Accepted Manuscript

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Jessica Hacking, Terry Bertozzi, Adnan Mousalli, Tessa Bradford, Michael Gardner

PII: S0145-305X(17)30672-9

DOI: 10.1016/j.dci.2018.02.012

Reference: DCI 3107

To appear in: *Developmental and Comparative Immunology*

Received Date: 15 December 2017

Revised Date: 10 February 2018

Accepted Date: 10 February 2018

Please cite this article as: Hacking, J., Bertozzi, T., Mousalli, A., Bradford, T., Gardner, M., Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard, *Developmental and Comparative Immunology* (2018), doi: 10.1016/j.dci.2018.02.012.

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Jessica Hacking[a] (corresponding author)

jessica.hacking@flinders.edu.au

Terry Bertozzi[b,c]

terry.bertozzi@samuseum.sa.gov.au

Adnan Mousalli[d]

amoussalli@museum.vic.gov.au

Tessa Bradford[a,b,c]

tessa.bradford@samuseum.sa.gov.au

Michael Gardner[a,b]

michael.gardner@flinders.edu.au

[a]College of Science and Engineering, Flinders University, Bedford Park SA 5042, Australia

[b]Evolutionary Biology Unit, South Australian Museum, Adelaide SA 5000, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

[d]Sciences Department, Museum Victoria, Carlton Gardens VIC 3053, Australia

Abstract

Characterisation of squamate major histocompatibility complex (MHC) genes has lagged behind other taxonomic groups. MHC genes encode cell-surface glycoproteins that present self- and pathogen-derived peptides to T cells and play a critical role in pathogen recognition. Here we characterize MHC class I transcripts for an agamid lizard (*Ctenophorus decresii*) and investigate the evolution of MHC class I in Iguanian lizards. An iterative assembly strategy was used to identify six full-length *C. decresii* MHC class I transcripts, which were validated as likely to encode classical class I MHC molecules. Evidence for exon shuffling recombination was uncovered for *C. decresii* transcripts and Bayesian phylogenetic analysis of Iguanian MHC class I sequences revealed a pattern expected under a birth-and-death mode of evolution. This work provides a stepping stone towards further research on the agamid MHC class I region.

Keywords: transcriptome assembly; Iguania, Agamidae; Ctenophorus decresii; MHC class I evolution

1. Introduction

The major histocompatibility complex (MHC) is a multigene family involved in pathogen recognition and immune response, and is one of the most diverse regions of the vertebrate genome (Piertney and Oliver, 2006). MHC genes encode cell surface glycoproteins that present self- and foreign-derived peptides to circulating T-lymphocyte cells (T cells). The evolution of MHC genes is complex and is thought to be governed primarily by the birth-and-death model of evolution in which loci are duplicated or lost, although concerted evolution via inter-locus gene conversion events may also play a role (Edwards and Hedrick, 1998; Nei and Rooney, 2005; Spurgin et al., 2011). These processes can occur over short time scales and it is apparent that MHC genes have undergone numerous independent expansion and diversification events throughout vertebrate evolution (Nei et al., 1997). The MHC is gene rich and is generally extremely polymorphic within loci (Janeway et al., 2001). Pathogen-mediated natural selection and sexual selection (MHC-associated mating) are considered to be the primary mechanisms maintaining these extraordinary levels of diversity (Edwards and Hedrick, 1998; Ejsmond et al., 2014; Milinski, 2006).

The MHC is divided into four classes based primarily on structural and functional differences (Janeway et al., 2001). Genes belonging to classes I and II are further separated into classical or non-classical genes based primarily on function and expression patterns (Alfonso and Karlsson, 2000; Janeway et al., 2001). The structure of classical MHC class I (hereafter MHC I) molecules is conserved among jawed vertebrates and includes a leader peptide, three α domains, and the transmembrane and cytoplasmic (Tm/Cyt) domains, all of which are encoded by a single gene (Kaufman et al., 1994). The α1 and α2 domains form the peptide binding cleft and contain amino acid positions that are directly involved in peptide binding, termed peptide binding regions (PBR) (Janeway et al., 2001). Classical MHC I molecules are anchored to the surface of somatic cells via the Tm/Cyt domains and display self-peptides and antigenic peptides derived primarily from intracellular pathogens to cytotoxic T cells. When a particular MHC molecule presents an antigenic peptide and is recognized by a T cell, an immune response is initiated, which usually involves lysis of the infected cell (Neefjes et al., 2011). Non-classical MHC class I genes are distinguished from classical genes by

low levels of allelic variation and restricted expression (Janeway et al., 2001). In mammals some non-classical MHC class I genes undertake important roles within the immune system, both at the cell surface and in secreted forms (Adams and Luoma, 2013).

The MHC has been thoroughly characterised in humans and model organisms, primarily due to the critical role that the region plays in organ and tissue transplantation (Garcia et al., 2012). MHC genes are also used in conservation genetic studies as a measure of population genetic health and adaptive potential (Sommer, 2005). However, many vertebrate groups, especially non-avian reptiles, are under-represented within the MHC literature and little is known regarding the mechanisms shaping MHC diversity in these taxa. The tawny dragon (*Ctenophorus decresii*) is a small (<30g) agamid lizard endemic to South Australia and provides a promising model system in which to investigate the mechanisms shaping MHC diversity. Male *C. decresii* use visual cues during social and sexual interactions (Gibbons, 1979; Osborne, 2005a, b; Yewers et al., 2016) and the species is host to external and intracellular parasites (Hacking et al. unpublished results); providing opportunities to investigate the roles of sexual selection and parasite-mediated selection in maintaining MHC diversity. Here, we characterised MHC I transcripts for *C. decresii* and investigated the evolutionary mechanisms playing a role in the generation of MHC I diversity within Iguanian lizards (Iguanidae, Agamidae, Chamaeleonidae, Dactyloidae and related families, Pyron et al., 2013).

2. Materials and methods

2.1. Sample collection

A single *C. decresii* individual was captured in Burra, South Australia (33°40’57.7”S, 138°56’16.8”E) in October 2012 and taken directly to Adelaide to be euthanized for tissue collection. Burra is located just north of the contact zone between the northern and southern clades of this species (McLean et al., 2014). The thymus and spleen were collected immediately after euthanasia and were stored separately in RNA Later (Qiagen, Venlo, Netherlands) at 4°C for 48 h and then at -80°C until required for RNA

extraction. The remainder of the specimen was accessioned into the South Australian Museum herpetology collection (SAMAR67384).

2.2. Transcriptome sequencing and MHC class I discovery

Total RNA was extracted using the Qiagen RNeasy mini kit. Sequencing libraries were then prepared using the TruSeq RNA Kit v1 using a polyA purification. These were then multiplexed with two other samples and sequenced (100 bp paired end) on a single lane of the HISEQ 2000. Extractions, library preparation and sequencing were carried out by Georgia Genomic Facility (GGF, University of Georgia, USA). Adaptor sequences and low quality reads were removed or trimmed using Trimmomatic ver. 0.22 (Bolger et al., 2014), with a minimum quality Phred score of 25 per 4bp sliding window and a minimum sequence length of 40bp. Assemblies were then constructed from the trimmed and filtered reads for each sample separately using the program Trinity v1 (r2013-02-25) (Haas et al., 2013) with default settings, followed by an assessment of gene completeness using BUSCO v1.22 (Simao et al., 2015) based on the OrthoDB 'vertebrata' database. Lastly, to identify putative *C. decresii* MHC I transcripts we performed local BLASTX (E-value ≤ 1e-10) searches (Altschul et al., 1997; Camacho et al., 2009) using predicted *Pogona vitticeps* MHC gene models (Georges et al., 2015) as our reference. Putative *C. decresii* MHC I transcripts were aligned manually with published MHC I sequences (table S1) to confirm expected MHC I structure and the presence of conserved sites (Kaufman et al., 1994).

Due to the high diversity and complex structure of the MHC region, traditional assembly methods may not be sufficient to obtain a robust assembly. To refine the MHC I assemblies, 75bp sequence fragments congruent with the putative antigen binding α2 domain for each unique sequence were iteratively re-assembled using the mirabait utility from MIRA v4.0.2 (Chevreux et al., 1999) as presented in Ansari et al. (2015) but using a kmer length of 31 (size of the search string) and requiring 50 matching kmers (number of matching search strings). Sequences were extended until sequence length stabilized or it was no longer possible to uniquely map reads. The resulting contigs were evaluated by concatenating each separated by 200 Ns and remapping cleaned reads using BWA (Li,

2013) with default settings and visualizing the resultant BAM file in IGV (Robinson et al., 2011). Read pairs spanning sequences were checked for accuracy and suspected chimeric reads removed.

2.3. Validation of MHC I transcripts and comparison with other vertebrates

Putative *C. decresii* MHC class I transcripts were translated and aligned with a subset of published full length MHC I amino acid sequences of other vertebrates (table S1) using MUSCLE (Edgar, 2004) implemented in MEGA ver. 6.06 (Tamura et al., 2011). The alignment was manually refined to ensure correct alignment of conserved regions. Coding domain boundaries were defined as per Koller and Orr (1985). Aligned *C. decresii* transcripts were validated as likely MHC I sequences by (i) confirming MHC I gene structure (leader peptide, α domains and Tm/Cyt domains), (ii) confirming concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences (Kaufman et al., 1994), and iii) confirming the absence of stop codons within coding regions. Two additional steps needed to further validate transcripts as likely classical MHC I sequences, which were beyond the scope of this study, are confirming polymorphism among individuals, and strong and widespread expression. Pairwise nucleotide and amino acid identity among validated MHC class I transcript sequences was calculated using Geneious ver. 8.1.7 (Kearse et al., 2012). Validated *C. decresii* MHC class I transcripts were named according to Klein et al. (1990); each unique nucleotide sequence was given the species identification prefix (Ctde) followed by U (Uno; class 1) and A (locus group/family designation), and a unique number (e.g. *Ctde-UA*001*). Once full-length MHC I genomic data are available for *C. decresii* locus designations may be defined (i.e. UA1 and UA2).

To investigate relationships among *C. decresii* MHC I transcripts and their position relative to other Iguanian lizards a Bayesian phylogenetic tree was constructed based on an alignment of validated full-length *C. decresii* MHC class I transcripts and all full-length Iguanian MHC I sequences available in GenBank (NCBI Resource Coordinators, 2016) (table S1). All sequences obtained from GenBank were validated as likely MHC I sequences via confirmation of expected MHC I structure and conserved sites. All squamate MHC I nucleotide sequences were translated before aligning with Muscle, implemented in MEGA ver. 6.06 and then untranslated for phylogenetic analysis. Tuatara

(*Sphenodon punctatus*) MHC I sequence was used as an outgroup. Only the three α domains (α1, α2 and α3) of these full-length sequences were used in phylogenetic analysis due to extreme variation at leader and Cyt/Tm domains inhibiting alignment. To determine optimal partitioning and the best model of evolution for Bayesian phylogenetic analysis, PartitionFinder2 (Guindon and Gascuel, 2003; Lanfear et al., 2012; Lanfear et al., 2017) was employed. Sequences were split into three data blocks representing codon positions and only models employed by MrBayes were considered, with the best model determined using AICc model selection. Bayesian phylogenetic analyses were undertaken in MrBayes ver. 3.2.6 (Ronquist et al., 2011), with one analysis employing the model of evolution and partitioning identified by PartitionFinder2, and another using mixed models with sequences partitioned by codon position. For both MrBayes analyses, two independent runs were performed, each with four Markov chains run for 20 million generations at a sample frequency of 1000 and a default burn-in period of 25%. Convergence diagnostics, including the standard deviation of split frequencies between runs, the potential scale reduction factor (PSRF) and the average effective sample size (ESS) were examined to confirm run convergence. FigTree ver. 1.4.2 (Rambaut, 2012) was used to annotate trees produced by MrBayes.

3. Results and Discussion

3.1. Transcriptome sequencing and MHC class I discovery

RNA extractions had RIN values of 9.1 and 9.8 for the thymus and spleen samples, respectively. Approximately 49 million paired-end reads were obtained for the thymus and approximately 51 million paired-end reads were obtained for the spleen. Of these reads, approximately 86% survived Trimmomatic filtering. Based on the Busco analysis, 73% and 76% of the vertebrata orthologous gene set were recovered and complete for the thymus (C:73%[D:15%],F:7.7%,M:19%,n:3023) and spleen (C:76%[D:13%],F:6.4%,M:16%,n:3023) transcriptomes, respectively. Combing the two transcriptome assemblies resulted in 82% of the vertebrata gene set being complete (C:82%[D:43%],F:5.1%,M:12%,n:3023). In total, eight putative *C. decresii* MHC I transcripts (*Ctde-UA*001 – Ctde-UA*008*) were discovered, based on unique α2 and α3 domains, indicating the

presence of at least four different loci. We were unable to confidently assign transcripts to specific loci based on data from a single individual. As a result, when naming transcripts all sequences were designated the letters 'UA' (MHC I locus group 'A') and no specific locus numbers were assigned (i.e. UA1, UA2, UA3 etc).

3.2. Validation of C. decresii MHC I transcripts

Of the eight putative *C. decresii* MHC I transcripts, six were validated as likely MHC I sequences after confirming (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences, and (iii) the absence of stop codons within coding regions (*Ctde-UA*001 – Ctde-UA*006*). Each of these transcripts contained complete leader peptide, α domains and Tm/Cyt domains and did not contain any premature stop codons. An amino acid alignment of putative *C. decresii* MHC I transcripts with published squamate, tuatara (*Sphenodon punctatus*) and human sequences, confirmed concordance with conserved regions and regions with predicted functions that are typical of MHC I molecules (table S2, fig. 1, Kaufman et al., 1994). Specifically, nine amino acid positions that bind C- and N-terminal residues of antigenic peptides are highly conserved among classical MHC class I molecules (Kaufman et al., 1994). All nine of these positions were identified within *C. decresii* MHC I transcripts and displayed conserved amino acids or low (≤2 changes) amino acid variability (site numbers: 45, 101, 126, 167, 187, 190, 191, 204 and 219). Similarly, amino acid positions involved in salt bridge formation within α1 and α2 were found to be conserved, with the histidine (H) residues at site 41 and 137, and the aspartic acid (D)/glutamic acid (E) residues at sites 69 and 163. The four cysteine (C) residues involved in intra-domain disulphide bridge formation within α2 and α3 (site numbers: 145, 209, 249 and 305) were conserved across all taxa included in the amino acid alignment. Finally, most vertebrates possess an NQS or NQT nitrogen-linked glycosylation acceptor site near the end of the α1 domain (Kaufman et al., 1994). Nitrogen-linked glycans play an important role in the folding and stability of classical, and likely also non-classical, MHC class I molecules (Ryan and Cobb, 2012, 2015). The amino acid sequence for nitrogen-linked glycosylation site consists of an asparagine (N), any other amino acid except proline, followed by serine (S) or threonine (T) (Ryan and Cobb, 2015). Most of the *C.*

decresii transcripts encoded NQS glycosylation sites, with three transcripts (*Ctde-UA*005* and *Ctde-UA*006*) encoding NHS instead. Together, these findings suggest that transcripts *Ctde-UA*001* – *Ctde-UA*006* produce molecules that undertake classical MHC I functions at the cell surface.

Classical MHC I gene structure (leader peptide, α domains and Tm/Cyt domains) could not be confirmed for two *C. decresii* putative MHC I transcripts; *Ctde-UA*007* and *Ctde-UA*008*. Both *Ctde-UA*007* and *Ctde-UA*008* ended prematurely; *Ctde-UA*007* was missing the 5' end of the α3 domain and all of the Tm/Cyt domains and *Ctde-UA*008* was missing both the α3 and the Tm/Cyt domains. Both *Ctde-UA*007* and *Ctde-UA*008* possessed the expected conserved sites associated with peptide binding at the α1 and α2 domains and did not contain any premature stop codons (fig. 1). These sequences could not be unambiguously resolved during transcriptome assembly, probably due to similarity with the other transcripts. This is not surprising given that MHC loci arise by periodic gene duplication and rearrangement events (Nei and Rooney, 2005). It is also possible that these transcripts are naturally truncated and encode non-classical MHC I molecules that are secreted and undertake peptide presentation away from the cell surface, given the lack of Tm/Cyt domains and the presence of conserved sites associated with peptide presentation (Carlini et al., 2016; Donadi et al., 2011; Glaberman et al., 2009). Given that *Ctde-UA*007* and *Ctde-UA*008* are likely truncated due to incomplete transcriptome assembly they were not assigned to a new locus group and were included in the 'UA' MHC I locus group.

Average nucleotide percent identity among the validated full-length *C. decresii* MHC I transcripts was 79.8%, with this differentiation primarily driven by variation at the α1 and α2 domains (table S2). Nucleotide identity within α1 and α2 was 81% and 82%, respectively, whereas nucleotide identity within α3 was 88%. Sequences *Ctde-UA*001* and *Ctde-UA*002* were identical only at the α1 domain, suggesting the possibility for exon shuffling recombination in the *C. decresii* MHC I region. Exon shuffling recombination occurs between entire exon regions with breaks within introns and has been found to occur in a range of vertebrates (Holmes and Parham, 1985; Wang et al., 2010; Zhao et al., 2013). Overall percent nucleotide identity between *Ctde-UA*001* and *Ctde-UA*002* is 95.6% and between *Ctde-UA*003* and *Ctde-UA*004 it is 95.1%. Nucleotide identity is slightly lower between

*Ctde-UA*005* and *Ctde-UA*006* at 93.7%. The high similarity between these pairs of transcripts indicates that they may be alleles at the same loci, suggesting five loci in total. Bayesian phylogenetic analysis of *C. decresii* transcripts (see section 3.3 for further details) suggest the presence of five or six loci. *Ctde-UA*001* and *Ctde-UA*002,* and *Ctde-UA*003* and *Ctde-UA**004 each clustered together but *Ctde-UA*005* and *Ctde-UA*006* didn't, with *Ctde-UA*006* located between *Ctde-UA*005* and a *P. vitticeps* sequence.

3.3. Phylogenetic analysis of Iguanian MHC class I genes

One of the Bayesian phylogenetic analyses of MHC I sequences was conducted with no partitioning of codon positions and a GTR+G model of evolution, according to PartitionFinder2 results. The other analysis was conducted using data partitioned by codon position and allowing mixed models of evolution. For both Bayesian analyses, convergence diagnostics indicated run convergence and high posterior probability values were obtained across most of the tree. The two resulting trees had identical topology but the tree resulting from the PartitionFinder2-informed analysis had slightly better overall posterior probability values and was therefore retained. Posterior probabilities were high (>0.95) for most branch nodes, but low for some outer nodes (fig. 2).

Full-length MHC I sequences were only available for three Iguania species (*Amblyrhynchus cristatus*, *Conolophus subcristatus* and *Iguana iguana*), one Dactyloidae species (*Anolis carolinensis*) and two Agamidae species (C. *decresii* and *Pogona vitticeps*, fig. 2, table S1). No orthologous relationships were observed within the Iguanidae clade, representing three genera that diverged 10 to 20 million years ago (Rassmann, 1997). Instead, MHC I sequences clustered by species, suggesting loss of ancestral diversity, recent gene duplications and potentially gene conversion events (concerted evolution, Glaberman and Caccone, 2008). In contrast, Agamidae MHC I sequences (*C. decresii* and *P. vitticeps*) displayed orthologous relationships despite similar divergence times to the iguanid species (Hugall et al., 2008). Given divergence of approximately 20 million years between *P. vitticeps* and *C. decresii* (Hugall et al., 2008), the separation of the agamid MHC I sequences into three orthologous clades suggests that at least two gene duplication events occurred greater than 20 million years ago (fig. 2). The conservation of MHC class I loci over such a time scale is not unusual; some

primate MHC I loci have been conserved for at least 46-66 million years (Piontkivska and Nei, 2003). The phylogenetic relationships within the Iguanidae and Agamidae families are consistent with the birth-and-death model of evolution; the continual gain and loss of genes (Edwards and Hedrick, 1998; Nei and Rooney, 2005). Concerted evolution via gene conversion may also play a role in generating diversity within Iguanidae and Agamidae, as evidenced by the patterns observed within Iguanidae and putative exon shuffling recombination in *C. decresii*.

4. Conclusions

A total of eight MHC I transcripts were isolated and characterised for *C. decresii* from a single individual using HiSeq next generation sequencing of thymus and spleen total RNA. Due to the complex nature of the MHC, the initial assembly was refined using an assembly technique capable of distinguishing highly similar transcripts. Six of the putative MHC I transcripts were validated as likely to encode classical MHC I molecules based on three criteria; (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function, and (iii) the absence of stop codons within coding regions. Two of the putative MHC I transcripts ended prematurely either due to transcriptome assembly restrictions (i.e. due to high sequence similarity among transcripts) or non-classical functionality (naturally truncated). Bayesian phylogenetic analysis indicated that a birth-and-death model of evolution is likely the main mechanism shaping MHC I diversity within Iguanian lizards. MHC I sequences from a wider range of squamates are required to obtain a clearer view of the mechanisms responsible for creating MHC I diversity within Squamata. This work provides a foundation for future work examining the mechanisms shaping diversity at the MHC class I region of agamid lizards.

Acknowledgments

We thank Talat Hojat Ansari and Sarah Pearson for assistance with laboratory work and data analysis, and Aaron Fenner for assistance with fieldwork. This work was funded by the Holsworth Wildlife Research Endowment and the Nature Foundation of South Australia, awarded to JH. Specimen collection and euthanasia were conducted according to the Flinders University Animal Welfare Committee (approval no. E379) and the South Australian Department of Environment, Water and Natural Resources regulations (permit no. U26225).

References

Adams, E.J., Luoma, A.M., 2013. The adaptable major histocompatibility complex (MHC) fold: Structure and function of nonclassical and MHC class I–like molecules. Annual Review of Immunology 31, 529-561.

Alfonso, C., Karlsson, L., 2000. Nonclassical MHC Class II Molecules. Annual Review of Immunology 18, 113-142.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

Ansari, T., Bertozzi, T., Miller, R., Gardner, M., 2015. MHC in a monogamous lizard - Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. Developmental and Comparative Immunology 53, 320-327.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Carlini, F., Ferreira, V., Buhler, S., Tous, A., Eliaou, J.-F., René, C., Chiaroni, J., Picard, C., Di Cristofaro, J., 2016. Association of HLA-A and Non-Classical HLA Class I Alleles. PLoS ONE 11, e0163570.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information, Proceedings of the German Conference on Bioinformatics (GCB) pp. 45-56.

Donadi, E.A., Castelli, E.C., Arnaiz-Villena, A., Roger, M., Rey, D., Moreau, P., 2011. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. Cellular and Molecular Life Sciences 68, 369-395.

Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Edwards, S., Hedrick, P., 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends in Ecology & Evolution 13, 305-311.

Ejsmond, M., Radwan, J., Wilson, A., 2014. Sexual selection and the evolutionary dynamics of the major histocompatibility complex. Proceedings of the Royal Society B: Biological Sciences 281, 1-8.

Garcia, M., Yebra, B., Flores, A., Guerra, E., 2012. The Major Histocompatibility Complex in Transplantation. Journal of Transplantation 2012, 1-7.

Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H.R., Holleley, C.E., Zhou, Y., Zhang, X., Matsubara, K., Waters, P., Graves, J.A.M., Sarre, S.D., Zhang, G., 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. GigaScience 4, 1-11.

Georges, A., O'Meally, D., Genomics@UC, 2016. The *Pogona vitticeps* genome browser (pvi1.1 Jan 2013). Institute for Applied Ecology, Canberra: University of Canberra.

Gibbons, J., 1979. The hind leg pushup display of the *Amphibolurus decresii* species complex (Lacertilia: Agamidae). Copeia 1, 29-40.

Glaberman, S., Caccone, A., 2008. Species-specific evolution of class I MHC genes in iguanas (order: Squamata; subfamily: Iguaninae). Immunogenetics 60, 371-382.

Glaberman, S., Du Pasquier, L., Caccone, A., 2009. Characterization of a Nonclassical Class I MHC Gene in a Reptile, the Galapagos Marine Iguana (*Amblyrhynchus cristatus*). PLOS ONE 3, e2859.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols 8, 1494–1512.

Holmes, N., Parham, P., 1985. Exon shuffling *in vivo* can generate novel HLA class I molecules. The EMBO Journal 4, 2849-2854.

Hugall, A.F., Foster, R., Hutchinson, M., Lee, M.S.Y., 2008. Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. Biological Journal of the Linnean Society 93.

Janeway, C.J., Travers, P., Walport, M., 2001. Immunobiology: The Immune System in Health and Disease, 5th ed. Garland Science, New York.

Kaufman, J., Salomonsen, J., Flajnik, M., 1994. Evolutionary conservation of MHC class I and class II molecules - different yet the same. Seminars in Immunology 6, 411-424.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Klein, J., Bontrop, R.E., Dawkins, R., Erlich, H.A., Gyllensten, U.B., Heise, E., Jones, P., Parham, P., Wakeland, E., Watkins, D., 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31, 217-219.

Koller, B.H., Orr, H.T., 1985. Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. Journal of immunology (Baltimore, Md. : 1950) 134, 2727-2733.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29, 1695-1701.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34, 772-773.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv:1303.3997v2 [q-bio.GN].

McLean, C.A., Stuart-Fox, D., Moussalli, A., 2014. Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. Journal of Evolutionary Biology 27, 2123-2137.

Milinski, M., 2006. The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. Annual Review of Ecology, Evolution, and Systematics 37.

NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 44, D7-D19.

Neefjes, J., Jongsma, M., Paul, P., Bakke, O., 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology 11, 823-836.

Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. PNAS 94.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annual Review of Genetics 39, 121-152.

Osborne, L., 2005a. Information content of male agonistic displays in the territorial tawny dragon (*Ctenophorus decresii*). Journal of Ethology 23, 189-197.

Osborne, L., 2005b. Rival recognition in the territorial tawny dragon (*Ctenophorus decresii*). Acta Ethologica 8, 45-50.

Pearson, S.K., Bull, C.M., Gardner, M.G., 2017. *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. Immunogenetics 69, 49-61.

Piertney, S., Oliver, M., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7-21.

Piontkivska, H., Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol 20.

Pyron, R.A., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evolutionary Biology 13, 93.

Rambaut, A., 2012. FigTree, 1.4.2 ed, Institute of Evolutionary Biology, University of Edinburgh.

Rassmann, K., 1997. Evolutionary Age of the Galápagos Iguanas Predates the Age of the Present Galápagos Islands. Molecular Phylogenetics and Evolution 7, 158-172.

Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331, 623-641.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature biotechnology 29, 24-26.

Ronquist, F., Teslenko, P., Ayres, D., Darling, S., Hohna, S., Larget, B., Liu, L., Suchard, M.H., JP, 2011. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 6, 539-542.

Ryan, S.O., Cobb, B.A., 2012. Roles for major histocompatibility complex glycosylation in immune function. Seminars in Immunopathology 34, 425-441.

Ryan, S.O., Cobb, B.A., 2015. Major Histocompatibility Complex: N-Glycosylation Form and Function, in: Taniguchi, N., Endo, T., Hart, G.W., Seeberger, P.H., Wong, C.-H. (Eds.), Glycoscience: Biology and Medicine. Springer Japan, Tokyo, pp. 643-648.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2, 16.

Spurgin, L.G., van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C., Richardson, D.S., 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Molecular ecology 20, 5213-5225.

Tamura, K., Peterson, D., Petersoin, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731-2739.

Wang, D., Zhong, L., Wei, Q., Gan, X., He, S., 2010. Evolution of MHC class I genes in two ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*). FEBS letters 584, 3331-3339.

Yewers, M.S.C., Pryke, S., Stuart-Fox, D., 2016. Behavioural differences across contexts may indicate morph-specific strategies in the lizard *Ctenophorus decresii*. Animal Behaviour 111, 329-339.

Zhao, M., Wang, Y., Shen, H., Li, C., Chen, C., Luo, Z., Wu, H., 2013. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evolutionary Biology 13, 113.

Tables and figures

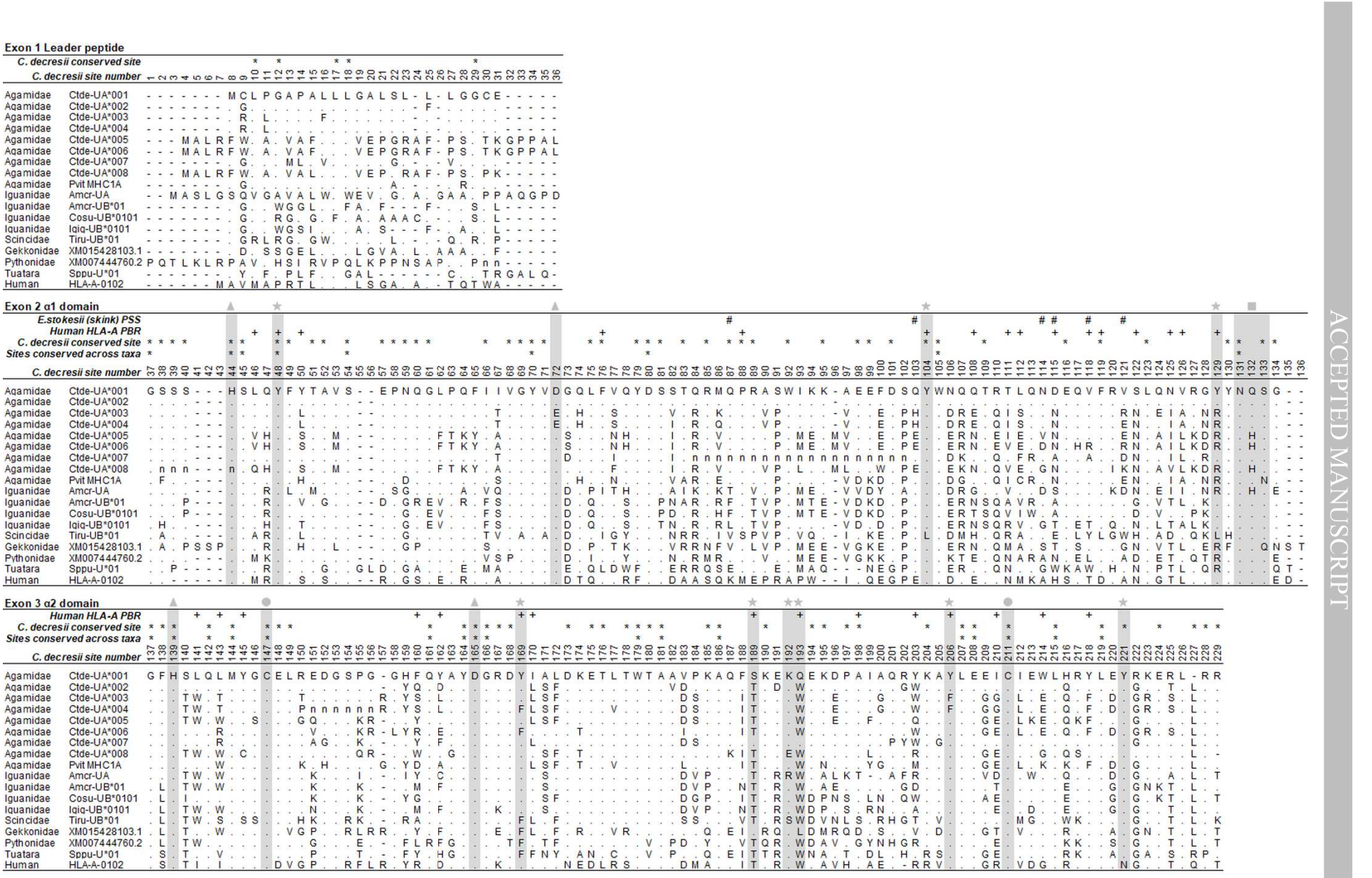

Figure 1. Amino acid alignment of tawny dragon (*Ctenophorus decresii*) MHC I transcripts with other squamates, tuatara (*Sphenodon punctatus*) and human. Coding domain separations are based on Koller and Orr (1985). Dots indicate identity with *Ctde-UA*001* and dashes indicate alignment gaps. ¥ = partial coding sequence. Residues with expected functions as per Kaufman et al. (1994) are shaded grey; *stars* = conserved peptide-binding residues of antigen N and C termini, *triangles* = salt bridge-forming residues, *circles* = disulphide bridge-forming cysteines, *square* = N-glycosylation site, CD8 = expected CD8 binding site. *Asterisks* represent conserved sites, across *C. decresii* sequences and across all taxa. Gidgee skink (*Egernia stokesii*) positively selected sites (PSS, putative peptide binding regions) (Pearson et al., 2017) and human peptide binding regions (PBR) (Reche and Reinherz, 2003) are indicated with *hashes* and *crosses*, respectively.

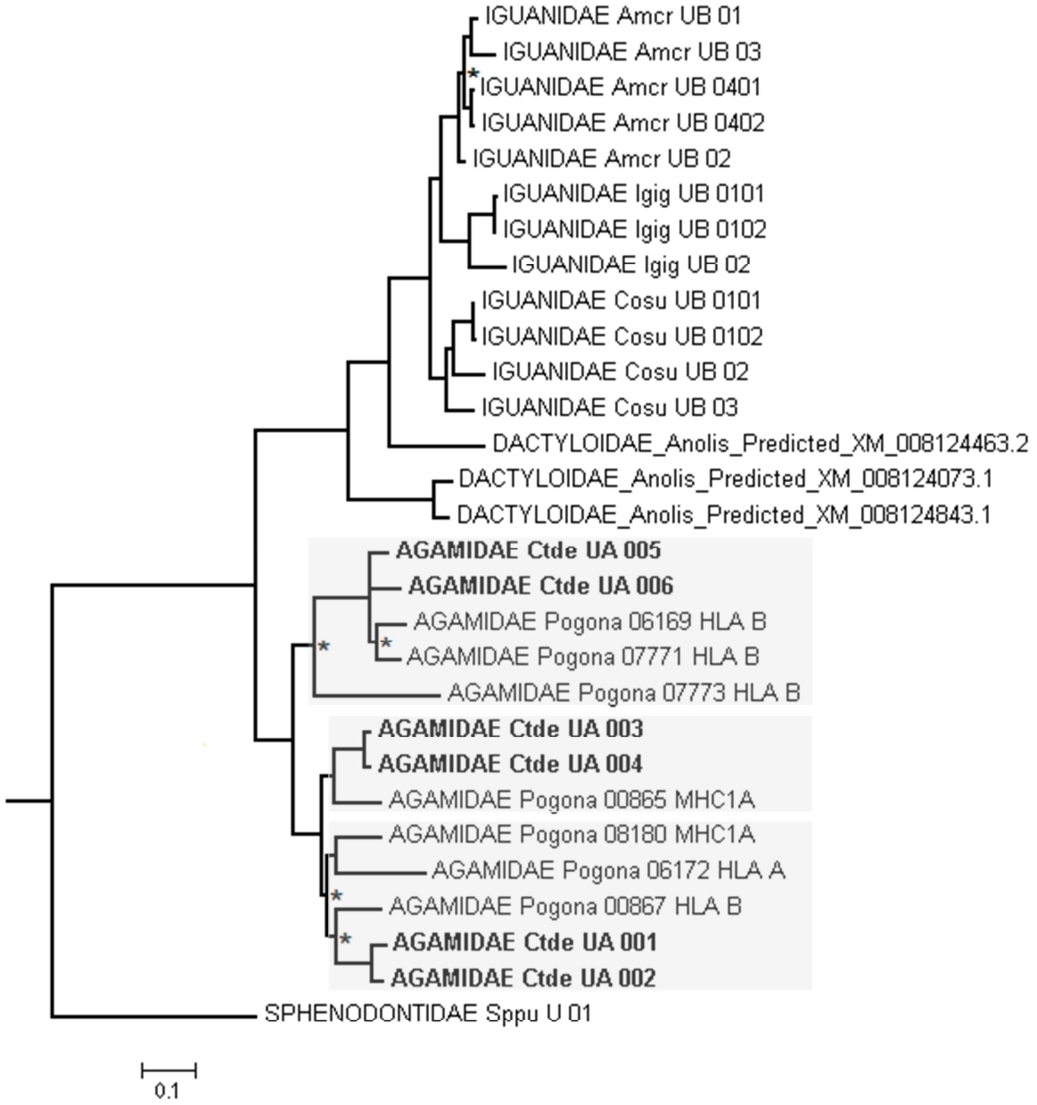

Figure 2. Bayesian phylogenetic tree of the α1, α2 and α3 domains of available full length Iguania (here, Iguanidae, Agamidae and Dactyloidae) MHC class I nucleotide sequence. Validated full length tawny dragon (*Ctenophorus decresii*) MHC class I transcripts are highlighted by bold text. The tree is rooted using tuatara (*Sphenodon punctatus*) MHC class I sequence. Three orthologous Agamidae clades are shaded grey. Nodes for which the Bayesian posterior probability <0.95 are indicated with an asterisk. The scale bar indicates the number of expected nucleotide substitutions per site. *Anolis* sequences were predicted by NCBI automated computational searches of genomic sequence (NCBI Resource Coordinators, 2016). The *Pogona* sequences were obtained from the Pogona Genome Project (Georges et al., 2015; Georges et al., 2016). Refer to table S1 for further information on sequences.

Highlights

- Six full length MHC class I transcripts were identified for an agamid lizard, *Ctenophorus decresii* (tawny dragon lizard).
- All full length transcripts were validated as likely to encode classical class I MHC molecules.
- Comparison among *C. decresii* MHC class I sequences and Bayesian phylogenetic analysis of Iguanian MHC I sequences revealed a primary role for the birth-and-death model of evolution and a potential secondary role for concerted evolution.