MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer

Hannah V. Siddle1,†, Jolanta Marzec1,†, Yuanyuan Cheng1, Menna Jones2 and Katherine Belov1,*

1Faculty of Veterinary Science, University of Sydney, New South Wales 2006, Australia
2School of Zoology, University of Tasmania, Tasmania 7001, Australia

Tasmanian devils face extinction owing to the emergence of a contagious cancer. Devil facial tumour disease (DFTD) is a clonal cancer spread owing to a lack of major histocompatibility complex (MHC) barriers in Tasmanian devil populations. We present a comprehensive screen of MHC diversity in devils and identify 25 MHC types and 53 novel sequences, but conclude that overall levels of MHC diversity at the sequence level are low. The majority of MHC Class I variation can be explained by allelic copy number variation with two to seven sequence variants identified per individual. MHC sequences are divided into two distinct groups based on sequence similarity. DFTD cells and most devils have sequences from both groups. Twenty per cent of individuals have a restricted MHC repertoire and contain only group I or only group II sequences. Counterintuitively, we postulate that the immune system of individuals with a restricted MHC repertoire may recognize foreign MHC antigens on the surface of the DFTD cell. The implication of these results for management of DFTD and this endangered species are discussed.

**Keywords:** Tasmanian devil; MHC; devil facial tumour disease; marsupial; cancer

1. INTRODUCTION

A contagious tumour, devil facial tumour disease (DFTD), threatens the world’s largest remaining marsupial carnivore, the Tasmanian devil (*Sarcophilus harrisii*) (McCallum & Jones 2006). Devils affected by DFTD develop lesions around the face and neck (Loh et al. 2006b), which can metastasize. Mortality of affected devils occurs owing to organ failure, secondary infection or an inability to feed (Loh et al. 2006a; Pyecroft et al. 2007). DFTD has affected more than 70 per cent of the total devil population, including the entire eastern half of Tasmania where it is invariably fatal, and has caused more than 90 per cent population decline. Extinction in the wild is considered to be a real possibility within 25–35 years (Lachish et al. 2007; McCallum et al. 2007), leading to an IUCN listing as ‘Endangered’ (http://www.iucnredlist.org/).

DFTD is a naturally occurring clonal tumour with a unique mode of transmission between individuals. DFTD cells in different individuals have identical, complex chromosomal rearrangements (Pearse & Swift 2006) and at microsatellite markers are genetically identical to each other but different to the host devil (Siddle et al. 2007a). Thus, the tumour is transmitted as an allograft, most probably through biting behaviour, rather than arising in each individual independently. Spread of such an infectious cell line between unrelated individuals should be prevented by major histocompatibility complex (MHC) Class I antigens on the surface of the tumour cell, as Class I molecules play a key role in self/non-self-recognition. Transplantation of cells with a different MHC type leads to rapid organ rejection in other marsupial species (Stone et al. 1997).

The extracellular α1 and α2 domains (exons 2 and 3) of the Class I gene are usually highly polymorphic, particularly at amino acid sites where the molecule interacts with foreign peptides known as the peptide binding region (Hughes & Nei 1988). Polymorphism in this region provides an individual and the population a greater chance of responding to a new pathogen in the environment (Hughes & Nei 1992). Genetic diversity at MHC loci has been used to measure the immunological fitness of a wild population and to estimate their ability to respond to new disease threats (O’Brien & Evermann 1988; Slade 1992; Hedrick et al. 2000).

Despite an efficient immune system (Woods et al. 2007; Kreiss et al. 2008), devils do not mount an immune response to DFTD. Our previous studies on 21 Tasmanian devils from eastern Tasmania showed that Tasmanian devils have reduced genetic diversity at MHC Class I (Siddle et al. 2007a). We hypothesized that a lack of MHC diversity allowed the tumour to be transmitted between individuals without immunological recognition (Siddle et al. 2007a).

DFTD was first observed at Mount William in far northeastern Tasmania in 1996 and has spread throughout the eastern half of the island. The disease front has now reached populations in the northwest that are partially geographically isolated and genetically different (Jones et al. 2004) from populations in the east. Here,
we investigate levels of MHC Class I diversity in Tasmanian devils from across their range and present data, which indicates that some animals are sufficiently different at their MHC from DFTD cells that they may recognize tumour MHC antigens as foreign.

2. MATERIAL AND METHODS

(a) Samples

Tables S1 and S3 in the electronic supplementary material summarize sample numbers, origin, MHC type and disease status for 387 devil samples.

(b) Amplification of MHC products

The MHC Class I α1 domains (exon 2) of 387 Tasmanian devil samples across Tasmania were PCR-amplified using the conditions and primers (P1 and P2) described by Siddle et al. (2007b) (P1 and P2 primer sites shown in the electronic supplementary material, figure S3) with minor modifications. PCR products were separated on a polyacrylamide gel by one strand conformation polymorphism analysis (OSCP) using the method described by Siddle et al. (2007a) with the following modifications as described by Miller et al. (2007) (detailed in the electronic supplementary material).

The PCR products from each distinct OSCP banding pattern were amplified using primers P3 and P2 (shown in the electronic supplementary material, figure S3) using the same PCR conditions as for OSCP typing, cloned and sequenced. For each individual, 20 clones were selected and sequenced. To confirm that null alleles were not the cause of copy number variation, additional sequencing was carried out on cDNA (blood) from an L-type individual using an alternative primer set (primers detailed in the electronic supplementary material, figure S3) and an M-type individual using 5′ rapid amplification of cDNA ends (RACE) (GeneRacer, Invitrogen). 5′ RACE was performed with a previously described α3 reverse primer (P4) located in the conserved α3 domain (Siddle et al. 2006) (P4 shown in the electronic supplementary material, figure S3). The optimum PCR conditions were as follows: 1 × buffer, 1.6 mm MgSO4, 200 μm dNTP, 0.5 μm 5′ RACE primer, 2 μm α3 reverse primer and 0.3 μl of high fidelity taq polymerase (Platinum DNA taq polymerase high fidelity, Invitrogen). PCR cycles were performed with initial denaturation at 94.0 °C for 3 min, followed by 30 cycles of 94.0 °C for 30 s, 65 °C for 30 s and 72 °C for 2 min, and a final extension at 72 °C for 10 min. A 750 bp band was cloned and 43 clones were selected and sequenced.

MEGA 4.0 (Kumar et al. 2004) was used to build a neighbour joining phylogenetic tree using Jukes–Cantor distance with 1000 bootstraps, replicating the methods used to build the phylogeny of marsupial Class I sequences described by Siddle et al. (2009). The GenBank accession numbers of sequences used in the analysis can be found in the electronic supplementary material.

3. RESULTS

(a) MHC diversity in Tasmanian devils

OSCP of the MHC Class I α1 domain of 387 Tasmanian devil samples revealed 25 different MHC types (types A–V and type 1) (electronic supplementary material, figure S1 and tables S1 and S2). Figure 1 shows the distribution of MHC types across Tasmania. A different pattern of MHC types is observed between eastern and northwestern Tasmanian devils. Limited MHC diversity is observed among devils in the east; over half the devils from eastern sample sites have MHC type A (24% of devils) or an identical MHC type to the tumour (1) (30% of devils). The most common devil MHC types are A, D, G or 1 and devils with these types have been shown to develop DFTD. Seventy per cent of eastern Tasmanian devils share these MHC types. By contrast, these common types constitute only 55 per cent of MHC types in the northwestern populations. Ten of the 25 MHC types are found only in the northwestern populations. Based on our sample set, no devils with MHC types B′, C, G′, H′, I, K, L, M, N, O, P, Q, S, U and V have developed DFTD (electronic supplementary material, table S1). The majority of these types are found only in the northwest, with the exception of C, G′, H′, S and V, which are found either in the east or across the range.

(b) Distribution of diversity among Class I sequences across Tasmania

To investigate the Class I alleles that constitute each MHC type, Class I sequences from each type were characterized (electronic supplementary material, tables S2 and S3). Based on sequencing of the MHC Class I α1 domain, a total of 53 nucleotide and 43 amino acid sequences were identified (amino acid alignment shown in the electronic supplementary material, figure S3). The sequences isolated all have a three amino acid deletion when compared with human leukocyte antigen-A (HLA-A), this deletion has been observed in other full length Class I transcripts isolated from marsupials (Miska & Miller 1999; Siddle et al. 2006).

Fourteen nucleotide sequences are unique to the northwest and seven are unique to the east (electronic supplementary material, figure S2). The 14 sequences unique to northwestern devils each have one or two amino acid substitutions when compared with the sequences found across the range. One substitution is at a site expected to interact with a bound peptide (position 52 of Sahai*S8) based on alignment with HLA-A (electronic supplementary material, figure S2). Four sequences were found in over half of all devils surveyed, with sequences 27 and 32 found in 88 per cent and 83 per cent of devils, respectively (electronic supplementary material, figure S4). Although there are four very common sequences, the majority of unique sequences are rare (found in less than 3% of the population). The most common sequences are found in devils across the range, but the less prevalent sequences are found predominantly in devils from the northwest of Tasmania (electronic supplementary material, figure S5).

The tumour has six Class I sequences; these are sequences 28, 32, 35, 46, 47 and 53 (electronic supplementary material, figure S5). Sequences 32 and 35 are prevalent among devils in both the northwest and east (found in 84% and 67% of devils, respectively). Sequence 35 is prevalent among eastern devils (87%), but
is only found in 60 per cent of northwest devils. Sequences 46, 47 and 53 are found in close to 30 per cent of eastern devils but only approximately 10 per cent of northwest devils. Although the tumour does not have one of the most common sequences, 27, sequences 53 and 28 differ by only one and two amino acids, respectively, from sequence 27.

Phylogenetic analysis demonstrates that the MHC Class I sequences fall into two clades (bootstrap of 92, electronic supplementary material, figure S6). Sequences within group I share between 94 and 99.5 per cent nucleotide identity and 88 and 98.7 per cent amino acid identity, while the sequences within group II share between 97 and 99.5 per cent nucleotide identity and 95 and 98.7 per cent amino acid identity. In comparison, the two groups of sequences share on average only 87 per cent nucleotide and 76 per cent amino acid identity. Most MHC types (including the tumour type) contain both

group I and group II sequences. A few key exceptions occur. Devils with types H and I contain only group I sequences. Individuals with these types are found in the northwest and in Bronte Park. Types B', L, M and V contain only group II sequences. Individuals with these types are found only in the northwest. DFTD has not been found in devils with any of these types.

The numbers of Class I sequences isolated from each individual vary between two and seven, suggesting that the number of MHC Class I loci ranges from one to at least four and a maximum of seven loci per individual. The possibility of null alleles was considered and different primer combinations and 5‘ RACE (using a conserved reverse primer located in the 3‘ domain) were used to capture any missed sequences from M type and L type individuals. Only previously identified MHC Class I cDNA domains belonging to group II were amplified. No group I sequences were identified. The RACE experiment also resulted in the amplification of a divergent Class I transcript. This transcript is a non-classical locus based on its phylogenetic relationship to other marsupial non-classical Class I genes and its lack of polymorphism (electronic supplementary material, figure S7). We concede that in individuals with a high number of MHC variants, some alleles may have been missed, given that only 20 clones per individual were sequenced.

4. DISCUSSION

Here, we describe levels of MHC Class I diversity across the Tasmanian devil range, and show that while there is a greater number of MHC Class I sequences in northwestern populations than in eastern populations, variation in devil MHC is owing to copy number variation rather than sequence polymorphism.

(a) MHC Class I diversity across the devil range

Devil MHC Class I diversity is limited. Although few investigations of MHC Class I polymorphism have been carried out on marsupials, MHC Class II β genes (DAB) have been examined in the brushtail possum (Trichosurus vulpecula) (Holland et al. 2008), and two species of South American marsupials, the Brazilian gracile mouse opossum (Gracilinanus microtarsus) and the grey slender mouse opossum (Marmosa incanus) (Meyer-Lucht et al. 2008). In the brushtail possum and gracile opossum, significant diversity was found at DAB loci, with 32 sequences (across a minimum of three loci) and 47 sequences (across a minimum of four loci), respectively. By contrast, only eight sequence variants were found across two loci in the slender mouse opossum. Here, we identified 53 sequence variants over at least four loci, but the level of sequence divergence between variants was low. In addition, common variants were seen in the majority of devils.

In the context of the spread of DFTD, differences between Class I sequences found in eastern and northwestern Tasmania are of interest. Fourteen unique Class I nucleotide sequences were identified in northwestern devils. However, these sequences differ from eastern sequences by only one or two amino acids and of the seven new amino acid substitutions found among northwestern sequences only one substitution is located within a putative peptide interaction site. Rejection of transmissible tumours in experimental models can occur owing to only minor changes in MHC antigens or minor histocompatibility antigens (Sanford et al. 1973; Cho et al. 2007). However, given that Tasmanian devils with DFTD do not always have identical MHC Class I sequences to the tumour, the limited number of additional amino acid substitutions identified among northwestern devils may not affect DFTD transmission.

The devil classical Class I sequences can be clearly divided into two subsets (group I and II) based on sequence similarity and evolutionary relatedness (electronic supplementary material, figures S2 and S6). The sequences within each clade are highly similar (greater than 88% amino acid identity), but between clades are different (on average 76% amino acid identity). It is likely that group I and group II sequences belong to distinct loci, but this remains to be confirmed.

(b) Tasmanian devils display allelic copy number variation

Individual devils have 0–6 group I sequence variants and 0–4 group II sequences. It is feasible that alleles may have been missed and that null alleles exist; however, the use of an alternative primer set and 5′ RACE using a conserved reverse primer suggests that the difference in sequence variants identified per individual is owing to copy number variation, and not methodological artefacts. Variation in the number of Class I sequences in individuals within a species is not unique to Tasmanian devils and has been reported in cattle, rats, mice, Rhesus macaques and pigs (Amadou et al. 1999; Otting et al. 2005; Roos & Walter 2005; Birch et al. 2006; Tanaka-Matsuda et al. 2009). Copy number variation is thought to replace allelic polymorphism in some species such as the Rhesus macaque (Otting et al. 2005) and RT1-CE region in the rat (Roos & Walter 2005). In the devil, it appears that either entire loci are missing from some haplotypes (as the same sequences were amplified from both DNA and cDNA) or that alleles at recently duplicated loci have become fixed. It is noteworthy that in the tammar wallaby, MHC Class I genes are not linked to the MHC Class II and III regions (Siddle et al. 2009).

Molecular dating of wallaby Class I genes indicates that gene movement probably occurred after the American and Australian marsupials diverged and before the Australian marsupial radiation. Thus, it is possible that devil Class I genes are also unlinked to the MHC.

The majority of devils across the range (approx. 80%) contain both group I and group II sequences, as do DFTD cells. Twenty per cent of individuals have restricted MHC repertoires: II and I types contain only group I sequences, while B’, L, M and V contain only group II sequences. The level of variation between the group I and group II sequences is significant. Based on this observation, we propose that the immune system of devils that have a restricted MHC repertoire should be able to recognize foreign MHC antigens on the surface of DFTD cells and mount an immune response against them. That is, individuals that have only group I alleles should recognize foreign group II molecules on the surface of DFTD cells, leading to immune response. Similarly, individuals with only group II sequences should recognize foreign DFTD group I molecules.
Interestingly, in our dataset, individuals with group I sequences only (types H and I—found only in the northwest and Bronte Park) or group II sequences only (B, L, M and V—found only in the northwest) have not developed DFTD. Most of these animals have not yet been exposed to the disease, as DFTD has not reached the northwest of Tasmania, but West Pencil Pine and Bronte Park are the first populations containing individuals with restricted MHC repertoires to be exposed to DFTD. We are monitoring West Pencil Pine to determine whether animals with restricted MHC repertoires succumb to disease. So far, there is no evidence of this (A. Lane, J. Marzec & K. Belov 2010, personal communication). Intriguingly, disease prevalence and population impacts appear to be lower than expected at West Pencil Pine (R. Hamede 2010, personal communication).

(c) Evidence for a selective sweep in the past
Microsatellite markers show low levels of genetic diversity across the devil range with lower levels of genetic diversity in the northwest than in the east (Jones et al. 2004). However, among eastern devils, a selective sweep appears to have affected MHC Class I genes leading to greater uniformity in allelic content and gene copy number. It is possible that a prior disease epidemic in eastern populations led to selection of MHC types with protective function. There is (unconfirmed) speculation in the literature that the larger dasyurids were affected by an epizootic disease, perhaps pleural pneumonia or a distemper-like disease in the early twentieth century (Guiler 1985, 1992; Guiler & Godard 1998).

(d) Management of Tasmanian devils in captivity and the wild
Based on the results presented here, it appears that animals with restricted MHC repertoires may be resistant to the disease and may survive the disease epidemic. If this hypothesis proves correct, it will have a significant impact on management of the species.

A Tasmanian devil insurance programme has been established to breed devils in captivity with a long-term aim of re-introducing them to the wild. The project aims to represent 95 per cent of all genetic diversity in the wild devil population and we argue that all MHC types should also be retained. It has been argued that individual genes should not be used as the sole determinants for selection of founders for captive breeding programmes as this can lead to loss of genetic diversity at other potentially important genes (Miller 1995). The counter-argument is that captive colonies need to be representative of genetic diversity in the wild and that they provide an opportunity to make sure that unique MHC alleles are not lost (Hughes 1991). The programme will not breed ‘resistant’ devils, as this would actually lead to a loss of genetic diversity in the species and would increase the vulnerability of the insurance population to new infectious diseases.

The implications of these results for management of the devils in the wild are counterintuitive. Traditionally, high MHC diversity is associated with increased immunological fitness. Here, it is possible that devils with restricted MHC diversity are more resistant to DFTD. If this proves to be correct, then genetic rescue programmes are likely to be detrimental. Mixing eastern and northwestern populations will lead to an increased chance that offspring contain both group I and group II sequences. Isolation of northwestern populations and disease suppression remains the best options for management of the disease in the wild.

Here, we confirm that overall levels of MHC diversity in Tasmanian devils are low, but present evidence of MHC variation owing to copy number differences. Most importantly, we show that most animals have MHC types that are similar to the DFTD MHC type, while some do not. We speculate that individuals with a restricted MHC repertoire may be capable of mounting an immune response to foreign MHC antigens on DFTD cells.

Ear biopsies from Tasmanian devils were collected under Animal Ethics permits DP1W07/08, UTAS AEC No. A0010296 and ANU AEC No. F.BTZ.11.00.

This project was funded by the Australian Research Council, an Eric Guiler grant, the Tasmanian Department of Primary Industries and Water, donations to the Veterinary Science Foundation and the Elizabeth Gabler Charitable Trust through the Trust Company. H.S. was supported by a UPA scholarship and Y.C. by an EIPRS scholarship. We thank the Save the Devil team for samples and insightful conversations. Particular thanks to Greg Woods, Alex Kreiss, Rodrigo Hamede, Amanda Lane, Shelly Lachish, Billie Lazenby, Dydee Mann, Chrissie Pukk, Fiona Hume, Jason Wiersma, Jim Richley, Clare Hawkins, Shaun Thurstans, Willow Farmer, Paul Humphrey and Erin Noonan. We also thank two anonymous reviewers for comments that improved this manuscript.

Data deposition footnote: Tasmanian devil MHC Class I sequences (Sahal*27-Sahal*85) have been assigned GenBank numbers GQ411435–GQ411493. Sahal*lb1 *b7 have been assigned GenBank numbers GU363942–GU363948.

REFERENCES

Holland, O. J., Cowan, P. E., Gleeson, D. M. & Chamley, L. W. 2008 High variability in the MHC Class II DA β chain of the brushtail possum (*Trichosurus vulpecula*).


