Supplementary Materials

MHC in newts illuminates the evolutionary dynamics of complex regions in giant genomes

W. Babik, K. Dudek, G. Palomar, M. Marszałek, G. Dubin, M. H. Yun, M. Migalska

Supplementary methods

Samples and RNAs. Adult *Pluerodeles waltl* from a laboratory colony in TU Dresden were mated, eggs and larvae developed at 19 °C and 12:12 L:D light regime. The following stages were sampled and preserved in RNAlater: young larvae (stages 38-44 Shi and Boucaut 1995), older larvae (stages 45-47), metamorphosing larvae (stages 48-49) and metamorphs (stage 50). Because *P. waltl* larvae are larger than those of other species, starting from stage 46, we sampled separately also intestine (middle part) and spleen. For all other species we sampled, during the breeding season, two presumably inseminated females from natural populations (Table S3). The females were kept in the laboratory until they deposited eggs, were then euthanized with MS222 and heart, intestine (middle part), liver, lungs, spleen and tail tissue were preserved in RNAlater and separate libraries were prepared for each tissue. Eggs and larvae developed in the laboratory at room temperature under the natural light regime. Larvae were fed ad libitum with *Artemia* nauplii and then with live chironomid larvae. Larvae were euthanized in MS222 and preserved in RNAlater; depending on their size, they were cut into two (anterior/posterior) or three (head/middle/tail) parts, preserved and further processed separately (Table S4).

RNA extraction and Illumina RNA-seq. Total RNA was extracted from tissues preserved in RNAlater using RNAzol® RT (Sigma) or, if only a very small amount of tissues was available, RNeasy kits (Qiagen). RNA-seq libraries were constructed using Novogene Plant and Animal Eukaryotic Strand Specific mRNA (WOBI) approach and sequenced (2 x 150 bp) on the Illumina NovaSeq platform by Novogene.

Full-length transcripts. PacBio HiFi reads were clustered into transcripts using SMRT Tools v. 13.1: lima, isoseq refine and isoseq cluster. Only high-quality (hq) Iso-Seq transcripts generated by this pipeline were used in subsequent analyses. A transcript is labelled hq if: 1) it is supported by full-length non-chimeric (FLNC) reads with both primers and poly(A) tail detected, 2) the consensus sequence has been polished using the underlying FLNC reads until the predicted per-base accuracy ≥ 99% (QV ≥ 30), and 3) it passes internal software filters against artifacts such as chimeras or concatemers.

Identification of MHC transcripts. First, sequences of the axolotl MHC-I (GenBank AAC60108) and MHC-II (XP_069491359.1, AAG42326, XP_069491364.1, XP_069491365.1) proteins were used as queries to identify MHC-I and MHC-II sequences in *P. waltl* reference transcriptome (Matsunami et al. 2019) using TBLASTN v. 2.15 (Camacho et al. 2009). Both axolotl and *P. waltl* MHC protein sequences were used to TBLASTN search the Iso-Seq transcripts from all seven newt species at the E-value threshold of 1e-30. TransDecoder v 5.7.1

(Haas et al. 2013) was used to identify Open Reading Frames (ORF) in Iso-Seq transcripts (both complete and partial ORFs were allowed, single best ORF per transcript). Because a large number of transcripts were identified (ca. 700 in the case of MHC-I) and visual inspection of their protein alignments indicated that sequences were often highly similar but differed by the presence/absence of entire exons, to reduce redundancy we clustered the proteins within each species based on sequence identity. Clustering was performed using Clusterize() from the R package DECIPHER (Wright 2016) at the protein sequence divergence threshold 0.1, ignoring regions containing gaps (penalizeGapLetterMatches = FALSE). The longest protein was selected as a representative of each cluster. Cluster representatives of the minimum length of 240 amino acids (aa) were aligned using DECIPHER::AlignSeqs(). The resulting alignments were visually inspected and sequences lacking more than half of any of α/β 1-3 domains, or showing extreme divergence in a part of the alignment and high similarity to at least one other sequence in the alignment were removed. While such sequences may represent genuine isoforms, they may also be artefacts of a limited-cycle PCR which is a part of Iso-Seq library preparation protocol.

Because some Iso-Seq clusters were represented by few HiFi reads and because some divergent MHC lineages contained representatives of only some species, we suspected that our Iso-Seq dataset may have missed some MHC sequences, especially those poorly expressed in the intestine. Therefore we supplemented the Iso-Seq dataset with sequences of transcripts obtained from de novo assemblies of short RNA-seq reads. Separate assemblies were obtained for each of five individuals per species, including adults, metamorphs, and, for some species, also advanced larvae (Table S2, S4), reads from all individual's RNA-seq libraries were pooled and assembled with Trinity v. 2.15 (Grabherr et al. 2011) using --SS_lib_type RF parameter suitable for our stranded RNA-seq libraries, with other settings set to default values. Trinity assemblies were searched for MHC sequences using TBLASTN with the axolotl and *P. waltl* MHC-I and MHC-II proteins as queries; sequences with hits of min. 90 aa alignment length and E-value < 1e-40 were retained. TransDecoder was used to identify the single best ORF per Trinity transcript and the resulting proteins were clustered together with representative sequences of Iso-Seq clusters. Although the majority of Trinity proteins fell into previously identified clusters, some new clusters which did not include Iso-Seq sequences and had only Trinity sequences of at least 240 aa appeared. Representative (longest) sequences from such clusters were visually examined for the signatures of chimaerism, and those that passed this check were added to the dataset. It is important to note here that Trinity sequences, resulting from de novo assembly of short reads, may not represent true MHC sequences because of the difficulties of reconstructing full-length transcripts of similar genes from multigene families from short reads. We decided to include such sequences because they represent, albeit imperfectly, extra MHC diversity present in individuals or tissues for which Iso-Seq was not available or lowly expressed.

The genomic organisation of the MHC region. For automatic annotation we used the available gene predictions for *P. waltl* (Brown et al. 2025) and performed de novo prediction with BRAKER3 (Gabriel et al. 2024) for the remaining three genomes. Because de novo gene prediction is a complex and computationally intensive task, especially in large genomes, such as those of newts, we limited our prediction to the MHC region(s) (both core and disparate). We added other continuous genomic regions, totalling 950 Mb in *L. helveticus* and *L. vulgaris* and 700 Mb in *T. cristatus*, to reach a sufficient number of well-supported genes required to train prediction algorithms in BRAKER3. The available version of *T. cristatus* genome assembly is already softmasked for repeats, while in the case of *L. helveticus* and *L. vulgaris* assemblies, we performed repeat identification and softmasking using RepeatModeler v. 2.0.6 (Flynn et al. 2020) and RepeatMasker v. 4.1.7 (repeatmasker.org). Both RNA-seq data from target species and vertebrate proteins from the OrthoDB database (Tegenfeldt et al. 2025) were used for gene prediction. Predicted protein sequences were extracted from the genome assemblies using gffread v 0.12.8. (Pertea and Pertea 2020) and annotated using eggNOG-mapper v2 (Cantalapiedra et al. 2021).

Because the initial mapping of cluster representatives to genomes indicated that automatic predictions of MHC genes were in many cases not accurate, MHC-I and MHC-II genes were annotated manually (also when present outside of the core MHC region). We used the following sources of information: i) gene models from automatic prediction, ii) mapping of all RNA-seq libraries from a target species to the genome with hisat2 v. 2.2.1 (Kim et al. 2019), iii) mapping of all cluster representatives (obtained as described above) from the genus to the genome with minimap2 v 2.28 (Li 2018), iv) mapping of all Iso-Seq transcripts from the species to the genome with minimap2. Tracks representing all the data were visualised in IGV v. 2.18.4 (Robinson et al. 2011), coordinates of coding sequences were identified and exported as .bed files. The retrieved sequences were checked for internal stop codons, and those that did not contain them were considered as putative MHC genes, forming the dataset G.

To identify putative MHC pseudogenes or gene fragments, we extracted for each species exons > 100 bp from all annotated MHC genes, and used BLASTN to search the genome at the E-value threshold of 1e-10 with scoring adjusted to divergent sequences (-reward 1 -penalty -1 -gapopen 3 -gapextend 2). Hits with at least 70% sequence identity covering at least 80% of the query sequence located outside manually annotated MHC genes were considered to represent pseudogenes or gene fragments. Complete annotations of the MHC genes and other genes in the MHC region are provided as .gff files in Supplemental Data.

MHC-I polymorphism. We did not analyze polymorphism of MHC-I-like genes as the primers used in previous studied did not amplify them. The alleles from Palomar et al. (2021) and Gaczorek et al. (2023) were BLASTN-mapped to the respective genomes at the E-value threshold of 1e-30. Because the majority of alleles produced multiple hits, we considered as valid all hits within 0.9 of the bitscore of the best hit. Each hit was given the weight 1/n, where n was the number of hits, allowing an allele matching multiple locations in the genome to be counted as a fraction in each, avoiding overestimation of polymorphism for similar genes. The primers used in Palomar et al. (2021) and Gaczorek et al. (2023) were developed from transcriptome data before genome assemblies became available, and thus were not guaranteed to amplify all genes. We therefore screened the primers against all gene sequences to identify potential mismatches in binding sites. For *L. vulgaris* and *T. cristatus*, we designed additional primers (Table S9), combined them with those previously published, and re-genotyped MHC-I in 24 individuals per species. PCR, library preparation, sequencing, and genotyping followed Palomar et al. (2021) and Gaczorek et al. (2023). In *L. vulgaris*, genotypes from the original and updated primer sets were nearly identical (2/125 alleles, 1.5%, were new). In *T. cristatus*, 17/126 alleles (13.5%) were new, yet estimates of polymorphism from the two primer sets were extremely highly correlated (R^2 = 99.8%; Fig. S9). For *P. waltl* we verified, using newly designed primers (Table S9), that *Plewa-06* and *Plewa-08* genes that previously did not amplify are not polymorphic. We conclude thus that our assessment of polymorphism of the MHC-I genes annotated in newt genomes is accurate.

Phylogenies of MHC proteins. The G+T dataset protein sequences, together with the axolotl or *Andrias* as outgroups, were aligned with DECIPHER::AlignSeq(), separately for MHC-I, MHC-IIA, MHC-IIB, MHC-DMA, MHC-DMB, as well as for MHC-I α1, α2, and α3 domains. The amino-acid evolution models were identified by ModelTest-NG v. 0.1.7 (Darriba et al. 2020); because for all alignments JTT + G4 or similar models were selected, we decided to use JTT + G4 model for all phylogenetic analyses. Maximum likelihood (ML) phylogenies were reconstructed in RAxML-NG 1.2.2 (Kozlov et al. 2019), and robustness of the obtained topologies was tested with 100 bootstrap replicates.

MHC-I and MHC-II expression through ontogeny and across tissues. Expression was estimated by mapping RNA-seq libraries to two types of references: (i) for the four species with the available genome assemblies, RNA-seq reads were mapped to genomes with hisat2 and reads mapped to the annotated MHC genes, pseudogenes and gene fragments were counted with featureCounts v. 2.0.8 (Liao et al. 2014) – this analysis provided detailed expression information for MHC sequences present in reference genomes, (ii) for all seven species, RNA-seq reads were mapped to the cluster representatives using Bowtie2 v. 2.5.1 (Langmead and Salzberg 2012) and the number of reads mapped to each reference was calculated with samtools idxstats – this analysis provided expression estimates for all full length MHC coding sequences identified in our samples. Expression was estimated separately for each library as Fragments Per Kilobase of transcript length per Million reads mapped (FPKM); in the analysis of all seven species we assumed mapping rate of 0.895 as this was the average for the four species with the available genome assemblies.

Gene tree – species tree reconciliation. The analysis was performed using Notung v. 3.0_25 beta (Durand et al. 2006) on two datasets that used different gene trees: (i) G, and (ii) G+T. Notung analyses used ML trees constructed using protein sequences with RAxML-NG (see above). Notung bootstrap edge weight threshold parameter, which indicates which branches are "weak", i.e., weakly supported, was set to 50, the cost of duplication and cost of loss were kept at the default values of 1.5 and 1.0. After reconciliation tree was adjusted to further optimise the number of duplications and losses by rearranging "weak" edges, that can be swapped in their placement without strongly contradicting the data.

Testing for gene conversion. Gene conversion was tested using coding DNA sequences of the G+T dataset after removing sequences assembled by Trinity from short RNA-seq reads, as some of them could have been computationally generated chimaeras that may produce a false recombination signal. The analyses were performed separately for MHC-I and MHC-I-like genes, because high sequence divergence between them makes conversion unlikely and because including numerous sequences would reduce the power of the test by inflating the number of pairwise comparisons that would necessitate stringent correction for multiple testing. MHC-I-like sequences from all species were analysed together due to their relatively small number. MHC-I sequences from each species were analysed separately as we were mostly interested in species-specific gene conversion events. The analyses were performed using geneconv v. 1.81a (Padidam et al. 1999) and were run both allowing no mismatches within conversion tracts ($g = 0$) and using the more permissive setting allowing for some mismatches ($g = 1$) and thus enabling detection of older conversion events.

Selective pressures acting on MHC-I and MHC-I-like genes. Tests of selective pressures were performed applying methods implemented in HyPhy v. 2.5.8 (Kosakovsky Pond et al. 2020), using protein-guided codon alignment of the G+T dataset obtained with DECIPHER:: AlignTranslation() and approximately ML phylogeny inferred with FastTree v. 2.1.11 (Price et al.

2010). RELAX (Wertheim et al. 2015) was used to test whether a change in the strength of selection occurred (i) along the long branches separating MHC-I-like1 and MHC-I-like2 from other sequences in MHC-I phylogeny, (ii) within MHC-I-like1 and MHC-I-like2 clusters. FUBAR (Murrell et al. 2013) was used to identify codons under positive/purifying selection, separately for MHC-I, MHC-I-like1 and MHC-I-like2. FUBAR analysis was limited to α1–α3 domains.

Structural modelling. Genomic sequences of MHC-I-like1 molecules were obtained for *L. vulgaris* and *P. waltl*, while MHC-I-like2 sequences were identified in *L. helveticus*, *L. vulgaris*, and *T. cristatus*. A putative *P. waltl* MHC-I-like2 locus was also identified on a different chromosome than the core MHC region. This locus contained an ORF spanning 20 exons, several of which appeared to encode domains atypical of MHC molecules. Given its likely chimeric and non-functional nature, this gene was excluded from structural modelling. Signal peptides were predicted using SignalP 6.0 (Teufel et al. 2021) and removed prior to structural modelling. MHC-I-like2 molecules contained short N-terminal sequences resembling leader peptides; however, not confidently identified by SignalP. Therefore, these regions were manually trimmed based on sequence alignment with classical MHC-I molecules. Genomic sequences of all MHC-I-like molecules and one putative classical molecule per species (*L. helveticus*: *Lihe-01*, *L. vulgaris*: *Livu-03*, *T. cristatus*: *Trcr-01*, *P. waltl*: *Plwa-01*), together with corresponding β2m, were modelled with the Alphafold3 Server (Abramson et al. 2024; default settings, access April 2025). Structures were visualized with UCSF ChimeraX 1.9 (Meng et al. 2023). A search for structural homology was performed with DALI server (Holm 2022).

Supplementary References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**: 493–500.

Almeida T, Ohta Y, Gaigher A, Muñoz-Mérida A, Neves F, Castro LFC, Machado AM, Esteves PJ, Veríssimo A, Flajnik MF. 2021. A highly complex, MHC-linked, 350 million-year-old shark nonclassical class I lineage. *J Immun* **207**: 824–836.

Brown T, Mishra K, Elewa A, Iarovenko S, Subramanian E, Araus AJ, Petzold A, Fromm B, Friedländer MR, Rikk L. 2025. Chromosome-scale genome assembly reveals how repeat elements shape non-coding RNA landscapes active during newt limb regeneration. *Cell genomics* **5**. doi: 10.1016/j.xgen.2025.100761.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 1–9.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* **38**: 5825–5829.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol* **37**: 291–294.

Durand D, Halldórsson BV, Vernot B. 2006. A Hybrid Micro–Macroevolutionary Approach to Gene Tree Reconstruction. *J Comput Biol* **13**: 320–335.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**: 9451–9457.

Gabriel L, Br\uuna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research* **34**: 769–777.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* **29**: 644–652.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li BO, Lieber M. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**: 1494–1512.

Holm L. 2022. Dali server: structural unification of protein families. *Nucl Acid Res* **50**: W210–W215.

Kaufman J, Salomonsen J, Flajnik M. 1994. Evolutionary conservation of MHC class I and class II molecules - different yet the same. *Sem Immun* **6**: 411–424.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotech* **37**: 907–915.

Kosakovsky Pond SL, Poon AF, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A. 2020. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol* **37**: 295–299.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**: 357–359.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Matsunami M, Suzuki M, Haramoto Y, Fukui A, Inoue T, Yamaguchi K, Uchiyama I, Mori K, Tashiro K, Ito Y. 2019. A comprehensive reference transcriptome resource for the

Iberian ribbed newt *Pleurodeles waltl*, an emerging model for developmental and regeneration biology. *DNA Research* **26**: 217–229.

Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF CHIMERAX : Tools for structure building and analysis. *Protein Sci* **32**: e4792.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* **30**: 1196–1205.

Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**: 218–225.

Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Research* **9**: ISCB-Comm.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **5**: e9490.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotech* **29**: 24–26.

Sammut B, Du Pasquier L, Ducoroy P, Laurens V, Marcuz A, Tournefier A. 1999. Axolotl MHC architecture and polymorphism. *Eur J Immunol* **29**: 2897–2907.

Shi DL, Boucaut JC. 1995. The chronological development of the urodele amphibian *Pleurodeles waltl* (Michah). *Int J Dev Biol* **39**: 427–441.

Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. 2025. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucl Acid Res* **53**: D516–D522.

Teufel F, Armenteros JJA, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Von Heijne G, Nielsen H. 2021. SignalP 6.0 achieves signal peptide prediction across all types using protein language models. *BioRxiv* 2021–06.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* **32**: 820–832.

Wright ES. 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. *R Journal* **8**: 352–359.

Supplementary Tables

Are in a separate Excel workbook.

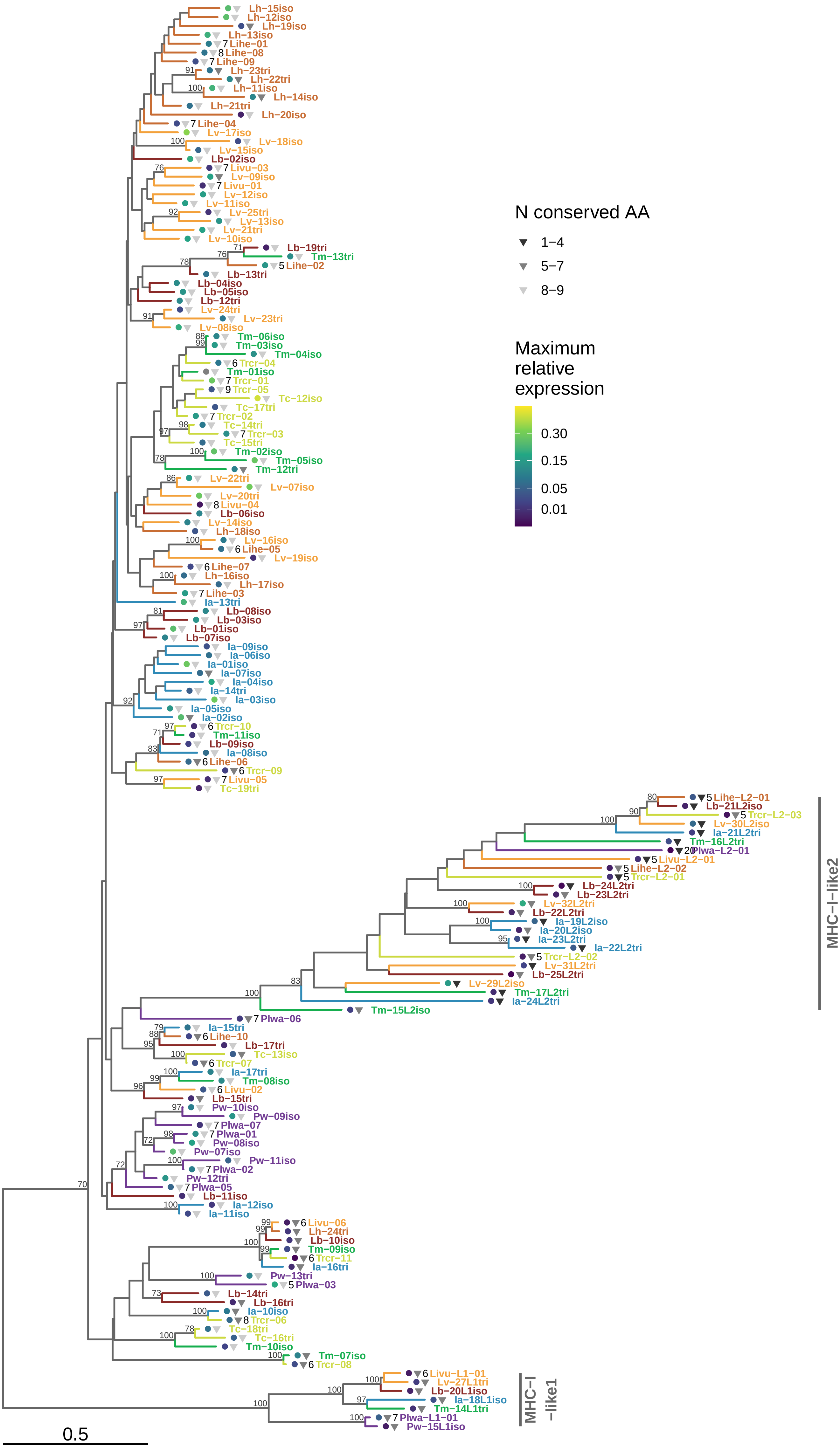

Lh-15iso
Lh-12iso
Lh-19iso
Lh-13iso
7 Lihe-01
8 Lihe-08
7 Lihe-09
Lh-23tri
Lh-22tri
Lh-11iso
Lh-14iso
Lh-21tri
Lh-20iso
7 Lihe-04
Lv-17iso
Lv-18iso
Lv-15iso
Lb-02iso
7 Livu-03
Lv-09iso
7 Livu-01
Lv-12iso
Lv-11iso
Lv-25tri
Lv-13iso
Lv-21tri
Lv-10iso
Lb-19tri
Tm-13tri
5 Lihe-02
Lb-13tri
Lb-04iso
Lb-05iso
Lb-12tri
Lv-24tri
Lv-23tri
Lv-08iso
Tm-06iso
Tm-03iso
Tm-04iso
6 Trcr-04
Tm-01iso
7 Trcr-01
9 Trcr-05
Tc-12iso
Tc-17tri
7 Trcr-02
Tc-14tri
7 Trcr-03
Tc-15tri
Tm-02iso
Tm-05iso
Tm-12tri
Lv-22tri
Lv-07iso
Lv-20tri
8 Livu-04
Lb-06iso
Lv-14iso
Lh-18iso
Lv-16iso
6 Lihe-05
Lv-19iso
6 Lihe-07
Lh-16iso
Lh-17iso
7 Lihe-03
Ia-13tri
Lb-08iso
Lb-03iso
Lb-01iso
Lb-07iso
Ia-09iso
Ia-06iso
Ia-01iso
Ia-07iso
Ia-04iso
Ia-14tri
Ia-03iso
Ia-05iso
Ia-02iso
6 Trcr-10
Tm-11iso
Lb-09iso
Ia-08iso
6 Lihe-06
6 Trcr-09
7 Livu-05
Tc-19tri
5 Lihe-L2-01
Lb-21L2iso
5 Trcr-L2-03
Lv-30L2iso
Ia-21L2tri
Tm-16L2tri
20 Plwa-L2-01
5 Livu-L2-01
5 Lihe-L2-02
5 Trcr-L2-01
Lb-24L2tri
Lb-23L2tri
Lv-32L2tri
Lb-22L2tri
Ia-19L2iso
Ia-20L2iso
Ia-23L2tri
Ia-22L2tri
5 Trcr-L2-02
Lv-31L2tri
Lb-25L2tri
Lv-29L2iso
Tm-17L2tri
Ia-24L2tri
Tm-15L2iso
MHC-I-like2
7 Plwa-06
Ia-15tri
6 Lihe-10
Lb-17tri
Tc-13iso
6 Trcr-07
Ia-17tri
Tm-08iso
6 Livu-02
Lb-15tri
Pw-10iso
Pw-09iso
7 Plwa-07
7 Plwa-01
Pw-08iso
Pw-07iso
Pw-11iso
7 Plwa-02
Pw-12tri
7 Plwa-05
Lb-11iso
Ia-12iso
Ia-11iso
6 Livu-06
Lh-24tri
Lb-10iso
Tm-09iso
6 Trcr-11
Ia-16tri
Pw-13tri
5 Plwa-03
Lb-14tri
Lb-16tri
Ia-10iso
8 Trcr-06
Tc-18tri
Tc-16tri
Tm-10iso
Tm-07iso
6 Trcr-08
6 Livu-L1-01
Lv-27L1tri
Lb-20L1iso
Ia-18L1iso
Tm-14L1tri
7 Plwa-L1-01
Pw-15L1iso
MHC-I -like1
0.5
N conserved AA
▼ 1-4
▼ 5-7
▼ 8-9
Maximum relative expression
0.30
0.15
0.05
0.01

Fig. S1. Phylogeny of MHC-I and MHC-I-like sequences. The RAxML-NG maximum likelihood tree constructed from protein sequences under the JTT+G4 amino-acid substitution model. The tree was rooted with *Andrias* sequence (GenBank AGY55973, not shown). Support values for clades with the minimum bootstrap support of 70% (100 bootstrap replicates) are shown. The tree contains sequences from the G+T dataset (as described in Materials and Methods) obtained from PacBio Iso-Seq, de novo assembly of RNA-seq and protein sequences predicted for manually annotated MHC genes in genome assemblies of *L. helveticus*, *L. vulgaris*, *P. waltl* and *T. cristatus*. For sequences predicted in genomes the number of protein-coding exons is shown. Circles at tippoints show maximum relative expression – the maximum fraction of total MHC-I expression (FPKM) attributable to a sequence across all RNA-seq libraries from postmetamorphic individuals within a species. Triangles show N conserved AA – the number of residues important for anchoring the termini of antigenic peptides containing amino acids that are conserved in classical MHC-I of most taxa. Sequences are color-coded according to the species and species phylogeny is in Fig. 1.

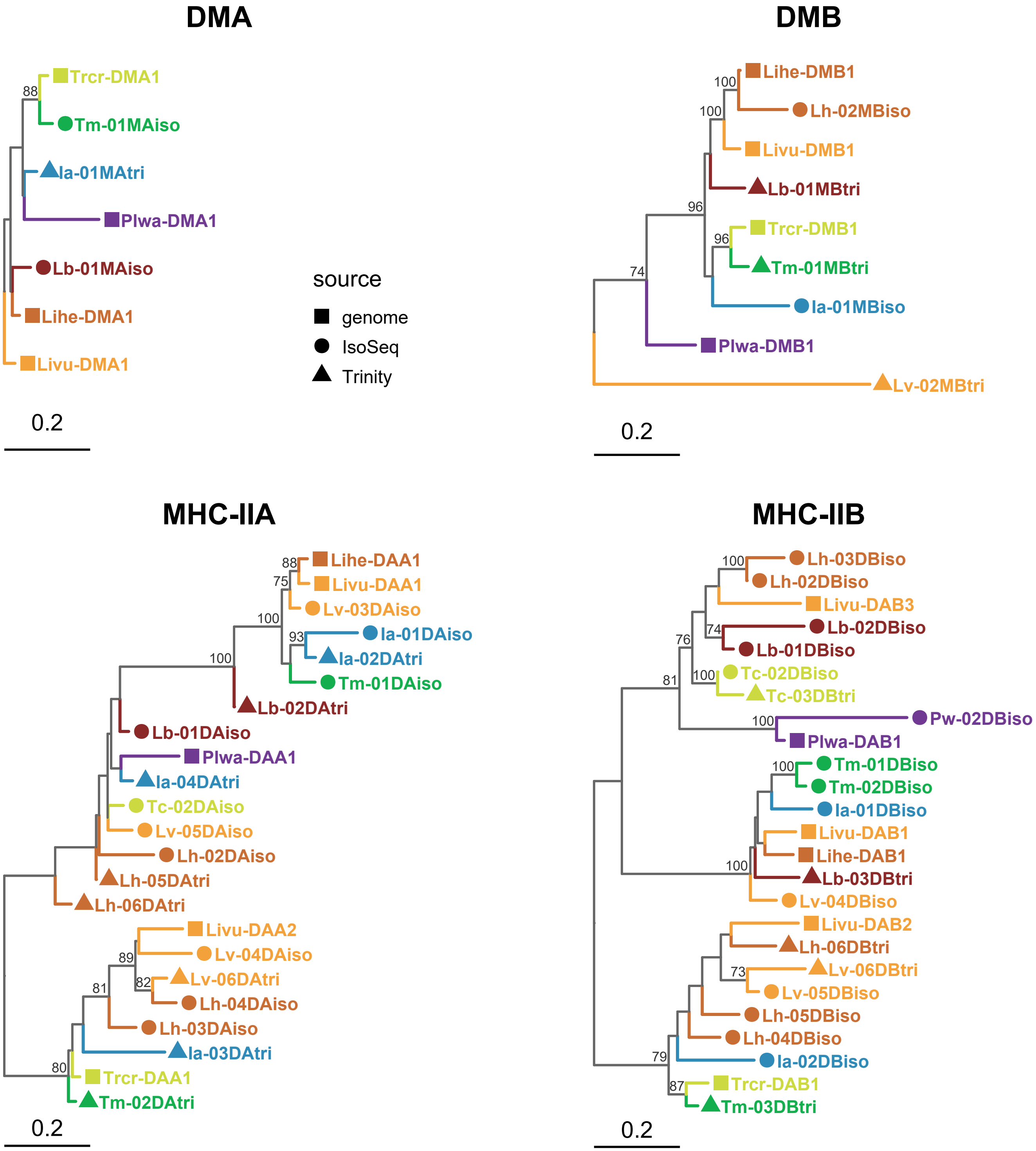
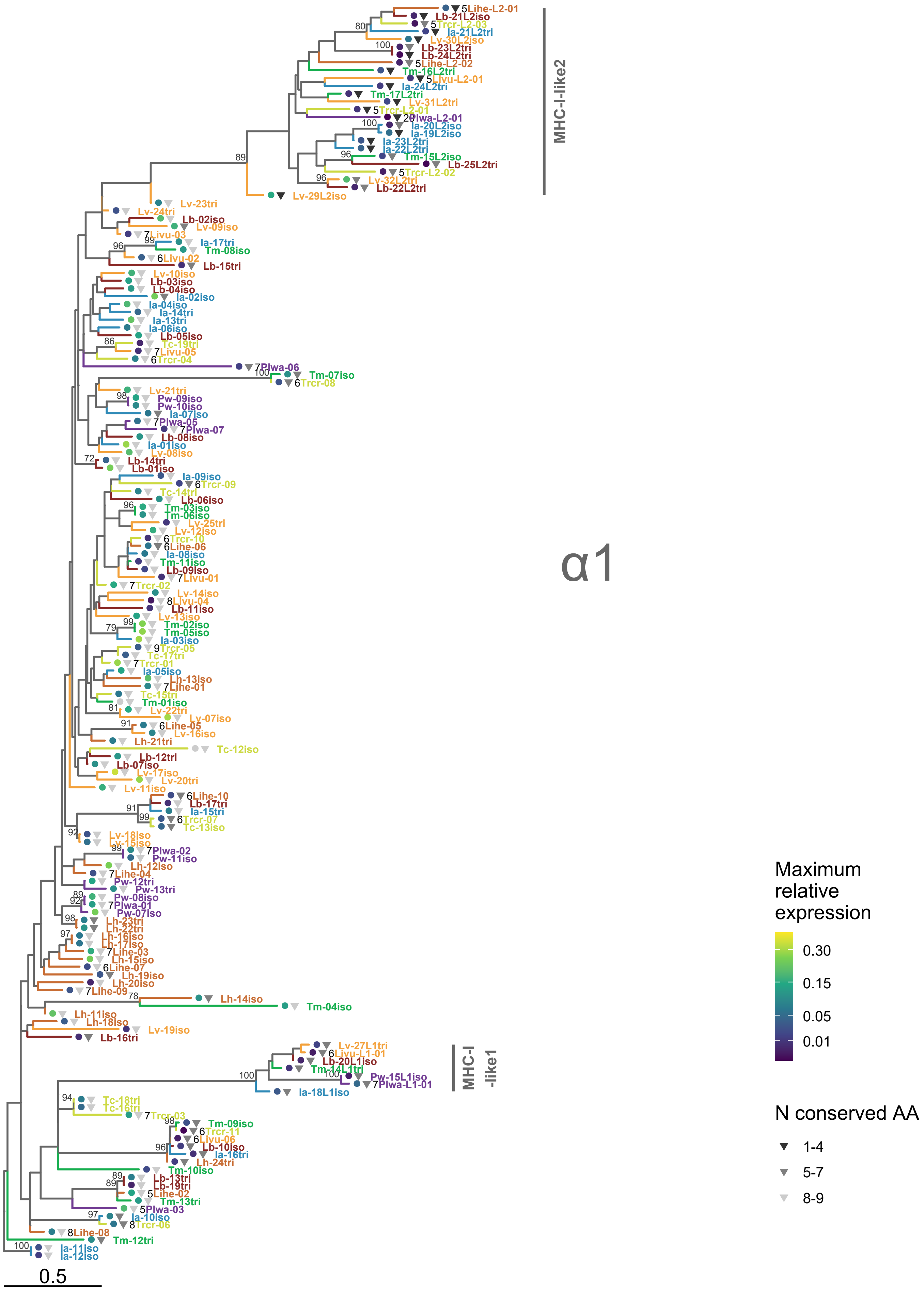
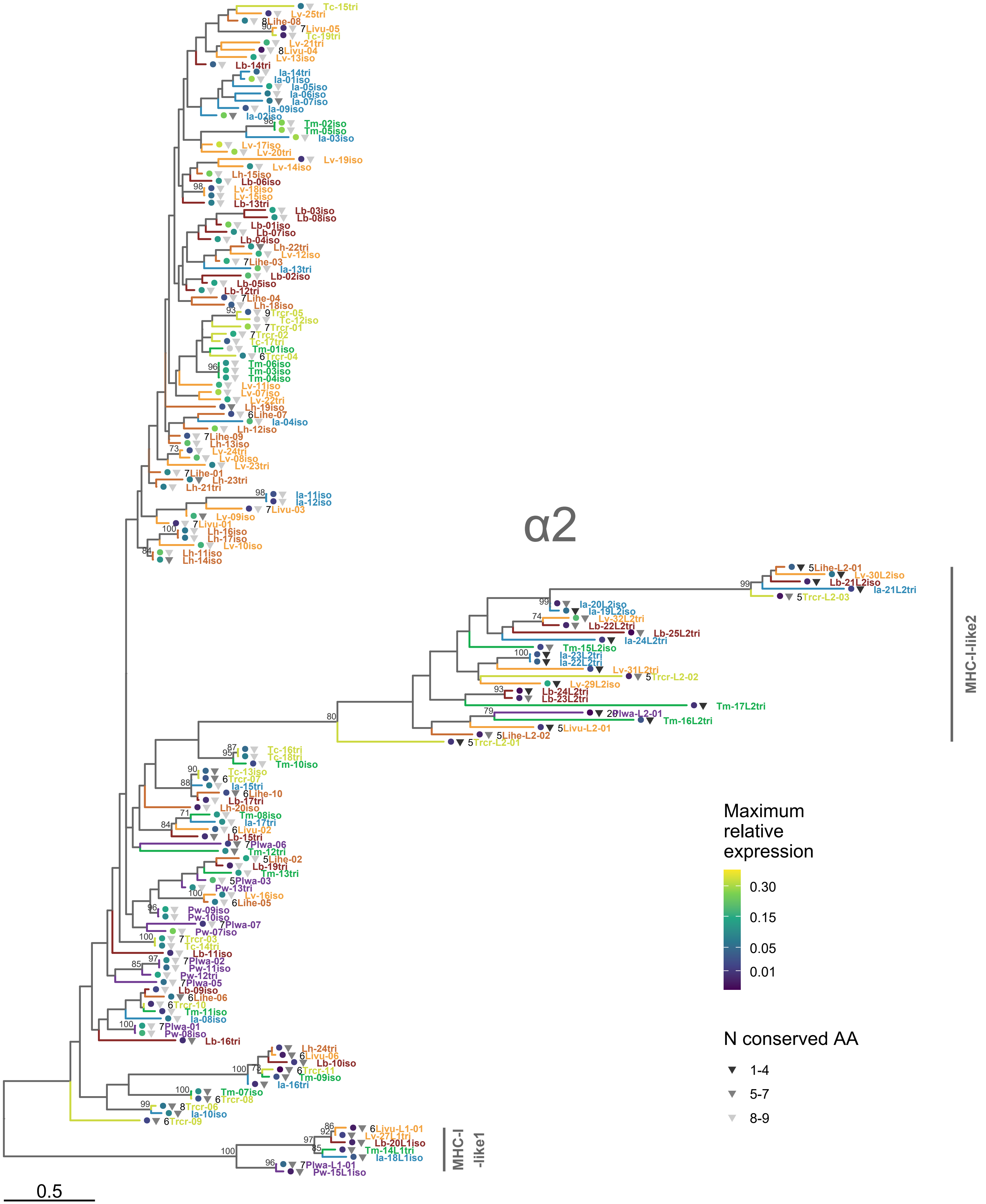

DMA
88
Trcr-DMA1
Tm-01MAiso
Ia-01MAtri
Plwa-DMA1
Lb-01MAiso
Lihe-DMA1
Livu-DMA1
0.2
source
■ genome
● IsoSeq
▲ Trinity
DMB
100
Lihe-DMB1
100
Lh-02MBiso
Livu-DMB1
Lb-01MBtri
96
96
Trcr-DMB1
Tm-01MBtri
74
Ia-01MBiso
Plwa-DMB1
Lv-02MBtri
0.2
MHC-IIA
88
Lihe-DAA1
75
Livu-DAA1
Lv-03DAiso
100
93
Ia-01DAiso
Ia-02DAtri
100
Tm-01DAiso
Lb-02DAtri
Lb-01DAiso
Plwa-DAA1
Ia-04DAtri
Tc-02DAiso
Lv-05DAiso
Lh-02DAiso
Lh-05DAtri
Lh-06DAtri
Livu-DAA2
89
Lv-04DAiso
81
82
Lv-06DAtri
Lh-04DAiso
Lh-03DAiso
Ia-03DAtri
80
Trcr-DAA1
Tm-02DAtri
0.2
MHC-IIB
100
Lh-03DBiso
Lh-02DBiso
Livu-DAB3
74
Lb-02DBiso
76
Lb-01DBiso
100
Tc-02DBiso
81
Tc-03DBtri
100
Pw-02DBiso
Plwa-DAB1
100
Tm-01DBiso
Tm-02DBiso
Ia-01DBiso
Livu-DAB1
Lihe-DAB1
100
Lb-03DBtri
Lv-04DBiso
Livu-DAB2
Lh-06DBtri
73
Lv-06DBtri
Lv-05DBiso
Lh-05DBiso
Lh-04DBiso
79
Ia-02DBiso
87
Trcr-DAB1
Tm-03DBtri
0.2

Fig. S2. Phylogenies of MHC-II sequences. Separate trees were constructed for α and β chains of DM and other class II proteins. The RAxML-NG maximum likelihood trees were constructed from protein sequences under the JTT+G4 amino-acid substitution model. The trees were rooted with *Andrias* or axolotl sequences (not shown). Support values for clades with the minimum bootstrap support of 70% (100 bootstrap replicates) are shown. The tree contains sequences from the G+T dataset (as described in Materials and Methods) obtained from PacBio Iso-Seq, de novo assembly of RNA-seq and protein sequences predicted for manually annotated MHC genes in genome assemblies of *L. helveticus*, *L. vulgaris*, *P. waltl* and *T. cristatus*. Sequences are color-coded according to the species and species phylogeny is in Fig. 1.



5Lihe-L2-01
Lb-21L2iso
80
5Trcr-L2-03
Ia-21L2tri
Lv-30L2iso
100
Lb-23L2tri
Lb-24L2tri
5Lihe-L2-02
Tm-16L2tri
5Livu-L2-01
Ia-24L2tri
Tm-17L2tri
Lv-31L2tri
5Trcr-L2-01
2Plwa-L2-01
100
Ia-20L2iso
Ia-19L2iso
Ia-23L2tri
Ia-22L2tri
96
Tm-15L2iso
Lb-25L2tri
5Trcr-L2-02
96
Lv-32L2tri
Lb-22L2tri
89
Lv-29L2iso
MHC-I-like2
Lv-23tri
Lv-24tri
Lb-02iso
Lv-09iso
7Livu-03
99
Ia-17tri
96
Tm-08iso
6Livu-02
Lb-15tri
Lv-10iso
Lb-03iso
Lb-04iso
Ia-02iso
Ia-04iso
Ia-14tri
Ia-13tri
Ia-06iso
Lb-05iso
86
Tc-19tri
7Livu-05
6Trcr-04
7Plwa-06
100
Tm-07iso
6Trcr-08
Lv-21tri
98
Pw-09iso
Pw-10iso
Ia-07iso
7Plwa-05
7Plwa-07
Lb-08iso
Ia-01iso
Lv-08iso
72
Lb-14tri
Lb-01iso
Ia-09iso
6Trcr-09
Tc-14tri
Lb-06iso
96
Tm-03iso
Tm-06iso
Lv-25tri
Lv-12iso
6Trcr-10
6Lihe-06
Ia-08iso
Tm-11iso
Lb-09iso
7Livu-01
7Trcr-02
Lv-14iso
8Livu-04
Lb-11iso
Lv-13iso
99
79
Tm-02iso
Tm-05iso
Ia-03iso
9Trcr-05
Tc-17tri
7Trcr-01
Ia-05iso
Lh-13iso
7Lihe-01
Tc-15tri
Tm-01iso
81
Lv-22tri
Lv-07iso
91
6Lihe-05
Lv-16iso
Lh-21tri
Tc-12iso
Lb-12tri
Lb-07iso
Lv-17iso
Lv-20tri
Lv-11iso
6Lihe-10
Lb-17tri
91
Ia-15tri
99
6Trcr-07
Tc-13iso
92
Lv-18iso
Lv-15iso
99
7Plwa-02
Pw-11iso
Lh-12iso
7Lihe-04
Pw-12tri
Pw-13tri
89
92
Pw-08iso
7Plwa-01
Pw-07iso
98
Lh-23tri
Lh-22tri
97
Lh-16iso
Lh-17iso
7Lihe-03
Lh-15iso
6Lihe-07
Lh-19iso
Lh-20iso
7Lihe-09
78
Lh-14iso
Tm-04iso
Lh-11iso
Lh-18iso
Lv-19iso
Lb-16tri
α1
Lv-27L1tri
6Livu-L1-01
Lb-20L1iso
Tm-14L1tri
100
100
Pw-15L1iso
7Plwa-L1-01
Ia-18L1iso
MHC-I
-like1
94
Tc-18tri
Tc-16tri
7Trcr-03
98
Tm-09iso
6Trcr-11
6Livu-06
96
Lb-10iso
Ia-16tri
Lh-24tri
Tm-10iso
89
89
Lb-13tri
Lb-19tri
5Lihe-02
Tm-13tri
97
5Plwa-03
Ia-10iso
8Trcr-06
8Lihe-08
Tm-12tri
100
Ia-11iso
Ia-12iso
0.5
Maximum relative expression
0.30
0.15
0.05
0.01
N conserved AA
▼ 1-4
▼ 5-7
▼ 8-9


Tc-15tri
Lv-25tri
8Lihe-08
90
7Livu-05
Tc-19tri
Lv-21tri
8Livu-04
Lv-13iso
Lb-14tri
Ia-14tri
Ia-01iso
Ia-05iso
Ia-06iso
Ia-07iso
Ia-09iso
Ia-02iso
98
Tm-02iso
Tm-05iso
Ia-03iso
Lv-17iso
Lv-20tri
Lv-19iso
Lv-14iso
Lh-15iso
Lb-06iso
98
Lv-18iso
Lv-15iso
Lb-13tri
Lb-03iso
Lb-08iso
Lb-01iso
Lb-07iso
Lb-04iso
Lh-22tri
Lv-12iso
7Lihe-03
Ia-13tri
Lb-02iso
Lb-05iso
Lb-12tri
7Lihe-04
Lh-18iso
93
9Trcr-05
Tc-12iso
7Trcr-01
7Trcr-02
Tc-17tri
Tm-01iso
6Trcr-04
96
Tm-06iso
Tm-03iso
Tm-04iso
Lv-11iso
Lv-07iso
Lv-22tri
Lh-19iso
6Lihe-07
Ia-04iso
Lh-12iso
7Lihe-09
Lh-13iso
73
Lv-24tri
Lv-08iso
Lv-23tri
7Lihe-01
Lh-23tri
Lh-21tri
98
Ia-11iso
Ia-12iso
7Livu-03
Lv-09iso
7Livu-01
100
Lh-16iso
Lh-17iso
Lv-10iso
84
Lh-11iso
Lh-14iso
α2
5Lihe-L2-01
Lv-30L2iso
99
Lb-21L2iso
Ia-21L2tri
5Trcr-L2-03
99
Ia-20L2iso
Ia-19L2iso
74
Lv-32L2tri
Lb-22L2tri
Lb-25L2tri
Ia-24L2tri
Tm-15L2iso
100
Ia-23L2tri
Ia-22L2tri
Lv-31L2tri
5Trcr-L2-02
Lv-29L2iso
93
Lb-24L2tri
Lb-23L2tri
Tm-17L2tri
79
2Plwa-L2-01
Tm-16L2tri
5Livu-L2-01
5Lihe-L2-02
5Trcr-L2-01
80
MHC-I-like2
87
95
Tc-16tri
Tc-18tri
Tm-10iso
90
Tc-13iso
88
6Trcr-07
Ia-15tri
6Lihe-10
Lb-17tri
Lh-20iso
71
Tm-08iso
Ia-17tri
84
6Livu-02
Lb-15tri
7Plwa-06
Tm-12tri
5Lihe-02
Lb-19tri
Tm-13tri
5Plwa-03
Pw-13tri
100
Lv-16iso
6Lihe-05
96
Pw-09iso
Pw-10iso
7Plwa-07
Pw-07iso
100
7Trcr-03
Tc-14tri
Lb-11iso
85
97
7Plwa-02
Pw-11iso
Pw-12tri
7Plwa-05
Lb-09iso
6Lihe-06
6Trcr-10
Tm-11iso
Ia-08iso
100
7Plwa-01
Pw-08iso
Lb-16tri
Lh-24tri
6Livu-06
Lb-10iso
100
78
6Trcr-11
Tm-09iso
Ia-16tri
100
Tm-07iso
6Trcr-08
99
8Trcr-06
Ia-10iso
6Trcr-09
86
92
6Livu-L1-01
Lv-27L1tri
97
Lb-20L1iso
85
Tm-14L1tri
Ia-18L1iso
100
96
7Plwa-L1-01
Pw-15L1iso
MHC-I
-like1
0.5
Maximum
relative
expression
0.30
0.15
0.05
0.01
N conserved AA
▼ 1-4
▼ 5-7
▼ 8-9

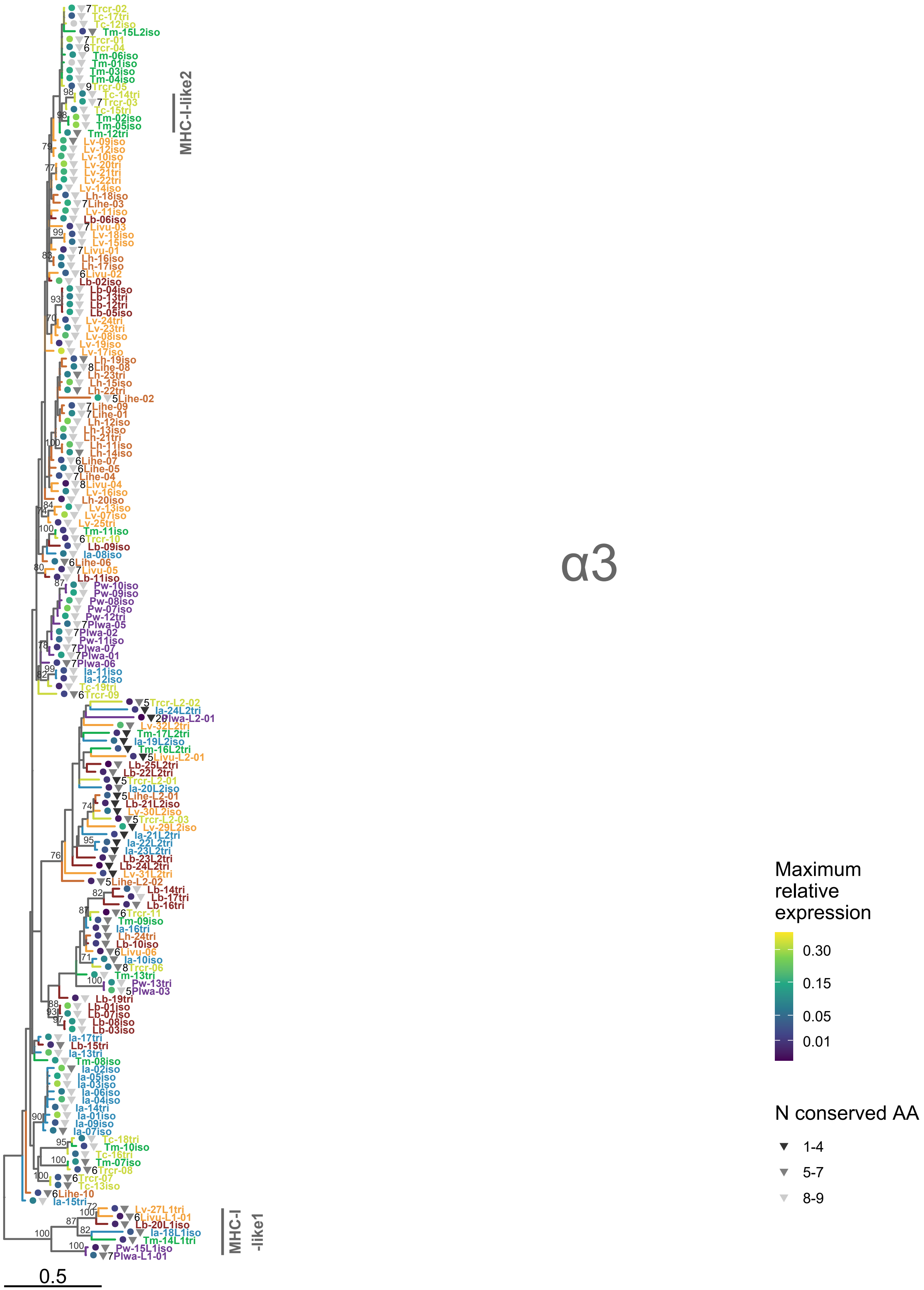

α3
MHC-I-like2
MHC-I -like1
Trcr-02
Tc-17tri
Tc-12iso
Tm-15L2iso
Trcr-01
Trcr-04
Tm-06iso
Tm-01iso
Tm-03iso
Tm-04iso
Trcr-05
Tc-14tri
Trcr-03
Tc-15tri
Tm-02iso
Tm-05iso
Tm-12tri
Lv-09iso
Lv-12iso
Lv-10iso
Lv-20tri
Lv-21tri
Lv-22tri
Lv-14iso
Lh-18iso
Lihe-03
Lv-11iso
Lb-06iso
Livu-03
Lv-18iso
Lv-15iso
Livu-01
Lh-16iso
Lh-17iso
Livu-02
Lb-02iso
Lb-04iso
Lb-13tri
Lb-12tri
Lb-05iso
Lv-24tri
Lv-23tri
Lv-08iso
Lv-19iso
Lv-17iso
Lh-19iso
Lihe-08
Lh-23tri
Lh-15iso
Lh-22tri
Lihe-02
Lihe-09
Lihe-01
Lh-12iso
Lh-13iso
Lh-21tri
Lh-11iso
Lh-14iso
Lihe-07
Lihe-05
Lihe-04
Livu-04
Lv-16iso
Lh-20iso
Lv-13iso
Lv-07iso
Lv-25tri
Tm-11iso
Trcr-10
Lb-09iso
Ia-08iso
Lihe-06
Livu-05
Lb-11iso
Pw-10iso
Pw-09iso
Pw-08iso
Pw-07iso
Pw-12tri
Plwa-05
Plwa-02
Pw-11iso
Plwa-07
Plwa-01
Plwa-06
Ia-11iso
Ia-12iso
Tc-19tri
Trcr-09
Trcr-L2-02
Ia-24L2tri
Plwa-L2-01
Lv-32L2tri
Tm-17L2tri
Ia-19L2iso
Tm-16L2tri
Livu-L2-01
Lb-25L2tri
Lb-22L2tri
Trcr-L2-01
Ia-20L2iso
Lihe-L2-01
Lb-21L2iso
Lv-30L2iso
Trcr-L2-03
Lv-29L2iso
Ia-21L2tri
Ia-22L2tri
Ia-23L2tri
Lb-23L2tri
Lb-24L2tri
Lv-31L2tri
Lihe-L2-02
Lb-14tri
Lb-17tri
Lb-16tri
Trcr-11
Tm-09iso
Ia-16tri
Lh-24tri
Lb-10iso
Livu-06
Ia-10iso
Trcr-06
Tm-13tri
Pw-13tri
Plwa-03
Lb-19tri
Lb-01iso
Lb-07iso
Lb-08iso
Lb-03iso
Ia-17tri
Lb-15tri
Ia-13tri
Tm-08iso
Ia-02iso
Ia-05iso
Ia-03iso
Ia-06iso
Ia-04iso
Ia-14tri
Ia-01iso
Ia-09iso
Ia-07iso
Tc-18tri
Tm-10iso
Tc-16tri
Tm-07iso
Trcr-08
Trcr-07
Tc-13iso
Lihe-10
Ia-15tri
Lv-27L1tri
Livu-L1-01
Lb-20L1iso
Ia-18L1iso
Tm-14L1tri
Pw-15L1iso
Plwa-L1-01
0.5
Maximum relative expression
0.30
0.15
0.05
0.01
N conserved AA
1-4
5-7
8-9

Fig. S3-S5. Phylogenies of MHC-I and MHC-I-like domains α1, α2, and α3. The RAxML-NG maximum likelihood tree constructed from protein sequences under the JTT+G4 amino-acid substitution model. The tree was rooted with *Andrias* sequence (GenBank AGY55973, not shown). Support values for clades with the minimum bootstrap support of 70% (100 bootstrap replicates) are shown. The tree contains sequences from the G+T dataset (as described in Materials and Methods) obtained from PacBio Iso-Seq, de novo assembly of RNA-seq and protein sequences predicted for manually annotated MHC genes in genome assemblies of *L. helveticus*, *L. vulgaris*, *P. waltl* and *T. cristatus*. For sequences predicted in genomes the number of protein-coding exons is shown. Circles at tippoints show maximum relative expression – the maximum fraction of total MHC-I expression (FPKM) attributable to a sequence across all RNA-seq libraries from postmetamorphic individuals within a species. Triangles show N conserved AA – the number of residues important for anchoring the termini of antigenic peptides containing amino acids that are conserved in classical MHC-I of most taxa. Sequences are color-coded according to the species and species phylogeny is in Fig. 1. Note, that in Fig. S5 (α3) tree, a single *T. marmoratus* MHC-I-like2 sequence did not cluster together with the remaining MHC-I-like2 sequences, so the labelled MHC-I-like2 clade does not include this sequence.

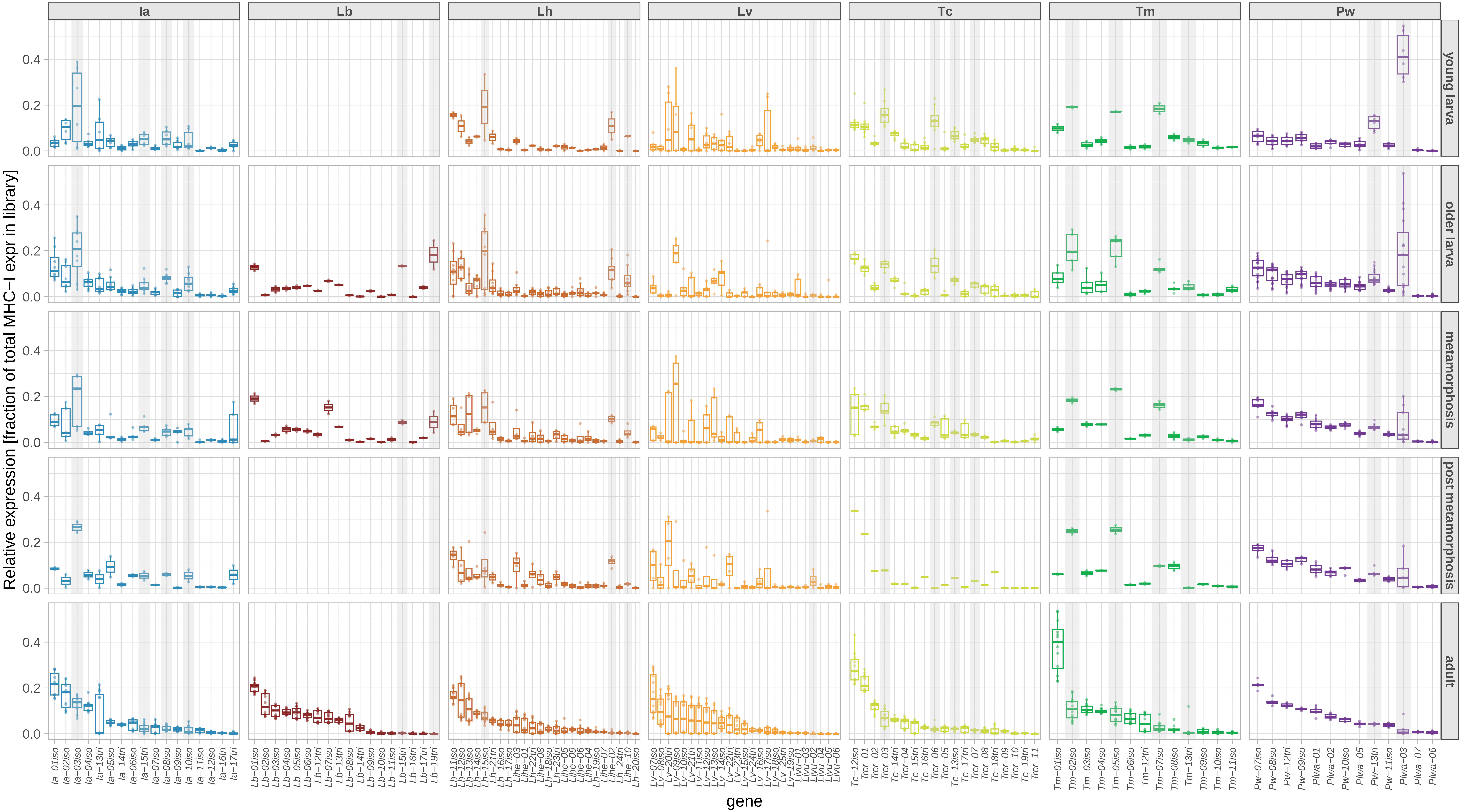

Ia
Lb
Lh
Lv
Tc
Tm
Pw
young larva
older larva
metamorphosis
post metamorphosis
adult
Relative expression [fraction of total MHC−I expr in library]
0.0
0.2
0.4
gene
Ia−01iso
Ia−02iso
Ia−03iso
Ia−04iso
Ia−13tri
Ia−05iso
Ia−14tri
Ia−06iso
Ia−15tri
Ia−07iso
Ia−08iso
Ia−09iso
Ia−10iso
Ia−11iso
Ia−12iso
Ia−16tri
Ia−17tri
Lb−01iso
Lb−02iso
Lb−03iso
Lb−04iso
Lb−05iso
Lb−06iso
Lb−12tri
Lb−07iso
Lb−13tri
Lb−08iso
Lb−14tri
Lb−09iso
Lb−10iso
Lb−11iso
Lb−15tri
Lb−16tri
Lb−17tri
Lb−19tri
Lh−11iso
Lh−12iso
Lh−13iso
Lh−14iso
Lh−15iso
Lh−21tri
Lh−16iso
Lh−17iso
Lihe−03
Lihe−01
Lh−22tri
Lihe−08
Lh−18iso
Lh−23tri
Lihe−05
Lihe−09
Lihe−06
Lihe−04
Lh−19iso
Lihe−07
Lihe−02
Lh−24tri
Lihe−10
Lh−20iso
Lv−07iso
Lv−08iso
Lv−20tri
Lv−09iso
Lv−10iso
Lv−21tri
Lv−11iso
Lv−12iso
Lv−13iso
Lv−14iso
Lv−22tri
Lv−23tri
Lv−15iso
Lv−24tri
Lv−16iso
Lv−17iso
Lv−18iso
Lv−25tri
Lv−19iso
Livu−01
Livu−03
Livu−02
Livu−04
Livu−05
Livu−06
Tc−12iso
Trcr−01
Trcr−02
Trcr−03
Tc−14tri
Trcr−04
Tc−15tri
Tc−16tri
Trcr−06
Trcr−05
Tc−13iso
Tc−17tri
Trcr−07
Trcr−08
Tc−18tri
Trcr−09
Trcr−10
Tc−19tri
Trcr−11
Tm−01iso
Tm−02iso
Tm−03iso
Tm−04iso
Tm−05iso
Tm−06iso
Tm−12tri
Tm−07iso
Tm−08iso
Tm−13tri
Tm−09iso
Tm−10iso
Tm−11iso
Pw−07iso
Pw−08iso
Pw−12tri
Pw−09iso
Plwa−01
Plwa−02
Pw−10iso
Plwa−05
Pw−13tri
Pw−11iso
Plwa−03
Plwa−07
Plwa−06

Fig. S6. Expression of the G+T dataset MHC-I sequences through ontogeny. Relative expression, i.e., FPKM for a particular cluster representative divided by the total MHC-I FPKM in a library. Boxplots show medians, interquartile and total ranges. Grey stripes indicate putative nonclassical (MHC-Ib) sequences identified as described in the text. Note, that for genes *Livu-02* and *Trcr-07* the pattern of higher relative expression in larval stages, visible when mapping RNA-seq data to genomes and used to classify them as MHC-Ib, is not clearly visible when mapping was performed to the broader set of G+T reference sequences. Species abbreviations: Ia – *I. alpestris*, Lb – *L. boscai*, Lh - *L. helveticus*, Lv - *L. vulgaris*, Pw – *P. waltl*, Tc – *T. cristatus*, Tm - *T. marmoratus*.

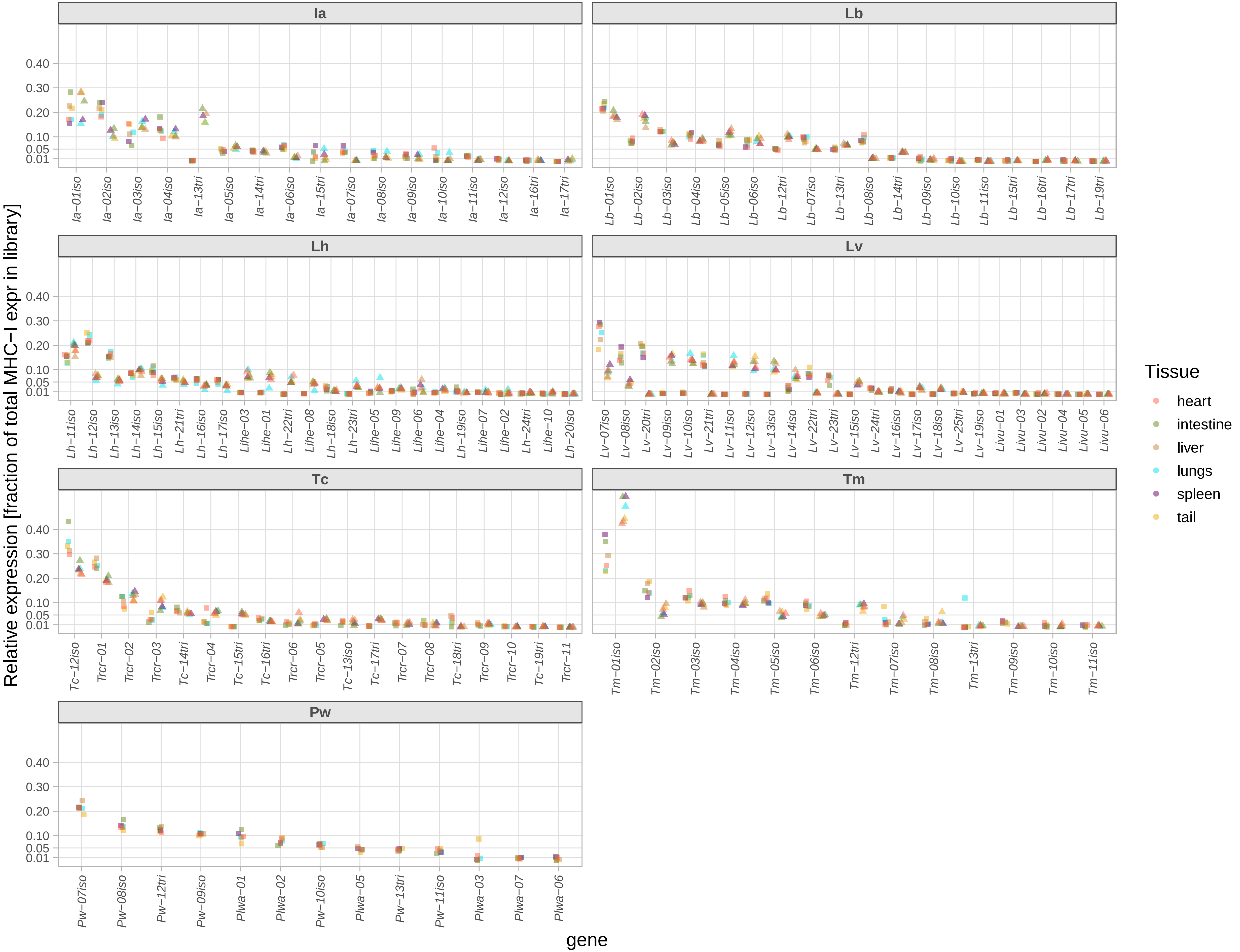

Ia
Lb
Lh
Lv
Tc
Tm
Pw
Relative expression [fraction of total MHC−I expr in library]
gene
0.40
0.30
0.20
0.10
0.05
0.01
Ia−01iso
Ia−02iso
Ia−03iso
Ia−04iso
Ia−13tri
Ia−05iso
Ia−14tri
Ia−06iso
Ia−15tri
Ia−07iso
Ia−08iso
Ia−09iso
Ia−10iso
Ia−11iso
Ia−12iso
Ia−16tri
Ia−17tri
Lb−01iso
Lb−02iso
Lb−03iso
Lb−04iso
Lb−05iso
Lb−06iso
Lb−12tri
Lb−07iso
Lb−13tri
Lb−08iso
Lb−14tri
Lb−09iso
Lb−10iso
Lb−11iso
Lb−15tri
Lb−16tri
Lb−17tri
Lb−19tri
Lh−11iso
Lh−12iso
Lh−13iso
Lh−14iso
Lh−15iso
Lh−21tri
Lh−16iso
Lh−17iso
Lihe−03
Lihe−01
Lh−22tri
Lihe−08
Lh−18iso
Lh−23tri
Lihe−05
Lihe−09
Lihe−06
Lihe−04
Lh−19iso
Lihe−07
Lihe−02
Lh−24tri
Lihe−10
Lh−20iso
Lv−07iso
Lv−08iso
Lv−20tri
Lv−09iso
Lv−10iso
Lv−21tri
Lv−11iso
Lv−12iso
Lv−13iso
Lv−14iso
Lv−22tri
Lv−23tri
Lv−15iso
Lv−24tri
Lv−16iso
Lv−17iso
Lv−18iso
Lv−25tri
Lv−19iso
Livu−01
Livu−03
Livu−02
Livu−04
Livu−05
Livu−06
Tc−12iso
Trcr−01
Trcr−02
Trcr−03
Tc−14tri
Trcr−04
Tc−15tri
Tc−16tri
Trcr−06
Trcr−05
Tc−13iso
Tc−17tri
Trcr−07
Trcr−08
Tc−18tri
Trcr−09
Trcr−10
Tc−19tri
Trcr−11
Tm−01iso
Tm−02iso
Tm−03iso
Tm−04iso
Tm−05iso
Tm−06iso
Tm−12tri
Tm−07iso
Tm−08iso
Tm−13tri
Tm−09iso
Tm−10iso
Tm−11iso
Pw−07iso
Pw−08iso
Pw−12tri
Pw−09iso
Plwa−01
Plwa−02
Pw−10iso
Plwa−05
Pw−13tri
Pw−11iso
Plwa−03
Plwa−07
Plwa−06
Tissue
heart
intestine
liver
lungs
spleen
tail

Fig. S7. Expression of MHC-I sequences in adult tissues. Relative expression, i.e., FPKM for a particular cluster representative divided by the total MHC-I FPKM in a library. RNA-seq libraries from different tissues are indicated with different colors and individuals with different shapes. Species abbreviations: Ia – *I. alpestris*, Lb – *L. boscai*, Lh - *L. helveticus*, Lv - *L. vulgaris*, Pw – *P. waltl*, Tc – *T. cristatus*, Tm - *T. marmoratus*. Sequence names correspond to those in Fig. 4, S1-S5. Note, that the cluster represented by the gene *Plwa-03* included also another gene annotated in the genome – *Plwa-04*.

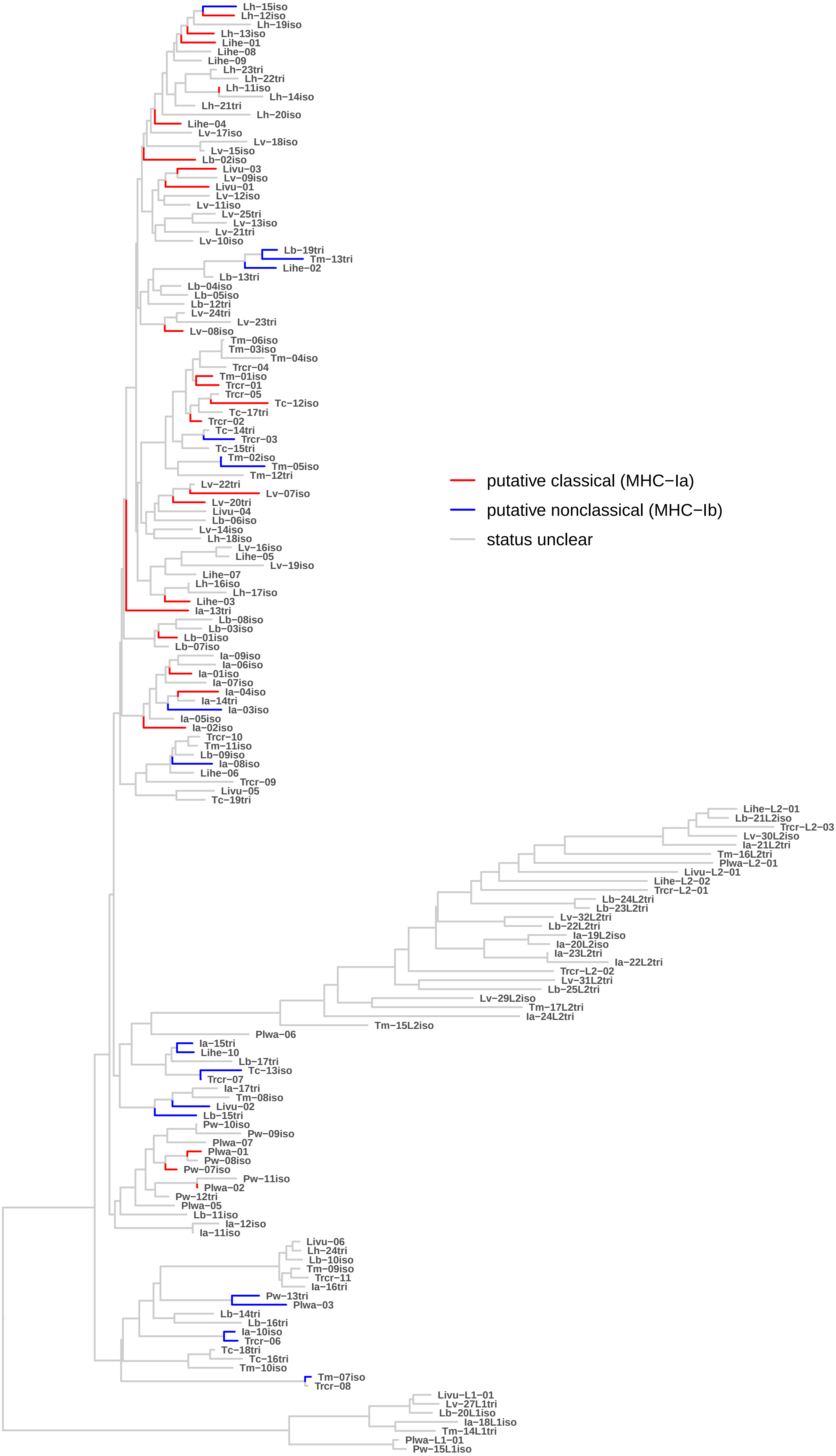

Lh−15iso
Lh−12iso
Lh−19iso
Lh−13iso
Lihe−01
Lihe−08
Lihe−09
Lh−23tri
Lh−22tri
Lh−11iso
Lh−14iso
Lh−21tri
Lh−20iso
Lihe−04
Lv−17iso
Lv−18iso
Lv−15iso
Lb−02iso
Livu−03
Lv−09iso
Livu−01
Lv−12iso
Lv−11iso
Lv−25tri
Lv−13iso
Lv−21tri
Lv−10iso
Lb−19tri
Tm−13tri
Lihe−02
Lb−13tri
Lb−04iso
Lb−05iso
Lb−12tri
Lv−24tri
Lv−23tri
Lv−08iso
Tm−06iso
Tm−03iso
Tm−04iso
Trcr−04
Tm−01iso
Trcr−01
Trcr−05
Tc−12iso
Tc−17tri
Trcr−02
Tc−14tri
Trcr−03
Tc−15tri
Tm−02iso
Tm−05iso
Tm−12tri
putative classical (MHC−Ia)
putative nonclassical (MHC−Ib)
status unclear
Lv−22tri
Lv−07iso
Lv−20tri
Livu−04
Lb−06iso
Lv−14iso
Lh−18iso
Lv−16iso
Lihe−05
Lv−19iso
Lihe−07
Lh−16iso
Lh−17iso
Lihe−03
Ia−13tri
Lb−08iso
Lb−03iso
Lb−01iso
Lb−07iso
Ia−09iso
Ia−06iso
Ia−01iso
Ia−07iso
Ia−04iso
Ia−14tri
Ia−03iso
Ia−05iso
Ia−02iso
Trcr−10
Tm−11iso
Lb−09iso
Ia−08iso
Lihe−06
Trcr−09
Livu−05
Tc−19tri
Lihe−L2−01
Lb−21L2iso
Trcr−L2−03
Lv−30L2iso
Ia−21L2tri
Tm−16L2tri
Plwa−L2−01
Livu−L2−01
Lihe−L2−02
Trcr−L2−01
Lb−24L2tri
Lb−23L2tri
Lv−32L2tri
Lb−22L2tri
Ia−19L2iso
Ia−20L2iso
Ia−23L2tri
Ia−22L2tri
Trcr−L2−02
Lv−31L2tri
Lb−25L2tri
Lv−29L2iso
Tm−17L2tri
Ia−24L2tri
Tm−15L2iso
Plwa−06
Ia−15tri
Lihe−10
Lb−17tri
Tc−13iso
Trcr−07
Ia−17tri
Tm−08iso
Livu−02
Lb−15tri
Pw−10iso
Pw−09iso
Plwa−07
Plwa−01
Pw−08iso
Pw−07iso
Pw−11iso
Plwa−02
Pw−12tri
Plwa−05
Lb−11iso
Ia−12iso
Ia−11iso
Livu−06
Lh−24tri
Lb−10iso
Tm−09iso
Trcr−11
Ia−16tri
Pw−13tri
Plwa−03
Lb−14tri
Lb−16tri
Ia−10iso
Trcr−06
Tc−18tri
Tc−16tri
Tm−10iso
Tm−07iso
Trcr−08
Livu−L1−01
Lv−27L1tri
Lb−20L1iso
Ia−18L1iso
Tm−14L1tri
Plwa−L1−01
Pw−15L1iso

Fig. S8. Position of putative classical and nonclassical MHC-I sequences on phylogeny from Fig. S1. Putative classical (MHC-Ia) and nonclassical (MHC-Ib) sequences were identified as described in the text and their branches are marked red and blue, respectively. The remaining sequences which status is unclear, as well as MHC-I-like sequences, are marked in gray.

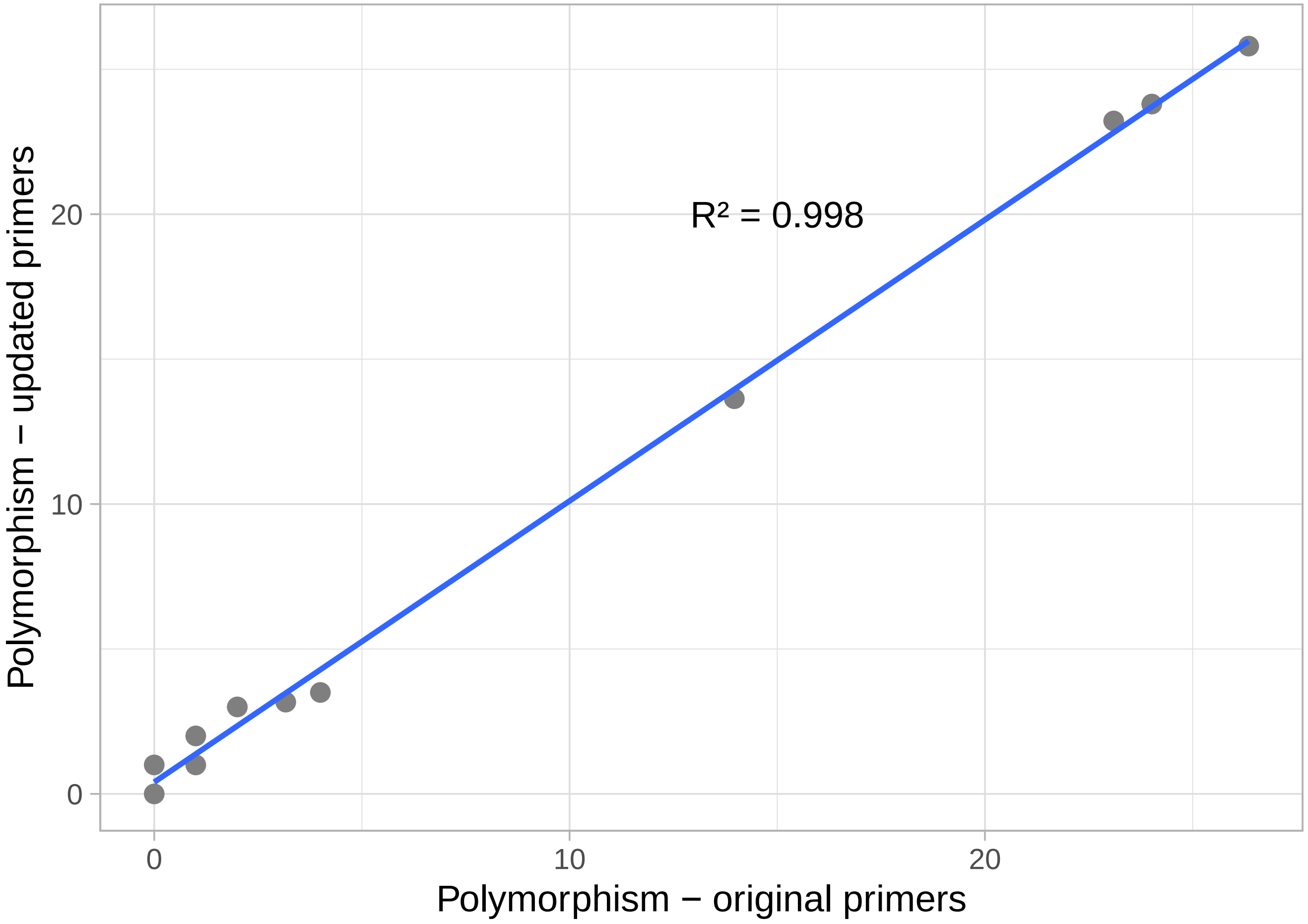

R² = 0.998
Polymorphism − updated primers
0
10
20
Polymorphism − original primers
0
10
20

Fig. S9. Polymorphism of MHC-I genes in *T. cristatus* estimated with previously used and updated primers. Twenty four individuals that were analysed previously by Palomar et al. (2021, original primers) were genotyped again using primers updated so that they would match exon 2 sequences of all genes annotated in the genome (updated primers). Polymorphism of each gene in both datasets was estimated as described in Materials and Methods and in Supplementary Materials.

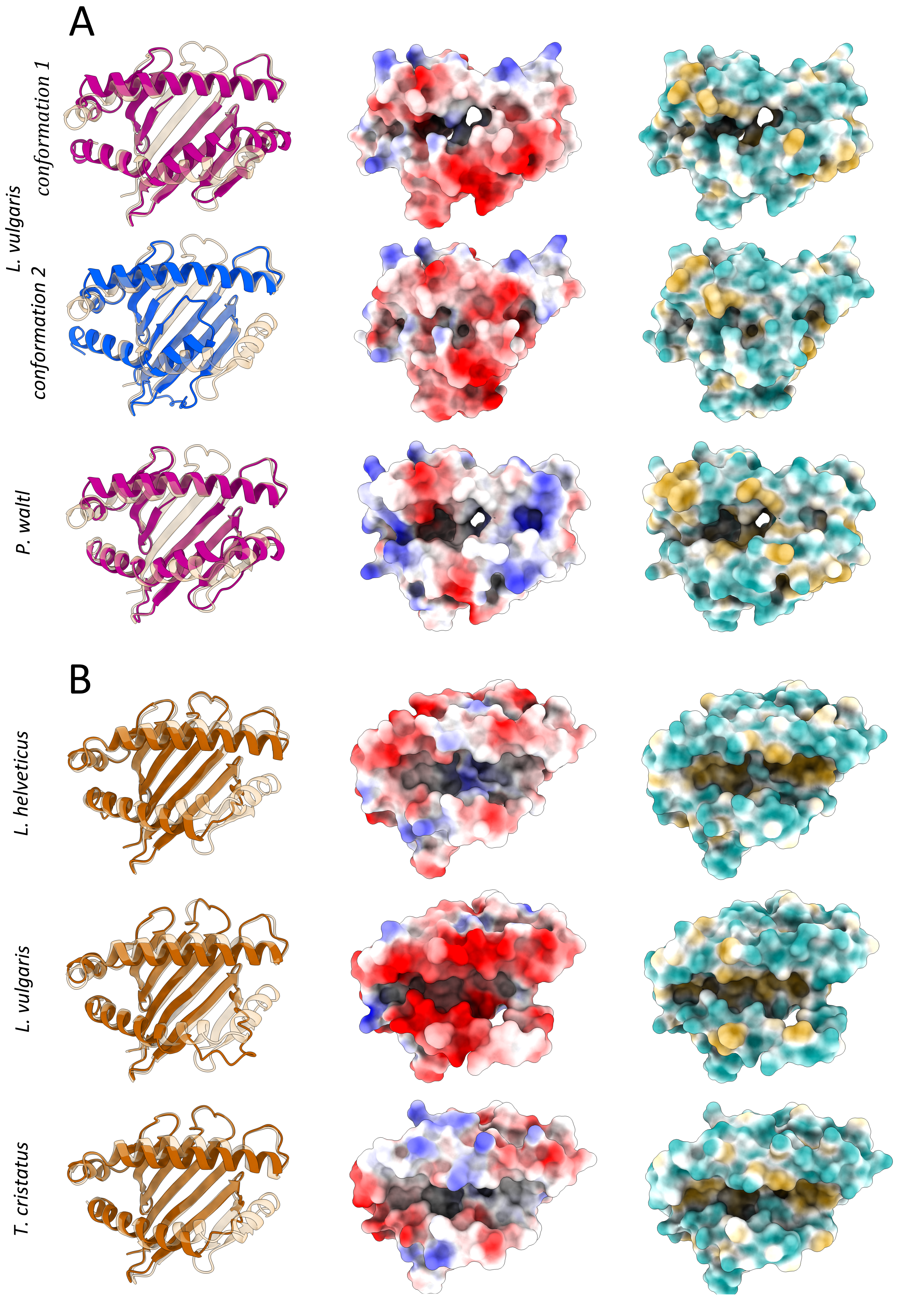

A
L. vulgaris
conformation 1
conformation 2
P. waltl
B
L. helveticus
L. vulgaris
T. cristatus

Fig. S10. Top views of α1-α2 domains of a) MHC-I-like1 and b) MHC-I-like2 molecules, generated with Alphafold3. Left hand-side structures: ribbon models of MHC-I-like molecules (magenta – top conformation of MHC-I-like1, blue – alternative MHC-I-like1 conformation, brown - top conformation of MHC-I-like2), overlaid on classical molecules of respective species. Middle structure: Coulombic electrostatic potential mapped on the molecular surface, with red for negative and blue for positive potential. Right hand-side structures: lipophilicity potential mapped on the molecular surface, with cyan most hydrophilic and goldenrod most lipophilic.

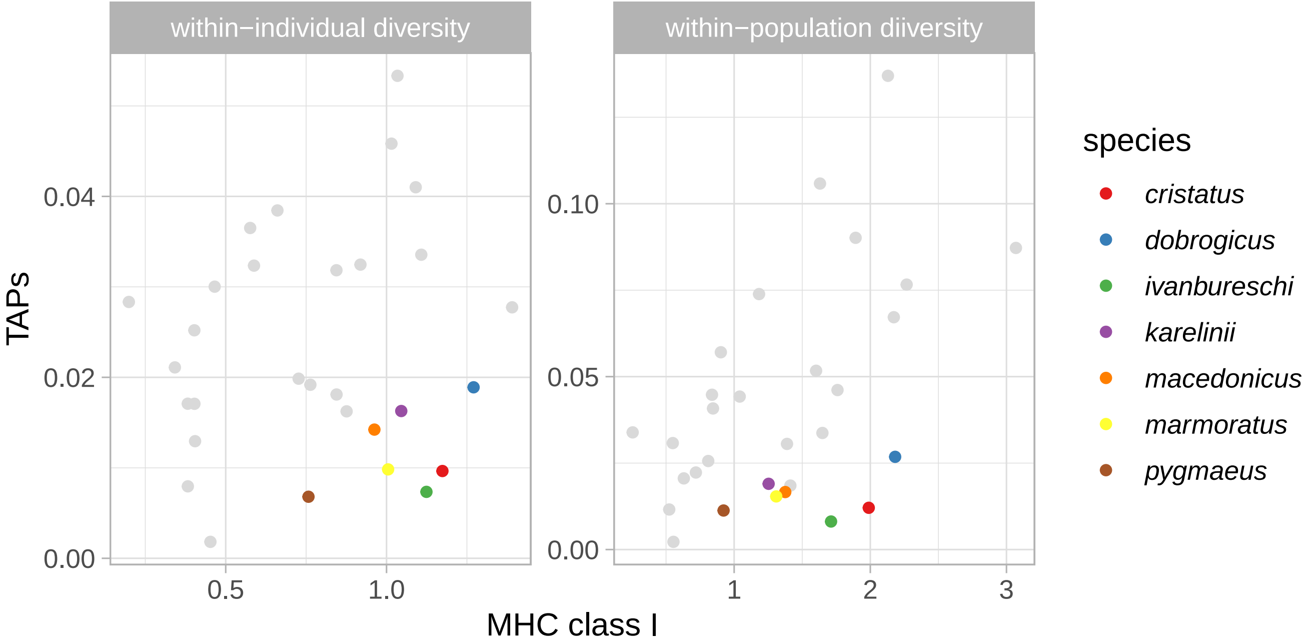

within−individual diversity
within−population diiversity
TAPs
0.00
0.02
0.04
0.00
0.05
0.10
0.5
1.0
1
2
3
MHC class I
species
cristatus
dobrogicus
ivanbureschi
karelinii
macedonicus
marmoratus
pygmaeus

Fig. S11. The relationship between MHC-I and *TAP* (*TAP1* + *TAP2*) diversity in *Triturus* newts. Estimates of phylogenetic alpha (within-individual) and gamma (within-population) diversity for 30 salamander species are from Palomar et al. (2021). There is a tendency for *Triturus* species (color-coded) to show low *TAP* diversity and little evidence for a correlation between MHC-I and *TAP* diversity, as opposed to the remaining 23 species representing 15 genera and six families.