Developmental and Comparative Immunology 84 (2018) 164–171

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Short communication

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

Jessica Hacking [a, *], Terry Bertozzi [b, c], Adnan Moussalli [d], Tessa Bradford [a, b, c], Michael Gardner [a, b]

[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

ARTICLE INFO

*Article history:*
Received 15 December 2017
Received in revised form
10 February 2018
Accepted 10 February 2018
Available online 16 February 2018

*Keywords:*
Transcriptome assembly
Iguania
Agamidae
Ctenophorus decresii
MHC class I evolution

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 characterise 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.



## 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 $\alpha$ domains, and the transmembrane and cytoplasmic (Tm/Cyt) domains, all of which are encoded by a single gene (Kaufman et al., 1994). The $\alpha$1 and $\alpha$2 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-

*Abbreviations:* MHC, major histocompatibility complex; PBR, peptide binding region; CNV, copy number variation.

* Corresponding author.
*E-mail addresses:* jessica.hacking@flinders.edu.au (J. Hacking), terry.bertozzi@samuseum.sa.gov.au (T. Bertozzi), amoussalli@museum.vic.gov.au (A. Moussalli), tessa.bradford@samuseum.sa.gov.au (T. Bradford), michael.gardner@flinders.edu.au (M. Gardner).

https://doi.org/10.1016/j.dci.2018.02.012
0145-305X/

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 (<30 g) 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/