 million years ago (Ma), while the basal divergence within each order exceeds 100 Ma (Marjanović and Laurin 2014; Hime et al. 2021). Targeted genotyping of key exons cannot, however, provide any information on the structure of noncoding (intronic) regions of MHC-I and MHC-II genes, and so far, we have virtually no comparative data on the evolution of MHC intron size in tetrapods, including amphibians. Despite these limitations, the scant available data suggest that the deep phylogenetic divergence may promote lineage- or species-specific macroevolutionary trajectories in MHC organization across amphibians, but this hypothesis cannot be effectively tested without a large-scale analysis of high-quality genomic data.

Here, we take advantage of available genomic data and reconstruct the macroevolution of the MHC across the major phylogenetic lineages of amphibians. The TGS technologies have been reported to provide a reliable picture of copy number variants, repetitive regions, and structural variants (Merker et al. 2018). TGS-based genomic resources have also been successfully used to reconstruct macroevolution of highly duplicated MHC region in other vertebrate groups (He, Liang, et al. 2022), and thus, we expected that the same approach should be well applicable to amphibian data. Our specific aims included: 1) identification of key characteristics and major evolutionary trajectories of MHC architecture in amphibians; 2) description of lineage-specific rearrangements within the amphibian MHC region; 3) quantification of gene duplication rates at the classical MHC genes (MHC-Ia and MHC-II genes) in amphibians; 4) testing for the widespread presence of multigene MHC-Ib families in amphibians; and 5) examination of intron sizes in amphibian classical MHC genes and their comparison across tetrapod lineages (see steps in fig. 1).

Results

Ancestral MHC Architecture in Amphibians

To infer the ancestral MHC in amphibians, we first used genome assemblies available for the basal amphibian

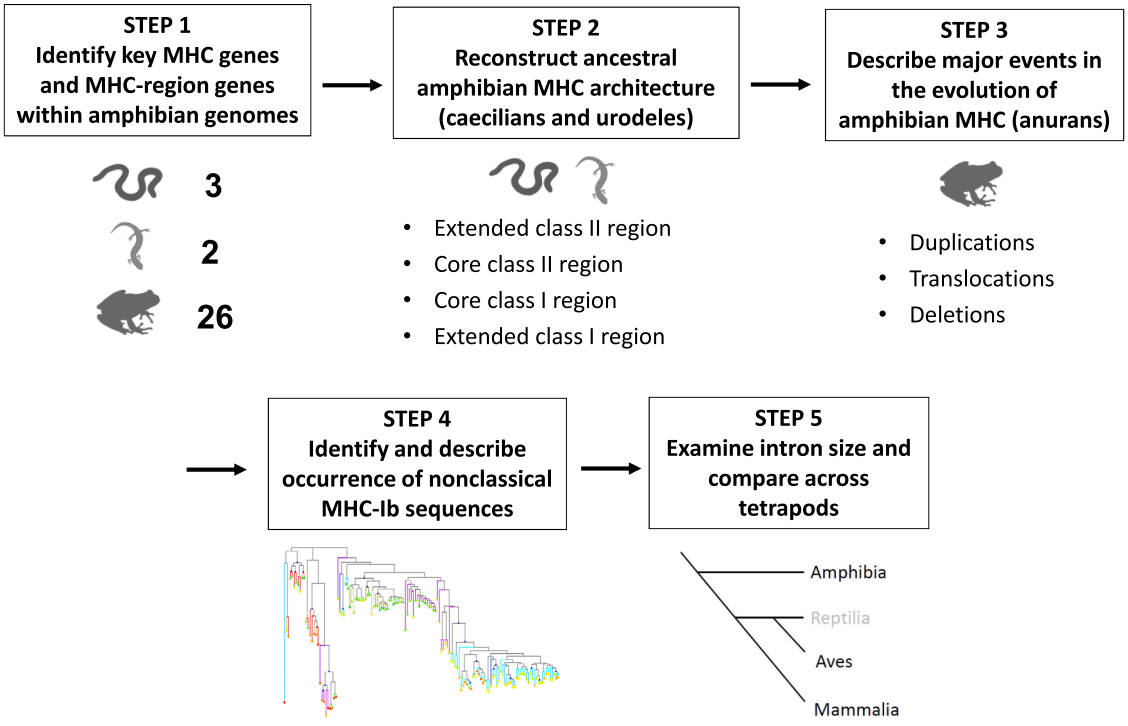

Fig. 1.—Key steps in the reconstruction of amphibian MHC evolution.

lineage (caecilians). Our analysis of caecilian genome assemblies indicated the presence of four well-distinguished MHC subregions (fig. 2). The most detailed physical map of the entire MHC region was obtained for *Microcaecilia unicolor*, in which we defined the following subregions: 1) extended class II subregion (KNSL2∼RXRB, tilde indicates the entire region between the two genes), 2) core class II subregion (MHC-IIA-IIB-BRD2, dash indicates tight linkage), 3) core class I subregion (MHC-I-PSMBs-TAPs), and 4) extended class I subregion (NOTCH∼FLOT1) (fig. 2). Although the core and extended subregions were found on separate scaffolds in the other two caecilian species (*Rhinatrema bivittatum* and *Geotrypetes seraphini*), the analysis of their genome assemblies can be interpreted as providing support for relatively conserved gene arrangement within subregions in caecilians. For example, BRD2 was located in the core class II subregion, while TAPs (TAP1 and TAP2) and PSMBs (PSMB8 and PSMB9) were located in the core class I subregion in all three caecilian species (fig. 2).

Second, we compared the caecilian MHC architecture with that of a more derived amphibian lineage (urodeles). In *A. mexicanum* and *Pleurodeles waltl*, we found a similar architecture of the entire MHC region consisting of four well-distinguished subregions. Gene arrangement was also consistent with caecilians, except for a fragment of extended class I subregion (C4, TCF19, and FLOT1) being directly linked with extended class II region in *A. mexicanum* (fig. 2). Location, orientation, and linkage of MHC region antigen-processing genes and other MHC region genes in *A. mexicanum* generally agreed with patterns reported previously by Schloissnig et al. (2021), although only a fraction of genes from the MHC region were included in our study (over 150 genes were mapped to the MHC region by Schloissnig et al. 2021). In contrast to Schloissnig et al. (2021), we found little evidence for an extensive duplication of MHC-I genes (see results below) in *A. mexicanum*, but this was primarily due to more stringent criteria of gene identification in our study (we only mapped presumably classical genes with complete exons 2–4; supplementary table S1 and supplementary material S1, Supplementary Material online). In *P. waltl*, we observed duplications of PSMB8-PSMB9 and TAP2 genes, while two zinc finger protein genes were located in the core class I region along with MHC-I genes.

Evolution of MHC Architecture in Anurans

The next step of our analyses was to establish whether the putative ancestral MHC architecture (as described in caecilians and urodeles) was retained in anurans. In this step, we focused on 21 anuran genome assemblies, where the linkage between different MHC subregions could be effectively resolved. Six anuran species (including five species with short-read genome assemblies) had MHC subregions on separate (unordered) scaffolds or showed incomplete

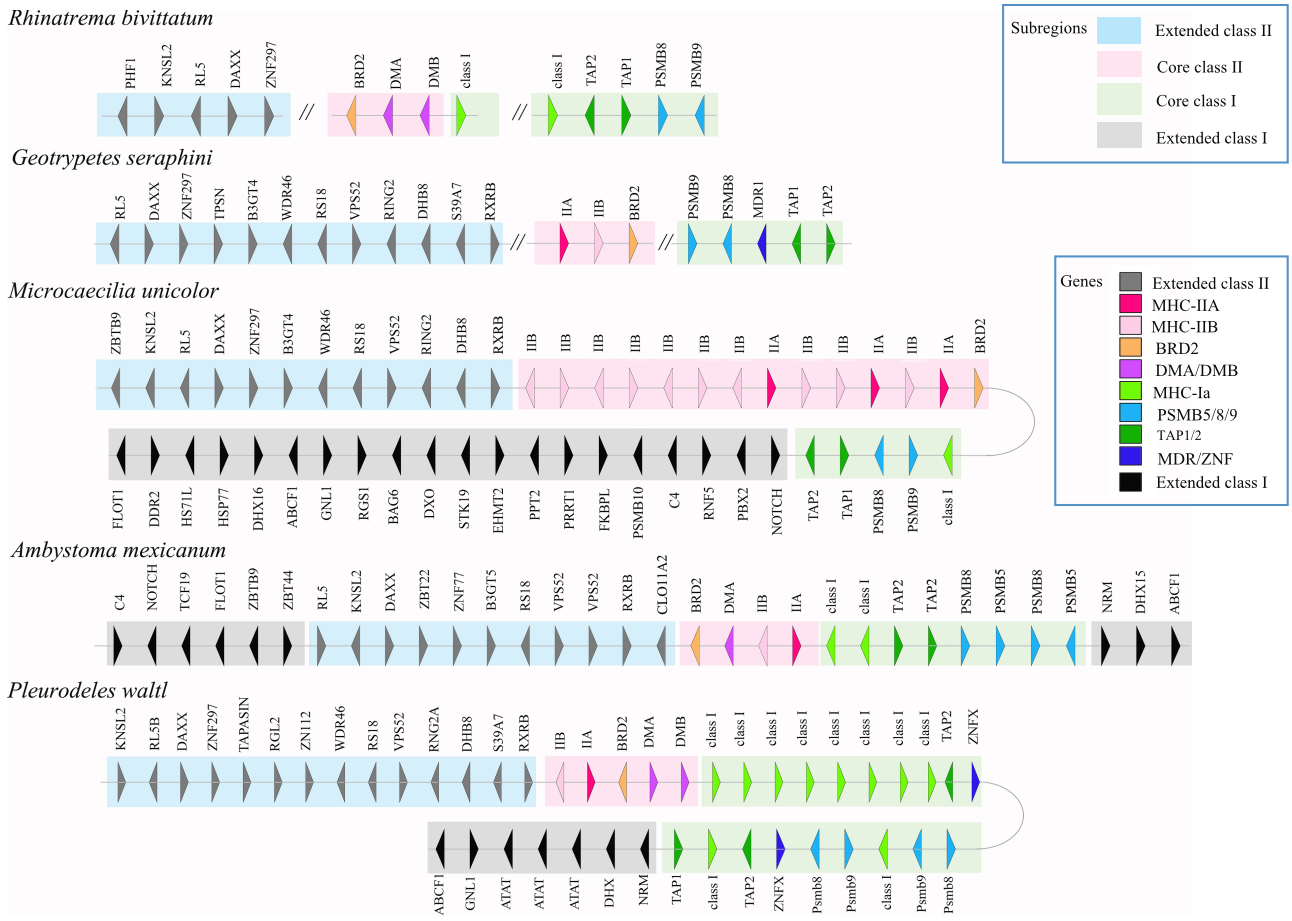

Fig. 2.—The MHC ancestral architecture in caecilian (*Rhinatrema*, *Geotrypetes*, and *Microcaecilia*) and urodele (*Ambystoma* and *Pleurodeles*) amphibians. Four MHC subregions were marked in different colors. Physical distances between genes and subregions were not reflected on the map, double slashes indicate separate scaffolds. Arrows indicate the direction of gene transcription.

core MHC class I or class II subregions and, thus, were excluded from this analysis (table 1). Physical maps constructed for anurans indicated that the ancestral MHC architecture was highly conserved in this lineage (fig. 3) and most MHC region antigen-processing genes and other MHC region genes had conserved location within subregions, for example, TAPs and PSMB family in the core class I subregion; DMA, DMB, and BRD2 in the core class II subregion; DAXX, KNSL2, ZNF297, and RXRB in the extended class II subregion; and ABCF1, TCF19, and C4 in the extended class I subregion (fig. 4). However, we also detected some notable differences between caecilian and anuran MHC architecture; for example, three genes (NOTCH, PBX2, and RNF5) were deleted from the extended class I subregion in anurans, and consequently, C4 gene (instead of NOTCH) marked the boundary between the extended and core MHC class I subregion in this lineage. In most anuran species, both core MHC subregions (class I and II) were adjacent to each other, and their average total span was 1.21 Mb (range: 0.86–3.29 Mb), resulting in more compact core MHC when compared with caecilians and urodeles (2.76 Mb in *Microcaecilia*, 6.19 Mb in *Ambystoma*, and 6.75 Mb in *Pleurodeles*).

Despite the generally conserved architecture of the entire MHC region, each subregion varied in length and in the number of MHC genes (fig. 3 and supplementary table S2, Supplementary Material online). In 9 of 15 anuran species, we found putative MHC-Ia (but not MHC-II) genes to be duplicated outside the core MHC region (either located at the same scaffold outside C4~RXRB region or on separate scaffolds). In most cases, we recorded duplicated tandems of MHC-I genes, without the core class I antigen-processing genes (TAPs and PSMBs) being duplicated alongside. In five species, we detected duplicated MHC-I genes in the proximity of the core region (the same scaffold) but located within the extended class I subregion (*Leptobrachium ailaonicum* and *Leptobrachium leishanense*), in the extended class II subregion (*L. ailaonicum*), in the core class II subregion (*Rhacophorus kio*), or outside the MHC region (*L. ailaonicum*, *Bufo bufo*, and *Bufo*

Table 1

The MHC Architecture in Caecilian (Gymnophiona), Urodele (Urodela), and Anuran (Anura) Amphibians

Order/Family	Species	Extended Class I Subregion (1)	Core Class I Subregion (2)	Core Class II Subregion (3)	Extended Class II Subregion (4)	Linked Subregions	Separated Subregions
Gymnophiona							
Dermophiidae	*Geotrypetes seraphini*	—	(Psmb8-Psmb9-Tap1-Tap2)	(BRD2-MHCII)	(RXRB-ZNF297-DAXX)	—	1, 2, 3
Rhinatrematidae	*Rhinatrema bivittatum*	—	a: (Psmb9-Psmb8-Tap1-Tap2-MHCI) b: MHCI	DMB-DMA-BRD2	(KNSL2-DAXX-ZNF297)	2(b)-3	2(a), 4
Siphonopidae	*Microcaecilia unicolor*	TUBB-ABCF1-C4-NOTCH	Tap2-Tap1-Psmb8-Psmb9-MHCI[a]	BRD2-MHCII	RXRB-ZNF297-DAXX-KNSL2	1-2-3-4	—
Urodela							
Ambystomatidae	*Ambystoma mexicanum*	a: ABCF1 b: TCF19-NOTCH-C4	Psmb5-Psmb8-Tap2-MHCI	MHCII-DMA-BRD2	RXRB-DAXX-KNSL2	1(a)-2-3-4-1(b)	—
Salamandridae	*Pleurodeles waltl*	ABCF1	Tap1-MHCI-Tap2-Psmb8-Psmb9-MHCI-Psmb8-Psmb9-MHCI	DMA/B-BRD2-MHCII	RXRB-ZNF297-DAXX-KNSL2	1-2-3-4	—
Anura							
Pipidae	*Hymenochirus boettgeri*	ABCF1-TUBB-TCF19-C4	Tap2-Tap1-Psmb8-Psmb9-MHCI	MHCII[b]	RXRB-KNSL2-TAPBP-ZNF297	1-2-3-4	—
	Xenopus laevis	ABCF1-TUBB-TCF19-C4	Tap2-Psmb8-Psmb9-MHCI	MHCII	RXRB-Tap1[c]-KNSL2-DAXX-ZNF297-TAPBP	1-2-3-4	—
	Xenopus tropicalis	ABCF1-TUBB-TCF19-C4	Tap2-Psmb8-Psmb9-MHCI	BRD2-MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP	1-2-3-4	—
Scaphiopodidae	*Spea multiplicata*	TCF19	a: MHCI-Tap2; b: Tap1-Psmb8	MHCII	TAPBP-ZNF297	1-2(a)-3-4-2(b)	—
Megophryidae	*Leptobrachium ailaonicum*	MHCI-TCF19-TUBB-ABCF1-MHCI-C4	Tap2-Tap1-Tap2-Psmb8-Psmb9-MHCI-Tap2-Tap1-Psmb8-MHCI[d,e]	BRD2-MHCII	KNSL2-DAXX-ZNF297-TAPBP	1-2-3-4	-
	Leptobrachium leishanense	TCF19-TUBB-ABCF1	a: Tap2-Tap1-Tap2-Psmb8-Psmb9 b: MHCI-Psmb9-Psmb8-Tap1-Tap2	BRD2-MHCII	a: RXRB b: (ZNF297-TAPBP-DAXX)	1[f]-2(a)-3-4-2(b)	4(b)
Myobatrachidae	*Limnodynastes dumerilii*	—	a: MHCI; b: (MHCI-Psmb8-Tap1-Tap2)	BRD2-MHCII	RXRB-KNSL2-DXAA-ZNF297-TAPBP	1-2(a)-3	2(b)
Dendrobatidae	*Oophaga pumilio*	—	(MHCI-Psmb9-Psmb8) (Tap2-Tap1)	—	(TAPBP-ZNF297) (DAXX-KNSL2)	—	—
Eleutherodactylidae	*Eleutherodactylus coqui*	—	(Tap1-Tap2) (Psmb8-Psmb9)	(MHCII)	(ZNF297-KNSL2)	-	2, 3, 4
Leptodactylidae	*Engystomops pustulosus*	ABCF1	MHCI	—	—	1-2	—
Bufonidae	*Bufo bufo*	TCF19-TUBB-ABCF1-C4	MHCI-Tap2-Tap1-Psmb8-Psmb9-MHCI[e]	BRD2-MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP	1-2-3-4	—
	Bufo gargarizans	TCF19-TUBB-ABCF1-C4	MHCI-Tap2-Tap1-Psmb8-Psmb9-MHCI[e]	BRD2-MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP	1-2-3-4	—
	Rhinella marina	-	a: (Psmb9-Psmb8-Tap1-Tap2-	BRD2-MHCII	(TAPBP-ZNF297-DAXX-KNSL2-RXRB)	2(b)-3	4

(continued)

Table 1 Continued

Order/Family	Species	Extended Class I Subregion (1)	Core Class I Subregion (2)	Core Class II Subregion (3)	Extended Class II Subregion (4)	Linked Subregions	Separated Subregions
			MHCI) b: MHCI				
Pyxicephalidae	*Pyxicephalus adspersus*	ABCF1-TUBB-TCF19-C4	MHCI-Tap2-Tap1-Psmb8-Psmb9-MHCI	MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP-BRD2[h]	1-2-3-4	-
Dicroglossidae	*Nanorana parkeri*	a: (ABCF1-TUBB-TCF19) b: (C4-MHCI)[i]	Tap1-Tap2-Psmb8-Psmb9-MHCI	BRD2-MHCII	RXRB-KNSL2-DXAA-ZNF297-TAPBP	2-3-4	1(a), 1(b)
Racophoridae	*Rhacophorus kio*	a: C4 b: ABCF1-TUBB-TCF19	MHCI-Psmb9-Psmb8-Tap1-Tap2-MHCI	MHCII-BRD2-MHCI[j]-MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP	1(a)-2-3-4-1(b)[k]	—
	Rhacophorus dugritei	a: C4 b: (ABCF1-TCF19)	a: MHCI(*9) b: MHCI c: (MHCI-Psmb9-Psmb8-Tap1-Tap2)	BRD2-MHCII	RXRB-KNSL2-DAXX-ZNF297-TAPBP	1(a)-2(a) 2(b)-3-4	1(b), 2(b), 2(c)
Ranidae	*Glandirana rugosa*	(TCF,19-TUBB-ABCF1)	a: (Tap2-Tap1-Psmb8-Psmb9-MHCI) b: MHCI	BRD2-MHCII	(RXRB-KNSL2-ZNF297-TAPBP)	2(b)-3	1, 2, 4
	Lithobates catesbeianus	—	(MHCI-Psmb8)	—	—	—	2
	Rana temporaria	C4-ABCF1-TCF19	MHCI-Tap2-Tap1-Psmb8-Psmb9-MHCI	BRD2-MHCII	a: (DAXX-RXRB-KNSL2-COL11A2-TUBB)[g] b: RXRB-KNSL2-ZNF297-TAPBP	1-4(a)-2-3-4(b)[l]	—

NOTE.—Species excluded from the table had no complete MHC genes or no linked other MHC region genes (*Bombina Variegata*, *Scaphiopus Couchii*, *S. Holbrookii*, *Spea Bombifrons*, and *Pipa Parva*) or only a single subregion (*Platyplectrum Ornatum*). Genes in parentheses indicate that linkage with other subregions was not established (separate scaffolds).

[a]Incomplete MHC-I gene (lacking some of major exons).
[b]BRD2 outside the extended class II subregion.
[c]Tap1 in the extended class II subregion.
[d]Additional MHC-I genes within the extended class II subregion.
[e]Additional MHC-I genes outside the extended class II subregion.
[f]Extended class I subregion in the same scaffold, but distance >40 Mb.
[g]Genes of one subregion in the same scaffold, but distance >40 Mb.
[h]BRD2 within the extended class II subregion.
[i]Additional MHC-I gene linked with C4 in the extended class I subregion.
[j]Additional MHC-I gene within the core class II subregion.
[k]ABCF1-TUBB-TCF19 linked with the extended class II subregion.
[l]The extended class II subregion between core and extended class I subregions.

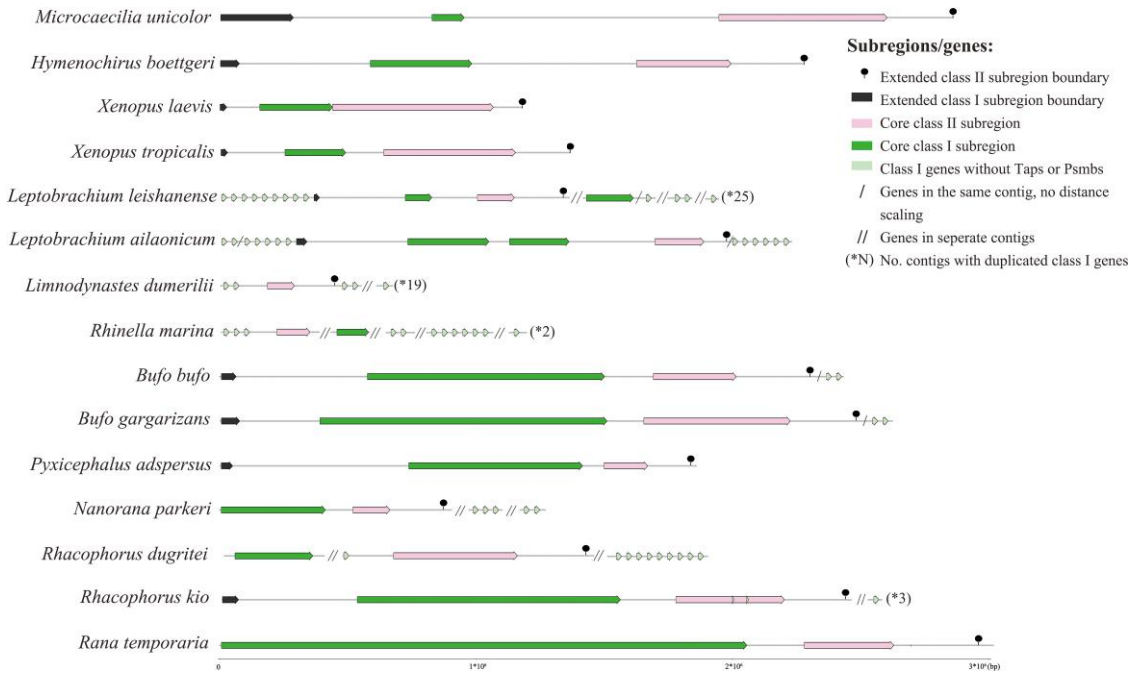

Fig. 3.—Physical maps of the core MHC class I and II subregions in anuran amphibians. *M. unicolor* (Gymnophiona) was used as a reference to an ancestral amphibian MHC architecture. All the subregions were identified based on the location of the MHC, MHC antigen-processing, or MHC-region-related genes targeted by Blast searches, as listed in the Materials and Methods section. C4 and RXRB genes (or neighboring genes in case of deletion events) marked the boundaries of the extended class I and class II subregions, respectively. Arrows indicate the direction of gene transcription. Note: Seven species from table 1 were excluded because of lacking information about the physical linkage between the MHC core and extended subregions (*Eleutherodactylus coqui*, *Engystomops pustulosus*, *Lithobates catesbeianus*, and *Oophaga pumilio*), insufficient information about linkage (*Glandirana rugosa*), or incomplete regions (*Spea multiplicata*).

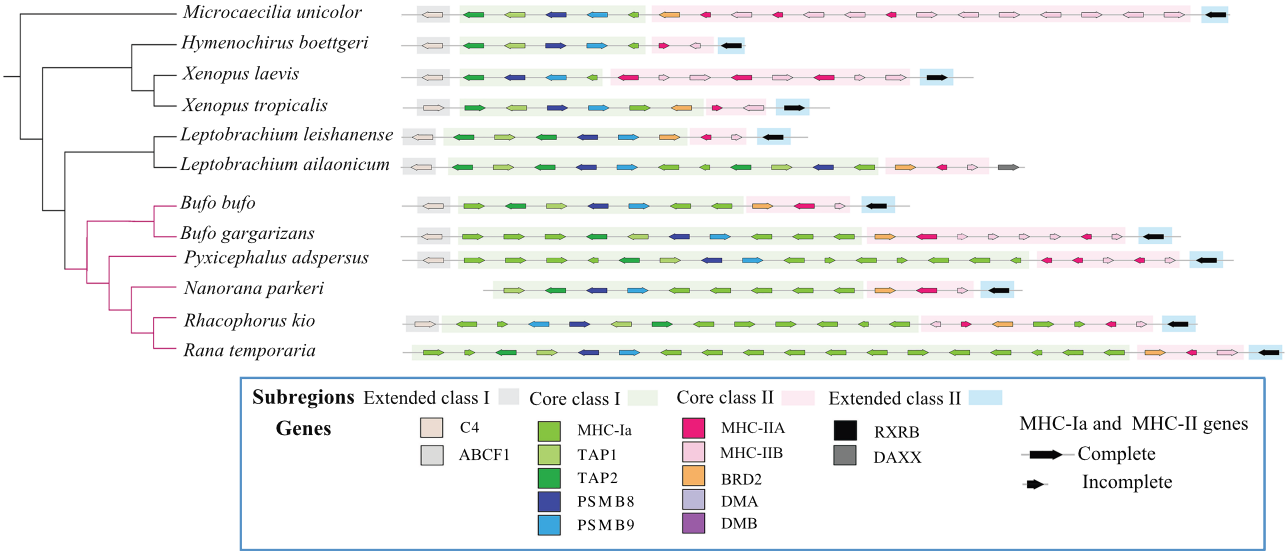

Fig. 4.—Gene arrangement in the core MHC subregions of anuran amphibians. *M. unicolor* (Gymnophiona) was used as a reference to the ancestral amphibian MHC architecture. Molecular distances between genes and subregions were not reflected on the map. Neobatrachia suborder was marked in red in the phylogeny. Arrows indicate the direction of gene transcription. Note: Three species from figure 3 were excluded because of lacking information about the physical linkage between the core class I and class II subregions (*R. marina* and *R. dugritei*) or about the location of MHC-I genes in the core MHC subregions (*Limnodynastes dumerilii*).

gargarizans). In both *Leptobrachium* species, we also found duplications of the entire core class I subregion (including TAPs and PSMBs). In *L. ailaonicum*, two duplicated core class I subregions were adjacent to each other forming a single cluster, while in *L. leishanense*, the duplicated core class I subregions were not adjacent and were located on different scaffolds (fig. 3). The number of conserved amino acids in positions binding peptide ends did not differ between the MHC-I genes within and outside the core region (paired samples Wilcoxon test, $P = 0.44$), so there is no evidence for an increased proportion of MHC-Ib genes outside the core region.

Duplications of presumable classical MHC genes within the core subregions were examined in 11 species. *Limmodynastes dumerilli*, *Rhacophorus dugritei*, and *Rhinella marina* were excluded from this analysis because of incomplete core MHC genes or missing information on physical linkage between core class I and class II subregions (see note in fig. 4 and supplementary tables S3 and S4, Supplementary Material online). We found a specific duplication pattern of the core MHC-I genes in six species from five anuran families (Bufonidae, Dicroglossidae, Pyxicephalidae, Ranidae, and Rhacophoridae), all belonging to a monophyletic clade (suborder) of Neobatrachia (marked purple in the phylogenetic tree in fig. 4). In five of these species (*Nanorana parkeri* excluded due to poor resolution), duplicated MHC-I genes were located in two separate positions within the core class I subregion, including one tandem located between PSMBs and the core class II subregion, and the second tandem located between TAPs and the extended class I subregion (fig. 4). In contrast, duplicated MHC-I genes in the two other anuran families (Megophryidae and Pipidae), which branched off before Neobatrachia, formed a single cluster within the core class I subregion (located between PSMBs and the core class II subregion) (fig. 4). This phylogenetic pattern suggests that the separation of duplicated MHC-I genes into two different clusters likely occurred in the ancestor of modern Neobatrachia after the split of the Megophryidae lineage. We also recorded several duplication events of MHC-II gene clusters within the core subregion, but they did not exhibit any clear phylogenetic pattern and were scattered across the phylogeny, suggesting an independent evolution (fig. 4). At the same time, most of the studied anuran species ($n = 7$ out of 11) had a single MHC-IIA and MHC-IIB gene cluster within the core region (fig. 4).

Except for the MHC-I and MHC-IIA/B genes, we detected infrequent duplication events of other target genes, for example, TAP2 genes were duplicated in Megophryidae (*L. ailaonicum* and *L. leishanense*). Occurrence of these evolutionary events indicates that in spite of generally conserved architecture, the MHC region of anuran amphibians is a hotspot of subtle and lineage-specific genetic reorganizations.

Copy Number of Putative Classical MHC Genes

To assess duplication rates, we quantified copy number of antigen-presenting genes (MHC-I, IIA, and IIB) within the core MHC region (supplementary tables S2 and S4, Supplementary Material online). In the caecilian *M. unicolor*, we recorded extensive duplication of MHC-IIB genes, retrieving ten complete core gene copies. Additional three copies were found outside the core subregion resulting in the total of 13 MHC-IIB copies. We only retrieved incomplete MHC-I (one copy) and MHC-IIA (three copies) genes in *Microcaecilia*, suggesting stronger constraints on duplication. In *Ambystoma*, we found no solid evidence for extensive gene duplications, as we retrieved only two complete MHC-I genes and a single gene copy of MHC-IIA and MHC-IIB. In contrast, we found support for extensive duplications of MHC-I genes in *P. waltl* ($n = 10$ complete copies) but no evidence for the duplication of MHC-IIA or MHC-IIB.

Anuran amphibians showed extensive duplications of MHC-I genes, except for three families, Pipidae, Dendrobatidae, and Scaphiopodidae (one gene copy in the core subregion). Within other anuran families, we recorded at least two MHC-I gene copies in MHC core subregion, and the most extensive duplication of the core MHC-I ($n = 14$ total gene copies including 12 complete ones) was recorded in *R. temporaria* (Ranidae). In anurans, MHC-I genes were often found outside the core MHC, and the highest total number of the MHC-I genes (located either within or outside the core subregion) was recorded in *L. leishanense* ($n = 42$ gene copies including 18 complete ones). The numbers of MHC-I gene copies predicted based on our analyses of anuran genome assemblies were consistently higher when compared with previously published estimates (based on the maximum number of different alleles detected per individual) (fig. 5). The number of MHC-II copies in anurans was much lower when compared with MHC-Ia (except *X. laevis*), ranging from one to six core MHC-IIA and IIB gene copies (including incomplete ones), with no additional copies outside the core subregion in most species.

Nonclassical MHC-I (MHC-Ib)

The phylogeny of the full-length MHC-I α1–α3 domains with the overlaid information about the number of conserved amino acids in positions binding peptide ends is highly informative concerning the evolution of MHC-Ia/Ib in amphibians (fig. 6). A large, well-supported cluster of *Xenopus* MHC-Ib XNC/SNC in the upper part of the tree also contains a single *Hymenochirus* sequence, suggesting that the expansion of this gene family occurred in the lineage leading to *Xenopus* (i.e. before the divergence of *X. laevis* and *X. tropicalis* approximately 40–100 Ma; (Cannatella 2015; Hime et al. 2021). Several presumably

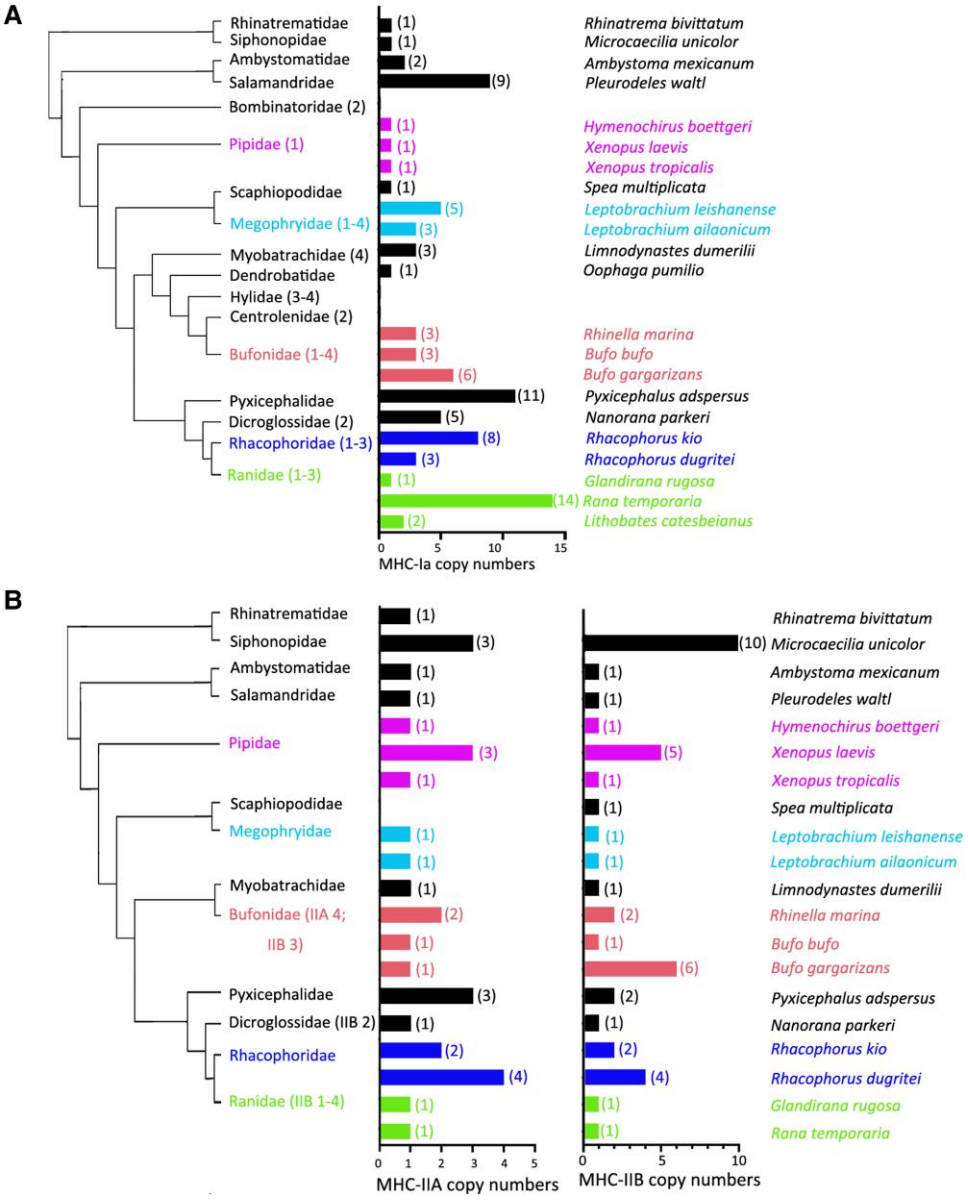

Fig. 5.—Copy number of MHC-I (*A*) and MHC-II (*B*) antigen-presenting genes in caecilian and anuran amphibians. Previously published estimates of copy numbers (based on the maximum number of alleles detected per individual) were shown for each family in parentheses (references listed in supplementary table S6, Supplementary Material online). Estimates from this study (based on the analyses of long-read genome assemblies) were shown for each species with horizontal bars (genes from core subregions only reported). Families with multiple species were marked in nonblack colors.

nonclassical MHC-Ib sequences from ranid frog cluster with the XNC/SNC with weak bootstrap support. The central part of the tree (fig. 6) contains numerous sequences with many conserved amino acids in positions binding peptide ends, which thus likely represent MHC-Ia. These sequences form three clusters, though only one of them, grouping three closely related families Bufonidae, Eleutherodactylidae, and Leptodactylidae, is well supported. The remaining two clusters comprise ranid frogs and *Xenopus*, respectively. The putative MHC-Ia sequences from caecilians form a separate well-supported clade (fig. 6). Among salamanders, the three sequences from the axolotl form a single cluster, which contains both putative MHC-Ia and MHC-b proteins, and all MHC-I sequences from *Pleurodeles* also cluster together. The lower part of the tree (fig. 6) contains several highly divergent sequences with few conserved amino acids in positions binding peptide ends. These divergent sequences, found in both anurans and caecilians, evolve quickly as indicated by long branches connecting them with the rest of the tree.

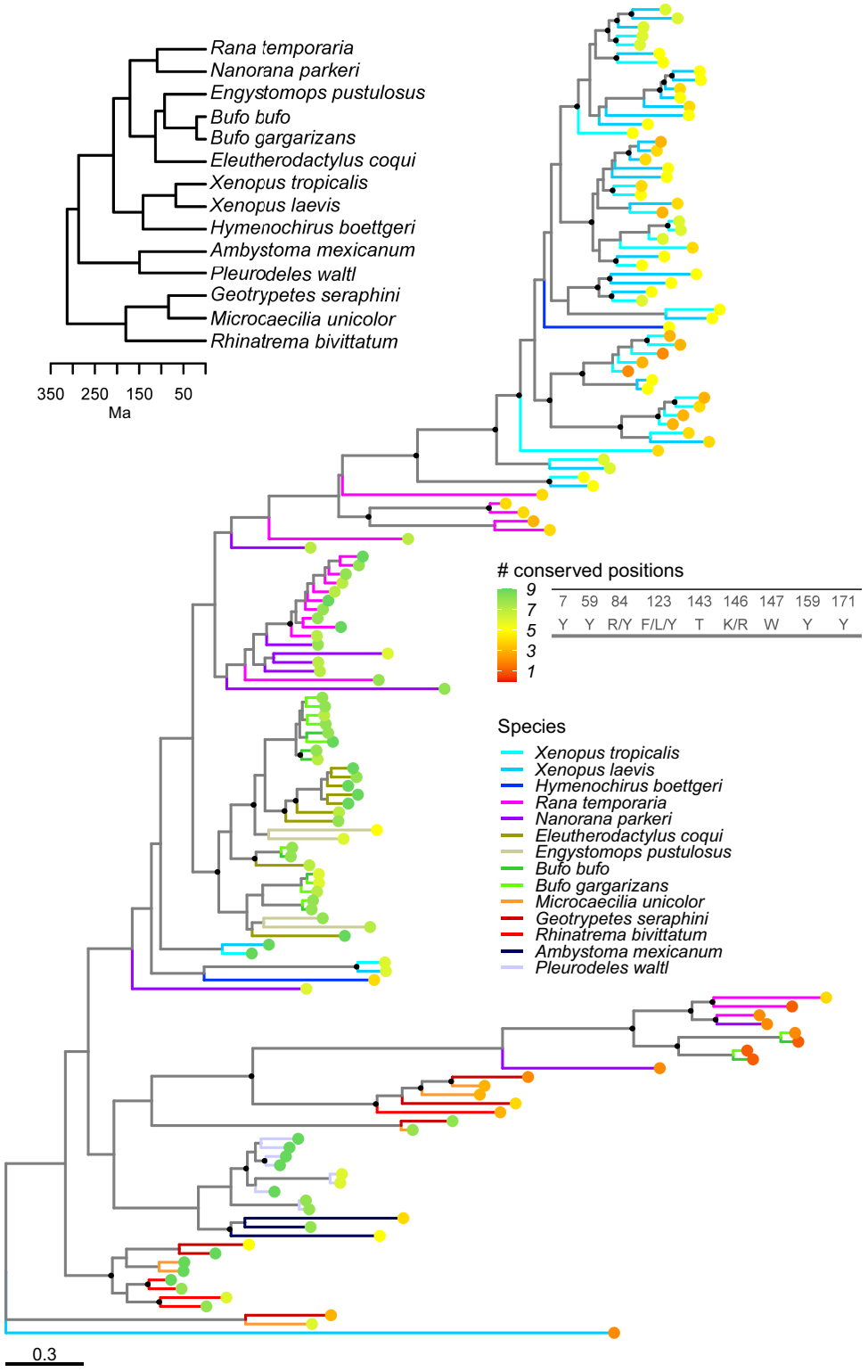

Fig. 6.—The maximum likelihood RAxML tree of full length a1–a3 domains of amphibian MHC-I proteins rooted with the highly divergent *X. laevis* MHC-Ib112 sequence (Ohta et al. 2019). The terminal branches were color-coded according to species. The colored dots at the end of the branches indicate the number of conserved amino acids in positions binding peptide ends (# conserved positions). The positions were numbered according to the HLA-A2 sequence (inset table, see also alignment in supplementary material S3, Supplementary Material online). Black dots indicate nodes with bootstrap support above 70%. Inset shows the time-tree of the included species; Ma, million years ago.

Overall, expansions of MHC-Ib originally described in the *Xenopus* lineage were much less apparent or absent in other amphibian lineages. It is, however, worth noting, that MHC-Ib appears to emerge continuously from classical MHC (Hughes and Nei 1989; Dijkstra et al. 2018). In amphibians, such a process is suggested by the presence of sequences with only five or six conserved amino acids nested deep within the putative classical clusters in both anurans and caecilians (fig. 6).

Intron Length of MHC Antigen-Presenting Genes

The last aim of our study was to examine the size of introns 2 and 3 of the classical MHC genes (putative MHC-Ia, IIA, and IIB) in amphibians and compare it with data available for other tetrapods (birds, mammals, and reptiles; basic statistics presented in supplementary table S5, Supplementary Material online). Median size of all introns selected for this analysis was >1 kb in amphibians. In general, MHC-I introns were longest (median length: 5,643 and 5,934 bp for introns 2 and 3, respectively), while MHC-II introns were considerably shorter, showing less difference between median lengths at MHC-IIB (2,950 and 2,754 bp for introns 2 and 3, respectively) than MHC-IIA (3,819 and 1,635 bp for introns 2 and 3, respectively). There was significant variation in MHC-I intron length between anuran families ($P < 0.001$), and both introns were found to be significantly longer in Ranidae than in most other anuran families, including Megophryidae and four Neobatrachia families (Bufonidae, Dicroglossidae, Myobatrachidae, and Pyxicephalidae) (all $P < 0.05$).

The MHC-I intron size varied significantly between tetrapod lineages ($P < 0.001$), being significantly longer in amphibians than birds and mammals (both $P < 0.001$) (fig. 7). There was no significant difference in MHC-I introns 2 and 3 length in amphibians ($P = 0.93$), although avian intron 2 was significantly longer than intron 3 ($P < 0.001$), while in mammals, we recorded the opposite pattern, with intron 3 being significantly longer than intron 2 ($P < 0.001$). We also found a positive association of MHC-I intron size and genome size across the lineages ($\beta = 0.46 \pm 0.11$; $P < 0.001$). There were significant differences in MHC-II intron lengths between amphibians (longer introns) and both birds and mammals (shorter introns) (both $P < 0.001$) (fig. 7). In amphibians, we recorded no significant differences in intron length between MHC-IIA and IIB ($P = 0.94$), although both avian and mammalian MHC-IIA introns were significantly shorter than MHC-IIB introns ($P = 0.009$ and $P < 0.001$ for birds and mammals, respectively). Finally, we found no significant differences in MHC-IIA/B introns 2 and 3 length in amphibians ($P = 0.08$), while avian and mammalian MHC-IIB intron 2 was significantly longer than intron 3 (both $P < 0.001$). Similar to amphibians, no significant differences were found between MHC-IIA introns 2 and 3 in birds ($P = 1.00$) and mammals ($P = 0.09$). There was no association between MHC-II intron length and genome size across all lineages ($P = 0.78$). Data from reptiles, although not included in formal analyses, provided support for relatively long MHC-I and MHC-IIB intron 2, with median lengths showing intermediate values (2,555 and 2,292 bp, respectively) between amphibians and birds/mammals (supplementary table S5, Supplementary Material online). Median lengths of all other MHC introns in reptiles were considerably shorter (173–523 bp) and roughly similar to the values observed in birds and mammals (supplementary table S5, Supplementary Material online).

Discussion

Ancestral Organization and Evolution

Our study of the genomic architecture and evolution of amphibian MHC primarily focused on anurans because of the availability of numerous genome assemblies from this group. However, the inclusion of the few available caecilian and urodele genomes provided insight into the ancestral architecture of amphibian MHC. Based on these scant data, we can infer that the ancestral MHC region in amphibians was probably simple, with a small number of antigen-presenting genes of both classes and the core MHC-I and MHC-II regions adjacent in the genome. The antigen-processing TAP and PSMB genes were located close to MHC-I they served, enabling coevolution between the interacting partners (Kaufman 2018; Palomar, Dudek, Migalska, et al. 2021), and MHC-IIA was tightly linked to MHC-IIB. The core region in the putative ancestral amphibian MHC region (class I and class II combined) was probably rather short, as its average length of approximately 1.2 Mb estimated for anurans is less than half of the combined length of human core MHC-I and MHC-II regions (ca. 2.7 Mb; Shiina et al. 2017), despite similar or larger size of anuran than mammalian genomes (Gregory 2022). Even the 6.2-Mb core MHC-I-II region in the axolotl and 6.75 Mb in *Pleurodeles* appear modest, considering their extreme genome sizes exceeding 20 Gb. This compact core MHC contrasts with the enormous size of the entire genomic MHC region in the axolotl, which spans approximately 100 Mb (Schloissnig et al. 2021).

A relatively simple organization of the ancestral amphibian MHC was proposed previously based on the analysis of *X. tropicalis* MHC region (Ohta et al. 2006), while the data from birds (He, Liang, et al. 2022), reptiles (Miller et al. 2015; Card et al. 2022; He, Zhu, et al. 2022), and marsupial mammals (Belov et al. 2006) suggest that the simple architecture has generally been maintained throughout the evolutionary history of tetrapods. Major architectural changes, such as the apparent large-scale inversion (Kaufman 2018) that inserted the class III region between class I and class II

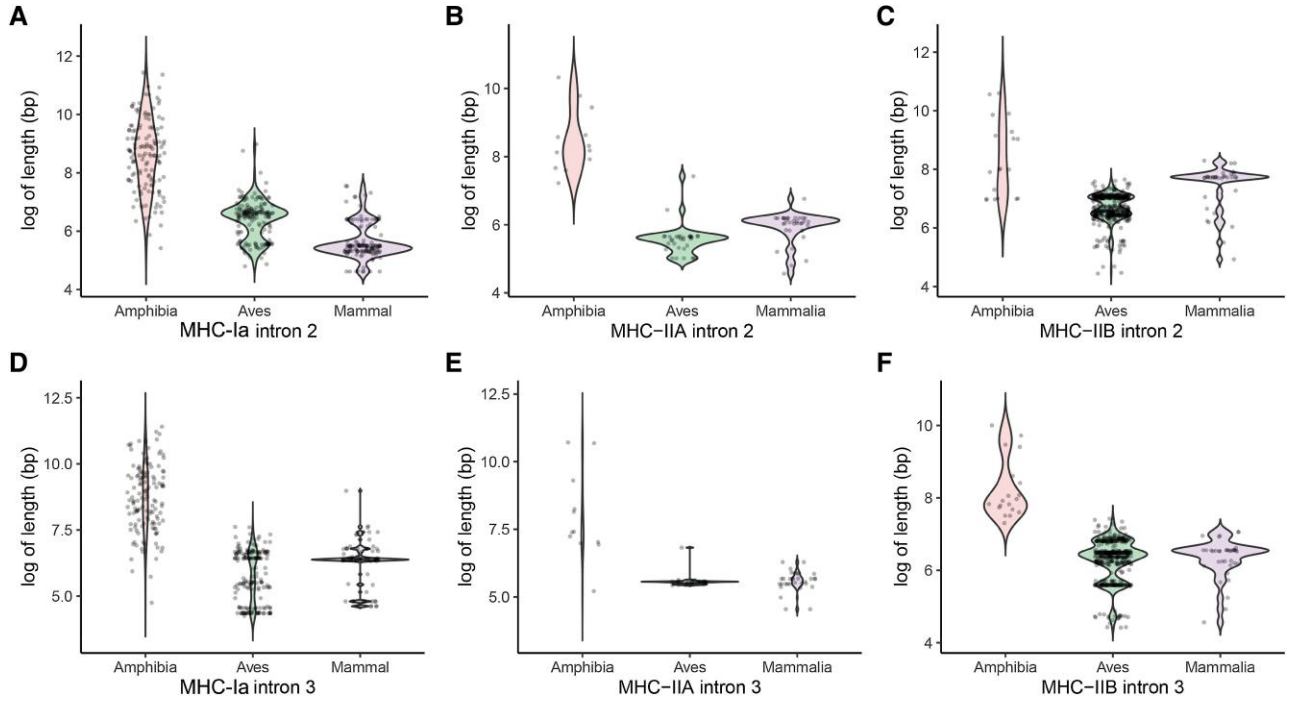

Fig. 7.—Intron length of putative classical MHC-I, IIA, and IIB genes in amphibians and two other tetrapod lineages (birds and mammals). Violin shapes indicate probability density of the data.

and broke the linkage between MHC-Ia and antigen-processing genes in eutherian mammals, or extreme expansions of classical MHC genes and translocations of MHC-IIA or TAP genes outside the core MHC, as seen in some birds (He, Liang, et al. 2022), were probably uncommon. The overall structure of the core MHC region and its gene content have also been well conserved through the evolution of amphibians. The lack of PSMB9 and TAP1 in the axolotl genome assembly is probably the assembly issue, as putative functional full-length coding sequences of both PSMB9 and TAP1 were successfully retrieved from axolotl transcriptome assembly in a recent study of urodele amphibians (Palomar, Dudek, Wielstra, et al. 2021).

While the evolutionary dynamics of the anuran MHC were not characterized by major architectural changes, we found extensive evidence for lineage-specific expansions of presumably classical MHC genes, most notably MHC-I in Neobatrachia, and for the appearance of a variable number of tandemly repeated copies of MHC-I genes outside the core MHC region in several lineages. The core class II region has been more conserved structurally, often retaining single MHC-IIA and MHC-IIB genes and only occasionally experiencing lineage-specific duplications.

Duplications and Copy Number Evolution

Classical MHC genes are known to be hyperpolymorphic, showing the greatest allelic diversity ever reported in vertebrates, with hundreds or even thousands alleles described in some species (e.g., over 16,000 allelic variants reported across three classical MHC-I genes in humans; Robinson et al. 2020). This extraordinary variation is maintained within populations via different forms of balancing selection (reviewed in Radwan et al. 2020), but individual-level variation is primarily determined by duplication processes, as the total number of duplicated genes defines the number of MHC variants that can be expressed within an organism. Our analysis of amphibian genome assemblies revealed an extensive, so far unrecognized variation in the MHC gene copy number across species and major evolutionary lineages.

Although some between-species variation in MHC-I gene copy was apparent across caecilians and urodeles, the evolution of anuran MHC-I clearly proceeded from low copy number in evolutionarily old clades (e.g., Pipidae) toward a rapid expansion in the more derived families. Our analyses revealed up to 12 complete core MHC-I gene copies in *Rana temporaria* (Ranidae), and even more duplicated MHC-I copies were recorded outside the core region in several species (e.g., *L. leishanense*, Megophryidae), similar to some eutherian mammals (Abduriyim et al. 2019). The evolution of MHC-II seems to have been much more constrained, with much less variation in copy number between species and no clear evolutionary trends. In fact, the highest duplication rate was found in a caecilian (ten core MHC-IIB copies), while in anurans, the number of core MHC-IIB gene copies varied between one and six,

with no additional copies outside the core subregion in most species. The evolution of MHC-IIA was even more constrained, as most of amphibian species (11 out of 16) had only a single copy. In some species (e.g., within Racophoridae), MHC-IIA and MHC-IIB genes were duplicated in tandems, resulting in the same copy number of both genes, but more frequently a single MHC-IIA copy was accompanied by multiple duplicated MHC-IIB copies. Although this kind of diversification in the MHC-II duplication patterns has already been described for birds and mammals (Yuhki et al. 2003; He, Liang, et al. 2022), our study of amphibian genomes shows that similar processes may have driven the evolution of MHC-II structure in older tetrapod lineages.

Previous estimates of MHC-I copy number variation in anurans (based on allele counts within individuals) were much lower (one to four gene copies per species, references in supplementary table S6, Supplementary Material online) when compared with the results of our genomic analyses. This difference prevailed even while using conservative protocols, that is, while discarding all incomplete gene copies retrieved from genome assemblies. Consequently, we may conclude that the analyses of high-quality genome assemblies (as conducted in this study) are likely to yield more reliable estimates of MHC gene copy number in nonmodel taxa than traditional approaches. At the same time, we are aware that genomic data may also contain varying level of assembling or annotation errors, especially in structurally complex regions containing numerous duplications of various ages, such as the MHC. For example, the annotation of the axolotl genome presented by Schlossing et al. (2021) revealed an overall synteny with the human MHC, but suggested much more extensive duplication of the core MHC-I genes (11 copies, supplementary data in Schloissnig et al. 2021) than those revealed in our study (only two copies). These discrepancies were caused by the presence of multiple incomplete copies of core MHC-I genes in the assembly (detailed information in supplementary table S1, Supplementary Material online), reflecting the presence of genuine gene fragments or imperfect assembly. Our analysis of axolotl genome also appeared to be more conservative than cDNA cloning approach by Sammut et al. (1999), suggesting the presence of 6–21 MHC-Ia gene copies in this species, likely resulting from sequencing artefacts (recognizing numerous similar variants represented by single clone sequences as true alleles). Although long-read sequencing and recent advances in bioinformatic (genome assembly) approaches enhance reliable resolution of the MHC region and other complex genomic regions with structural variants (Vekemans et al. 2021), we acknowledge that available genomic resource may contain some sequencing and assembly artefacts. For example, two somewhat divergent allele sequences may be possibly merged into a single haplotype containing an apparent duplication (haplotig), when sequencing data originating from heterozygous (outbred) individuals (Guan et al. 2020).

Taking all our results into account, we showed that the evolution of the MHC-I in anuran amphibians has been highly dynamic in terms of duplication processes. This finding is in line with the results of our recent analyses of MHC-I diversity in urodele amphibians, revealing unexpectedly high copy number variation (2–22 gene copies) across 30 species (Minias et al. 2022). In general, extensive between-species differences in the MHC gene copy number are thought to arise through the birth-and-death processes, where new gene copies are created by repeated gene duplication, some being maintained in the genome for long evolutionary times, while the others being deleted or becoming nonfunctional (Nei et al. 1997). Thus, the number of functional classical MHC gene copies is the net differences between birth and death processes, likely evolving in response to varying selection pressures, such as pathogen richness and the rate of pathogen evolution (Bentkowski and Radwan 2019). It remains to be tested whether the same selective forces have driven the evolution of MHC copy number variation in anuran amphibians.

The Evolution of Nonclassical MHC-I

The sequences of putative MHC-Ib were found in all examined species. These included both deep lineages showing the general lack of conservation of amino acids in positions binding peptide ends and sequences more similar to the classical MHC. However, across most anuran lineages, we did not find evidence for strong expansion of MHC-Ib family comparable with that previously described in *Xenopus* (Goyos et al. 2011; Edholm et al. 2014). The demonstration of a key role of *Xenopus* MHC-Ib *XNC4* and *XNC10* genes in larval immunity against mycobacteria and ranaviruses, conferred by unconventional T cells (Edholm et al. 2013, 2018), led to a general hypothesis linking MHC-Ib with the immunity of vertebrates developing from early stages in the aquatic environment. An efficient adaptive immune response requires a large pool of diverse lymphocytes and takes time to mount. Both these requirements may pose a serious challenge for many fish and amphibians that hatch at a very small size, are composed of relatively few cells, and continue their development in the pathogen-rich aquatic environment. The full-fledged adaptive immune response in such small animals may be constrained by the low number of available lymphocytes and the time needed to develop the immune repertoire (Flajnik et al. 1986). It has been hypothesized that the use of unconventional lymphocytes that carry receptors of limited diversity and are restricted by MHC-Ib would be an efficient solution in the face of such constraints (Edholm et al. 2014; Robert and Edholm 2014). Unconventional lymphocytes would confer some

specificity toward pathogens, providing a quick and efficient defense mechanism when the number of lymphocytes is limited. The benefits of such a system may be particularly pronounced in amphibians that undergo metamorphosis, a dramatic event that profoundly remodels their immune system (Rollins-Smith 1998).

If the mechanism proposed by Edholm et al. (2014) and Robert and Edholm (2014) was indeed widespread, we would expect numerous MHC-Ib genes in amphibians with the aquatic larva. We did not, however, find large expansion of MHC-Ib genes in taxa other than *Xenopus*, which leads to at least two possible explanations. First, it is possible that just a handful of MHC-Ib may be sufficient for efficient larval immunity, and additional selection pressures have driven the expansion of *Xenopus* MHC-Ib. Second, the key role of MHC-Ib in larval immunity may be restricted to *Xenopus*, and different mechanisms operate in other amphibians. Regardless of whether any of this explanations hold, the phylogenetic analysis of MHC-I sequences suggests that nonclassical genes may continuously emerge from classical genes in amphibians in response to as yet unknown selection pressures.

Intron Size

Intron size within the MHC has rarely been investigated in nonmodel organisms, especially in the comparative framework. Here, we demonstrated a clear expansion of MHC intron size in amphibians compared with birds and mammals, and we found that this may be partly driven by larger size of amphibian genomes (only for MHC-I). Median intron size was significantly smaller for the MHC-II than MHC-I in amphibians, and we found no association between genome size and MHC-II intron size, suggesting that intronic regions in both MHC classes may have been shaped by different evolutionary forces. In general, a positive association between intron size and genome size is thought to be driven by proliferation of repetitive elements (including transposable elements) in noncoding DNA, and it pertains on a large evolutionary scale across eukaryotes, although its slope may vary considerably between evolutionary lineages (Vinogradov 1999; Elliott and Gregory 2015). For example, tetrapods have longer introns than expected for their genome sizes, and within amniotes, mammals have longer introns than birds and nonavian reptiles, suggesting that evolution of intron size may be adaptively driven by powered flight and metabolic rate (Zhang and Edwards 2012). In contrast, passerine birds were found to have longer MHC-I introns when compared with humans (Arnaiz-Villena et al. 2010), indicating that evolutionary trajectories of the MHC intron size may not well fit into general genome-wide patterns. It requires further investigation whether exceptionally long introns, as we found in the amphibian MHC, are also characteristic for other immune and nonimmune genes in this lineage. Scant empirical data provide little support for this hypothesis, as an average intron size in frogs (across nearly 5,000 genes) was 1.6 and 1.9 times shorter than in mice and humans, respectively (Nowoshilow et al. 2018).

Conclusions

In conclusion, we inferred a relatively simple organization of the ancestral amphibian MHC, with tight linkage between class I and class II. This simple picture has been complicated by lineage specific expansions and translocations, more extensive within class I, resulting in a considerable diversity of the MHC region architectures in anurans. We did not find evidence for an extensive expansion of nonclassical MHC-Ib genes outside the genus *Xenopus*. The intronic regions of amphibian presumably classical MHC genes were much longer when compared with birds and mammals, but these differences can only be partly attributed to the differences in genome sizes between the taxa. These results are based on a still limited number of the available amphibian genomes, mostly focusing on anurans, while salamanders and caecilians were poorly represented. As more amphibian genomes become available in the future, comparative studies should augment the picture presented here with the identification of ecological and life-history traits driving various aspects of the genomic MHC organization in amphibians.

Materials and Methods

Compilation of Genomic Data

Publicly available amphibian genomic data ($n = 32$ species) were downloaded from the Genome Database of the National Center for Biotechnology Information (NCBI, Bethesda, MD, USA) and the China National GeneBank DataBase (CNGBdb, China), as accessed on April 2, 2022 (except for *P. waltl* accessed in October 2022; Brown et al. 2022). Specifically, we retrieved genome assemblies of three caecilian (Gymnophiona), two urodele, and 27 anuran species (supplementary table S3, Supplementary Material online). Most genome assemblies used in this study were based on long-read sequencing (TGS) technologies, primarily PacBio sequencing, yielding good or very good assembly statistics (mean contig N50: 5,836 ± 2,562 kb). The data were complemented with several short-read genome assemblies (supplementary table S3, Supplementary Material online), mostly of moderate quality, but these resources were almost exclusively used to extract short sequences (e.g., for intron size analyses), rather than to infer the architecture of the MHC region. For example, the reconstruction of ancestral amphibian MHC was based entirely on TGS data, while in the analyses of physical maps and gene arrangement within the core

MHC region, we used only three anuran short-read assemblies of relatively high quality (contig N50 > 10 kb). Although MHC architecture of *A. mexicanum* and *X. tropicalis* was previously described based on genome assemblies (Ohta et al. 2006; Schloissnig et al. 2021), we kept these species in our analyses for comparative purposes. If multiple versions of genome assemblies were available per species, we chose the latest version based on long-read TGS sequencing. Genomic data of the mimic poison frog *Ranitomeya imitator* were excluded from the analyses, as we found that the assembly could have been contaminated with mammalian (rat) sequences. This resulted in the final data set of $n = 31$ species.

Searching for Putative Classical MHC Genes, MHC Region Antigen-Processing Genes, and Other MHC Region Genes

In order to retrieve the key presumably classical MHC genes (MHC-I, MHC-IIA, and MHC-IIB), MHC region antigen-processing genes, and other MHC region genes from genome assemblies, we used two sets of query sequences (Step 1 in fig. 1). First, to retrieve genes encoding the three MHC antigen-presenting molecules (MHC-I, MHC-IIA, and MHC-IIB), we: 1) downloaded all available amphibian sequences of targeted genes from the NCBI GenBank Nucleotide Database, 2) aligned the sequences by gene and taxonomic affiliation (order), 3) selected representative sequences (excluded sequences with >80% pairwise similarity), and 4) generated conserved consensus query sequences with 50% consensus threshold (ambiguities retained only if the frequency of the most common residue was <50%). All query sequences for classical MHC genes are provided in supplementary material S2, Supplementary Material online. As no caecilian MHC sequences were available in GenBank, query sequences for this group were generated based on data from urodele species. Our query sequences allowed for effective searches of targeted genes across species, and our previous research in birds has supported the reliability of our protocols (He et al. 2020). Sequence and alignment processing was conducted in Geneious v.2022.1.1 software (Biomatters Ltd., Auckland, New Zealand).

We also used genome assemblies to retrieve a set of MHC region antigen-processing and other MHC region genes, which were used to identify the location of four basic MHC subregions (extended class I, core class I, core class II, and extended class II), following the structure of human MHC as described by Kulski et al. (2002) and Shiina et al. (2017). Using the information on the architecture of the MHC region in humans (Shiina et al. 2017), chicken (Kaufman et al. 1999), and two amphibian species (*A. mexicanum* and *X. tropicalis*) (Ohta et al. 2006; Schloissnig et al. 2021), we initially selected 21 genes for Blast searches in amphibians. One other MHC region gene (MDR1) was a posteriori selected for Blast searches after our preliminary examination of MHC linkage maps in caecilians, giving a total of 22 of MHC region antigen-processing and other MHC region genes. As all these genes are much more conserved than classical (antigen-presenting) MHC genes, we used a single available amphibian sequence per gene for Blast searches. The predicted location of each gene within the MHC region and GenBank accession numbers of query sequences were provided in supplementary table S7, Supplementary Material online.

All BlastN search procedures were conducted with *E*-values < 1e−5. Criteria used to identify highly variable presumably classical MHC genes included 60% similarity with query sequences and 80% coverage of the entire exon length. Genes containing all three major exons (2, 3, and 4), including the ones that encode extracellular domains and are involved in peptide binding, were recognized as complete (otherwise recognized as incomplete). Other exons were not included in our consensus query sequences because of short length (<100 bp) and aligned poorly making them difficult to identify across divergent species. The BlastN search approach focusing on exons 2–4 proved reliable in the previous analyses of avian MHC macroevolution (He et al. 2020; He, Liang, et al. 2022). Localization of MHC region antigen-processing and other MHC region genes within subregions (see the Results section for definition of subregion borders) was identified based on the most complete gene copies, and upon retrieval of multiple gene copies from assemblies, we excluded the ones with incomplete exons from the analyses. In the tetraploid *X. laevis*, we exclusively focused on the chromosome with a better resolution of MHC structure (chr. 8L). All genes and gene fragments retrieved from the genomes were listed in supplementary table S3, Supplementary Material online.

Identification of Nonclassical MHC-I Genes

Nonclassical MHC-I (MHC-Ib) genes usually show low polymorphism and expression limited to certain tissues (Adams and Luoma 2013), which could not be directly evaluated for most of amphibian species investigated here based on the genome assemblies and predicted protein sequences. Therefore, in our assessment of the putative Ia/Ib status, we focused on the conservation of the nine amino acid residues in positions binding peptide ends in the MHC peptide-binding groove (fig. 6; Kaufman et al. 1994; Sammut et al. 1999; Almeida et al. 2021).

MHC-Ib sequences can be highly divergent from MHC-Ia; for example, *Xenopus* MHC-Ib proteins may show only approximately 35% amino acid identity with MHC-Ia (Flajnik et al. 1993; Edholm et al. 2014). Therefore, rather than directly searching the genome assemblies, we searched, where available, the databases of

full-length predicted protein sequences available at NCBI to identify putative MHC homologues. We focused on species with the genome assemblies characterized by scaffold N50 > 500 kb and the available predicted protein sequences, which reduced the data set to 14 species representing six anuran, three caecilian, and two urodele families. The protein sequence databases were searched with BlastP using three MHC-I proteins as queries: *X. laevis* MHC-Ia (NCBI accession: AAA16064), *X. laevis* MHC-Ib XNC1 (AAA16359), and human HLA-A (QKG86484); the *E*-value threshold was set to 1e–10, and other BlastP parameters were kept at their default values. Redundancy in the results was removed by retaining only the longest protein among those with overlapping genomic coordinates. The sequences were aligned with MAFFT, the proteins with incomplete α1–α3 domains were removed, and for the remaining sequences (all trimmed to α1–α3 domains), the number of conserved amino acids in positions binding peptide ends was calculated (the observed range, 1–9).

Phylogeny of the α1–α3 protein sequences was reconstructed in raxmlGUI 2.0 (Edler et al. 2021) using RAxML v. 8.2 (Stamatakis 2014) under the JTT+G amino-acid evolution model as identified by ModelTest-NG v. 0.1 (Darriba et al. 2020). The robustness of the obtained topology was tested with 100 rapid bootstrap replicates.

Architecture of Core and Extended MHC Subregions

We explored amphibian MHC architecture in three major steps. First, we described the ancestral MHC architecture in Gymnophiona (Step 2 in fig. 1) and compared it with the patterns observed in a more derived group of Urodela. In general, Gymnophiona are basal to the sister groups of Urodela and Anura, and the divergence between these two major lineages (Gymnophiona vs. Urodela/Anura) exceeds 300 Ma (Hime et al. 2021; Kligman et al. 2023). Consequently, any similarities in the MHC architecture between Gymnophiona and other lineages (e.g., Urodela) are likely to constitute an ancestral state in amphibians. The MHC antigen-processing and MHC-region related genes in Gymnophiona and Urodela were identified using the online version of the program Maker (http://www.yandell-lab.org/software/mwas.html), taking UniProt/Swiss-Prot as the protein reference. To visualize the structure of the core and extended MHC subregions in these lineages, we extracted the regions 20-kb upstream and downstream of the targeted genes. We plotted gene location using *gggenes*, an extension for the *ggplot2* package (Wickham et al. 2016) in R 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria). Only genes detected in at least two species or those known to be located within the human MHC region were visualized. Second, we reconstructed major evolutionary events, that is, translocations, inversion, and deletions between and within (if possible) major amphibian lineages (Step 3 in fig. 1). We also explored duplication patterns of the putatively classical MHC genes, focusing exclusively on species with complete core MHC class I and II subregions ($n = 14$ species). Phylogeny of our study species was reconstructed based on the VertLife webserver (http://vertlife.org; Jetz and Pyron 2018). We primarily focused on the variation in MHC-I copy numbers across species and families, as duplicated MHC-II tandems had a relatively poor resolution in most genomes.

Intron Size of MHC Antigen-Presenting Genes

We also aimed to examine the structure of classical MHC-I and MHC-IIA/B genes retrieved from amphibian genome assemblies. Here, we focused on the size of key introns (2 and 3) from each gene, as exon size is highly conserved across vertebrates despite some small indels in the regions coding for the peptide-binding groove (e.g., in birds (Minias et al. 2018) and bats (Abduriyim et al. 2019)). We retrieved most information for the MHC-I ($n = 138$ estimates per intron), while data on the MHC-II intron size were much less abundant ($n < 20$ estimates per intron). Thus, the MHC-I intron length was selected for a comparison between seven anuran families, including Megophryidae and the remaining six families from Neobatrachia suborder. Families with one to two measurements of intron lengths were not included in this analysis.

Finally, we aimed to compare the MHC intron size between amphibians and amniote lineages (birds, mammals, and reptiles) using available genomic data from NCBI (mammals and reptiles) and previously published data sets (birds; He et al. 2020). For this purpose, we compiled estimates of MHC intron length in 38 avian and 40 mammalian species. The data on intron size in reptiles were insufficient for any formal comparisons (only one to six estimates originating from one to three species for all introns except for MHC-I intron 3; supplementary table S5, Supplementary Material online) and thus excluded from testing. We also excluded axolotl from these analyses due to extraordinary genome size (28,206.9 Mb) and outlying intron length (up to ca. 86 kb at MHC-II). Due to extensive genomic resources for some mammalian species (e.g., humans), we randomly selected five sequences per species to retrieve data on MHC intron size in this lineage. We also used NCBI database to compile information on genome size for all our study species across all lineages.

Intron length was log-transformed to improve normality and entered as a response variable in general linear mixed models. In the comparisons of MHC-I intron length between anuran families, we entered intron identity (2 vs. 3) and family as fixed factors. Comparisons on intron length between tetrapod lineages were conducted separately for MHC-I and MHC-II genes. Intron identity and lineage

were entered as fixed factors, while gene identity (IIA vs. IIB) was entered as an additional fixed factor in the analysis of MHC-II. Log genome size was entered as a covariate to test whether variation in intron size between lineages can be attributed to whole genome expansion or contraction. Species identity was included as a random factor in each analysis to avoid pseudoreplication resulting from multiple measurements of the same taxa. We also tested for interactions between fixed factors (excluded from the models if nonsignificant), and significance of differences between factor levels (if more than two) was assessed using Tukey post hoc tests. All analyses were conducted using *lme4* R package (Bates et al. 2014).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We thank Dr. Ying Zhu (Southwest Minzu University, China) and three anonymous reviewers for constructive comments on the manuscript.

Author Contributions

Ke He, Wiesław Babik, and Piotr Minias designed this study. All authors analyzed the data. Ke He, Wiesław Babik, and Piotr Minias wrote the initial draft. All authors revised the draft for intellectual content and approved the final manuscript.

Data availability

GenBank numbers for genomic data are presented in Supplementary Material.

Literature Cited

Abduriyim S, Zou DH, Zhao H. 2019. Origin and evolution of the major histocompatibility complex class I region in eutherian mammals. Ecol Evol. 9:7861–7874.

Adams EJ, Luoma AM. 2013. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I–like molecules. Ann Rev Immunol. 31:529–561.

Almeida T, et al. 2021. A highly complex, MHC-linked, 350 million-year-old shark nonclassical class I lineage. J Immunol. 207:824–836.

Andersson L, Lundén A, Sigurdardottir S, Davies CJ, Rask L. 1988. Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. Immunogenetics 27:273–280.

Arnaiz-Villena A, et al. 2010. Songbirds conserved sites and intron size of MHC class I molecules reveal a unique evolution in vertebrates. Open Ornithol J. 3:156–165.

Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. J Stat Softw. 67:1–48.

Belov K, et al. 2006. Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. PLoS Biol. 4:e46.

Bentkowski P, Radwan J. 2019. Evolution of major histocompatibility complex gene copy number. PLoS Comput Biol. 15:e1007015.

Brown T, et al. 2022. Sequencing and chromosome-scale assembly of the giant *Pleurodeles waltl* genome. bioRxiv: 2022.2010.2019.512763. doi: 10.1101/2022.10.19.512763.

Cannatella D. 2015. *Xenopus* in space and time: fossils, node calibrations, tip-dating, and paleobiogeography. Cytogenet Genome Res. 145:283–301.

Card DC, et al. 2022. Structure and evolution of the squamate major histocompatibility complex as revealed by two *Anolis* lizard genomes. Front Genet. 13:979746.

Darriba D, et al. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 37:291–294.

de Sá ALA, et al. 2019. The marine mammal class II major histocompatibility complex organization. Front Immunol. 10:696.

Dijkstra KK, et al. 2018. Generation of tumor-reactive T cells by co-culture of peripheral blood lymphocytes and tumor organoids. Cell 174:1586–1598.

Edholm E-S, et al. 2013. Nonclassical MHC class I-dependent invariant T cells are evolutionarily conserved and prominent from early development in amphibians. Proc Natl Acad Sci U S A. 110:14342–14347.

Edholm E-S, Banach M, Hyoe Rhoo K, Pavelka Jr MS, Robert J. 2018. Distinct MHC class I-like interacting invariant T cell lineage at the forefront of mycobacterial immunity uncovered in *Xenopus*. Proc Natl Acad Sci U S A. 115:E4023–E4031.

Edholm E-S, Grayfer L, Robert J. 2014. Evolution of nonclassical MHC-dependent invariant T cells. Cell Mol Life Sci. 71:4763–4780.

Edler D, Klein J, Antonelli A, Silvestro D. 2021. raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol Evol. 12:373–377.

Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Phil Trans R Soc B. 370:20140331.

Flajnik MF, et al. 1986. Major histocompatibility complex-encoded class I molecules are absent in immunologically competent *Xenopus* before metamorphosis. J Immunol. 137:3891–3899.

Flajnik M, et al. 1993. A novel type of class I gene organization in vertebrates: a large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian *Xenopus*. EMBO J. 12:4385–4396.

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

Garrido F, Algarra I. 2001. MHC antigens and tumor escape from immune surveillance. Adv Cancer Res. 83:117–158.

Gemmell NJ, et al. 2020. The tuatara genome reveals ancient features of amniote evolution. Nature 584:403–409.

Geraghty DE, Daza R, Williams LM, Vu Q, Ishitani A. 2002. Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. Immunol Rev. 190:69–85.

Goyos A, Sowa J, Ohta Y, Robert J. 2011. Remarkable conservation of distinct nonclassical MHC class I lineages in divergent amphibian species. J Immunol. 186:372–381.

Gregory TR. 2022. Animal genome size database. Available at http://www. genomesize.com.

Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36:2896–2898.

He K, Liang C, et al. 2022. Reconstructing macroevolutionary patterns in avian MHC architecture with genomic data. Frontiers Genet. 13: 823686.
He K, Minias P, Dunn PO. 2020. Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. Genome Biol Evol. 13:evaa270.
He K, Zhu Y, et al. 2022. Major histocompatibility complex genomic investigation of endangered Chinese alligator provides insights into the evolution of tetrapod major histocompatibility complex and survival of critically bottlenecked species. Front Ecol Evol. 10: 1078058.
Hime PM, et al. 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Syst Biol. 70:49–66.
Hughes AL, Nei M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci U S A. 86:958–962.
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. 2:850–858.
Joffre OP, Segura E, Savina A, Amigorena S. 2012. Cross-presentation by dendritic cells. Nat Rev Immunol. 12:557–569.
Kaufman J. 2018. Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. Annu Rev Immunol. 36:383–409.
Kaufman J, Milne S, Göbel T, Walker BA, Beck S. 1999. The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923–925.
Kaufman J, Salomonsen J, Flajnik M. 1994. Evolutionary conservation of MHC class I and class II molecules—different yet the same. Semin Immunol. 6:411–424.
Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Lips KR, Zamudio KR. 2012. Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. Heredity (Edinb). 109:146–155.
Kligman BT, et al. 2023. Triassic stem caecilian supports dissorophoid origin of living amphibians. Nature 614:102–107.
Krasnec KV, Sharp AR, Williams TL, Miller RD. 2015. The opossum MHC genomic region revisited. Immunogenetics 67:259–264.
Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H. 2002. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev. 190:95–122.
Li C, Chen L, et al. 2019. A high-density BAC physical map covering the entire MHC region of addax antelope genome. BMC Genomics 20: 479.
Li Y, Ren Y, et al. 2019. Chromosome-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis. Gigascience 8:giz114.
Malmstrøm M, et al. 2016. Evolution of the immune system influences speciation rates in teleost fishes. Nature Genet. 48:1204–1210.
Marjanović D, Laurin M. 2014. An updated paleontological timetree of lissamphibians, with comments on the anatomy of Jurassic crown-group salamanders (Urodela). Hist Biol. 26:535–550.
Mayassi T, Barreiro LB, Rossjohn J, Jabri B. 2021. A multilayered immune system through the lens of unconventional T cells. Nature 595:501–510.
Merker JD, et al. 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet Med. 20: 159–163.
Miller HC, et al. 2015. Major histocompatibility complex genes map to two chromosomes in an evolutionarily ancient reptile, the tuatara *Sphenodon punctatus*. G3 5:1439–1451.
Minias P, Palomar G, Dudek K, Babik W. 2022. Salamanders reveal novel trajectories of amphibian MHC evolution. Evolution 76: 2436–2449.
Minias P, Pikus E, Whittingham LA, Dunn PO. 2018. A global analysis of selection at the avian MHC. Evolution 72:1278–1293.
Murphy K, Weaver C. 2016. Janeway's immunobiology: Garland Science.
Nakatani Y, et al. 2021. Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. Nat Comm. 12:4489.
Nei M, Gu X, Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A. 94:7799–7806.
Nowoshilow S, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. Nature 554:50–55.
Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF. 2006. Ancestral organization of the MHC revealed in the amphibian *Xenopus*. J Immunol. 176:3674–3685.
Ohta Y, Kasahara M, O'Connor TD, Flajnik MF. 2019. Inferring the "primordial immune complex": origins of MHC class I and antigen receptors revealed by comparative genomics. J Immunol. 203: 1882–1896.
Palomar G, Dudek K, Migalska M, et al. 2021. Coevolution between MHC class I and antigen-processing genes in salamanders. Mol Biol Evol. 38:5092–5106.
Palomar G, Dudek K, Wielstra B, et al. 2021. Molecular evolution of antigen-processing genes in salamanders: do they coevolve with MHC class I genes? Genome Biol Evol. 13:evaa259.
Plasil M, Futas J, Jelinek A, Burger PA, Horin P. 2022. Comparative genomics of the major histocompatibility complex (MHC) of felids. Front Genet. 13:829891.
Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. 2020. Advances in the evolutionary understanding of MHC polymorphism. Trends Genet. 36:298–311.
Reed KM, Settlage RE. 2021. Major histocompatibility complex genes and locus organization in the Komodo dragon (*Varanus komodoensis*). Immunogenetics 73:405–417.
Rhie A, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592:737–746.
Robert J, Edholm E-S. 2014. A prominent role for invariant T cells in the amphibian *Xenopus laevis* tadpoles. Immunogenetics 66: 513–523.
Robinson J, et al. 2020. IPD-IMGT/HLA database. Nucleic Acids Res. 48: D948–D955.
Rollins-Smith LA. 1998. Metamorphosis and the amphibian immune system. Immunol Rev. 166:221–230.
Salomonsen J, et al. 2003. The properties of the single chicken MHC classical class II alpha chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. Immunogenetics 55:605–614.
Sammut B, et al. 1999. Axolotl MHC architecture and polymorphism. Eur J Immunol. 29:2897–2907.
Schloissnig S, et al. 2021. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. Proc Natl Acad Sci U S A. 118:e2017176118.
Session AM, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature 538:336–343.
Shiina T, Blancher A, Inoko H, Kulski JK. 2017. Comparative genomics of the human, macaque and mouse major histocompatibility complex. Immunology 150:127–138.
Siddle HV, et al. 2009. MHC-linked and un-linked class I genes in the wallaby. BMC Genom. 10:310.

Siddle HV, et al. 2011. The tammar wallaby major histocompatibility complex shows evidence of past genomic instability. BMC Genom. 12:421.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Star B, et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.
Swann JB, Holland SJ, Petersen M, Pietsch TW, Boehm T. 2020. The immunogenetics of sexual parasitism. Science 369:1608–1615.
Vekemans X, et al. 2021. Whole-genome sequencing and genome regions of special interest: lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. Mol Ecol. 30:6072–6086.
Vinogradov AE. 1999. Intron–genome size relationship on a large evolutionary scale. J Mol Evol. 49:376–384.
Wang B, Ekblom R, Bunikis I, Siitari H, Höglund J. 2014. Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. BMC Genom. 15:180.
Wang B, Ekblom R, Strand TM, Portela-Bens S, Höglund J. 2012. Sequencing of the core MHC region of black grouse (*Tetrao tetrix*) and comparative genomics of the galliform MHC. BMC Genom. 13:553.
Westerdahl H, et al. 2022. The genomic architecture of the passerine MHC region: high repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. Mol Ecol Res. 22:2379–2395.
Wickham H, et al. 2016. ggplot2: create elegant data visualisations using the grammar of graphics. R package v2. Available at https://github.com/tidyverse/ggplot2.
Yuhki N, et al. 2003. Comparative genome organization of human, murine, and feline MHC class II region. Genome Res. 13: 1169–1179.
Zhang Q, Edwards SV. 2012. The evolution of intron size in amniotes: a role for powered flight? Genome Biol Evol. 4:1033–1043.
Zhang W, Luo Z, Zhao M, Wu H. 2015. High genetic diversity in the endangered and narrowly distributed amphibian species *Leptobrachium leishanense*. Integr Zool. 10:465–481.
Zhou Y, et al. 2021. Platypus and echidna genomes reveal mammalian biology and evolution. Nature 592:756–762.
Zhu R, et al. 2014. Extensive diversification of MHC in Chinese giant salamanders *Andrias davidianus* (Anda-MHC) reveals novel splice variants. Dev Comp Immunol. 42:311–322.

Associate editor: Jay Storz