New Insights into the Role of MHC Diversity in Devil Facial Tumour Disease

Amanda Lane[1,9], Yuanyuan Cheng[1,9], Belinda Wright[1,2], Rodrigo Hamede[3], Laura Levan[1,4], Menna Jones[3], Beata Ujvari[1], Katherine Belov[1]*

1 Faculty of Veterinary Science, University of Sydney, Sydney, Australia, 2 Department of Biological Sciences, Macquarie University, Sydney, Australia, 3 School of Zoology, University of Tasmania, Hobart, Australia, 4 School of Veterinary Medicine and Science, University of Nottingham, Nottingham, United Kingdom

Abstract

Background: Devil facial tumour disease (DFTD) is a fatal contagious cancer that has decimated Tasmanian devil populations. The tumour has spread without invoking immune responses, possibly due to low levels of Major Histocompatibility Complex (MHC) diversity in Tasmanian devils. Animals from a region in north-western Tasmania have lower infection rates than those in the east of the state. This area is a genetic transition zone between sub-populations, with individuals from north-western Tasmania displaying greater diversity than eastern devils at MHC genes, primarily through MHC class I gene copy number variation. Here we test the hypothesis that animals that remain healthy and tumour free show predictable differences at MHC loci compared to animals that develop the disease.

Methodology/Principal Findings: We compared MHC class I sequences in 29 healthy and 22 diseased Tasmanian devils from West Pencil Pine, a population in north-western Tasmania exhibiting reduced disease impacts of DFTD. Amplified alleles were assigned to four loci, *Saha-UA*, *Saha-UB*, *Saha-UC* and *Saha-UD* based on recently obtained genomic sequence data. Copy number variation (caused by a deletion) at *Saha-UA* was confirmed using a PCR assay. No association between the frequency of this deletion and disease status was identified. All individuals had alleles at *Saha-UD*, disproving theories of disease susceptibility relating to copy number variation at this locus. Genetic variation between the two sub-groups (healthy and diseased) was also compared using eight MHC-linked microsatellite markers. No significant differences were identified in allele frequency, however differences were noted in the genotype frequencies of two microsatellites located near non-antigen presenting genes within the MHC.

Conclusions/Significance: We did not find predictable differences in MHC class I copy number variation to account for differences in susceptibility to DFTD. Genotypic data was equivocal but indentified genomic areas for further study.

Introduction

The Tasmanian devil (*Sarcophilus harrisii*) is threatened with extinction in the wild due to the emergence of a contagious cancer known as Devil Facial Tumour Disease (DFTD). Since its emergence 15 years ago, devil populations have declined by at least 95% in the longest diseased areas and there are concerns that DFTD will lead to the extinction of this endangered marsupial carnivore within 35 years [1,2]. DFTD is highly unusual in that it is transmitted as an infectious cell line (an allograft) [3–5] which spreads as devils bite and injure each other during the course of communal feeding on carcasses and during social and reproductive interactions [6,7]. The primary DFTD tumours usually begin on the face or inside the mouth and quickly develop into large globular tumours, that ulcerate and become friable and metastasize to internal organs [4,8]. DFTD is invariably fatal, usually within 3–6 months of the clinical presentation of visible tumours [9]. This has clear impacts, not just on devil population size and growth rates [10], but also age class structuring [11] and important life history traits including dispersal [12], sex ratios [11] and age at first reproduction [13].

One of the great challenges in managing this disease is that the cancer and its host are highly similar, since DFTD evolved from a Tasmanian devil Schwann cell (Schwann cells are part of the peripheral nervous system) [14]. Tasmanian devils are renowned for their lack of diversity across the genome [15,16] including at a key gene region involved in immune response, the Major Histocompatibility Complex (MHC) [5]. The MHC plays a

critical role in self/non-self recognition. MHC molecules can recognise antigenic peptides or cancer antigens and present them to cytotoxic T cells for destruction. MHC molecules can also act as antigens themselves, which is why MHC typing is used to select appropriate donors for tissue transplantation. Due to low levels of genetic diversity, devils frequently share the same MHC antigens as the DFTD tumour and hence it has been suggested that the tumour MHC antigens are not recognised by the devils as non-self [5]. Consequently, individual devils do not mount an immune response to DFTD [17]. We previously suggested that both the naturally existing contagious cancers, DFTD and Canine Transmissible Venereal Tumour (CTVT) evolved in populations of devils and wolves respectively that lacked MHC diversity [18]. It was postulated that these animals were more likely to be able to accept grafts from unrelated (but MHC identical) animals due to a lack of histocompatibility barriers. However, recent skin graft experiments have shown that MHC-similar devils will reject skin grafts [19]. This demonstrates that the low MHC diversity of devils does not necessarily mean that these animals are unable to identify foreign tissues. However, it does not address the possible influence that increased diversity at the MHC may have on reducing disease transmission.

Our research group previously surveyed sequence polymorphism in Tasmanian devils across their range and identified a total of 53 MHC class I sequence variants [20]. Between two and seven MHC class I variants were amplified from each individual, but without genomic information it was not possible to assign variants to loci. Recent sequencing of bacterial artificial chromosomes containing MHC inserts has allowed us to elucidate the gene content and organisation of the devil MHC. The Tasmanian devil MHC contains three classical class I genes (Saha-UA, Saha-UB and Saha-UC) [21]. These loci share extremely high amino acid identity in both exons (>98.3%) and introns (>97.7%) and account for all sequence variants which we previously designated "group 1" [20]. The sequences previously classified as "group 2" belong to a single locus designated Saha-UD, which shows some features of a nonclassical class I gene such as significantly lower levels of polymorphism [21]. Comparison of devil MHC haplotypes revealed a deletion within the Saha-UA gene, rendering it a pseudogene in certain haplotypes [21]. This deletion explains, to some extent, why a differing number of sequence variants have been detected between individuals.

It was previously suggested that individuals with a more restricted MHC complement (presence of only "group 1" or only "group 2" alleles) may be better able to recognize the foreign antigens on the surface of DFTD cells [20]. This is because the tumour (which contains genes which encode both group 1 and group 2 proteins) will therefore contain foreign MHC antigens which are not present in the host. We test this hypothesis by using individuals from a region of north-western Tasmania, West Pencil Pine ('WPP' hereafter). There is population genetic structuring across the island of Tasmania reflecting reduced gene flow between devils in the large, well connected eastern population and the more remote north-west region which is partially geographically isolated by large swathes of unsuitable alpine and wet forest habitat [22]. Devils in this region are genetically distinct from those in the east of Tasmania at both neutral loci [15,16] and at MHC genes [20]. The devils at WPP are the first population to show reduced disease effects at both population and individual levels, with low disease prevalence, increased survival time of infected individuals, and little indication of changes in population size, population growth rate or age structure, four years after the disease was first detected [23]. Here we test the hypothesis that MHC class I differences can explain the variant epidemiology and

host effects of DFTD at WPP. We do this by comparing the MHC profiles of devils that have presented with tumours with those of healthy older devils in the population (>3 years of age) who have not developed tumours. The rationale for selecting older individuals for the sample of uninfected devils was that they have had maximum time within the population to come into contact with DFTD without developing the disease. In eastern populations, virtually all devils have become infected and died of DFTD before the age of three [1,10]. Data on contact rates and biting behavior of devils at WPP suggests that it is unlikely that mature adults that are mating and fighting would not have come into contact with DFTD by this age [6, R.H. unpublished data], particularly as the population is shown to act as one social unit [24].

We discover that mismatches in previously published polymerase chain reaction (PCR) primers result in inconsistent amplification of MHC alleles and turn to MHC-linked microsatellites to look for differences between the affected and unaffected groups, which proves to be an efficient alternative to cloning-and-sequencing based MHC typing in investigating genetic diversity and natural selection at MHC loci [25–27].

Results and Discussion

Tasmanian Devil MHC Class 1 Sequence

By establishing the presence/absence of alleles at the Saha-UA and Saha-UD loci via PCR assays, we detected no significant differences in MHC class I copy number between 22 diseased devils and 29 healthy controls from WPP. We propose that previous findings to the contrary were the result of inconsistent primer amplification. Sequencing of the MHC class I peptide binding region identified 27 different alleles including six variants previously undescribed. The six new alleles all clustered neatly within what was formerly referred to as the "group 1" clade of sequences (according to Siddle et al) [20] (Saha-UA, UB and UC loci according to Cheng et al) [21] and have been assigned GenBank accession numbers JN397396–JN397401.

Alleles were assigned to four loci (Saha-UA, UB, UC and UD) based on phylogenetic analysis (Table 1). Sequences associated with the Saha-UD locus were more divergent from the other loci, but overall sequence variation was still low and characterized by nucleotide identities of between 85–99% and amino acid identities of 74–100%. Saha-UD had the fewest alleles (four) and the highest within group nucleotide identities of 98–99% and amino acid identities of 96–100%. Locus Saha-UB had the largest number of alleles (ten) and the lowest within-group nucleotide identities of 96–99% and amino acid identities of 91–99%. Saha-UA and UC had six and seven alleles respectively and intermediate levels of nucleotide diversity. When alleles were translated into amino acids, all were non-synonymous with the exception of SahaI*32 and SahaI*39, which produced identical amino acid profiles (Table S1).

PCR primer efficiency was assessed for each locus, revealing the highest amplification failure rate of 43.14% at Saha-UD (Table 1). Locus specific primers for Saha-UD showed that all individuals had at least one allele and therefore previously reported copy number variation at this locus ('group 2') was the result of primer mismatches. Primer design represents one of the major technical challenges in MHC studies [28,29]. In our case, the primer mismatches (two in Saha-UA, UB and UC, four in Saha-UD) were identified only after recent sequencing of BAC contigs for the devil MHC regions [21]. In the devil, copy number variation exists between different MHC haplotypes due to a deletion at the Saha-UA locus, making it more difficult to resolve whether differing

Table 1. Tasmanian devils sequenced using the MHC class I primers developed by Siddle et al [41] and screened for a deletion at the *Saha-UA* locus in this same gene region.

Disease status	Name	Deletion at *Saha-UA* locus	Alleles at *Saha-UA* locus		Alleles at *Saha-UB* locus		Alleles at *Saha-UC* locus		Alleles at *Saha-UD* locus	Missing data
Healthy	Allende	✓	Sahal*35		Sahal*49		Sahal*27			Saha-UD
	Baguette	χ	Sahal*29	Sahal*38			Sahal*27	Sahal*28	Sahal*32	
	Cartagena	χ	Sahal*35		Sahal*46	Sahal*49	Sahal*27	Sahal*96	Sahal*32	
	Chicomunita	χ	Sahal*35	Sahal*93	Sahal*46	Sahal*49	Sahal*27	Sahal*28		Saha-UD
	Chinquihue	χ			Sahal*36	Sahal*37				Saha-UA, UC, UD
	Delfina	χ	Sahal*29	Sahal*35			Sahal*27	Sahal*28		Saha-UB, UD
	Elvira	✓	Sahal*35		Sahal*49		Sahal*27		Sahal*32	
	Estrella	✓	Sahal*35		Sahal*49		Sahal*27			Saha-UD
	Evaristo	χ	Sahal*29				Sahal*27	Sahal*28	Sahal*32	Saha-UB
	Gengibre	χ			Sahal*48		Sahal*27	Sahal*53	Sahal*32	Saha-UA
	Huenchullan	✓	Sahal*35		Sahal*49		Sahal*27		Sahal*32	
	Iquique	χ	Sahal*35				Sahal*28		Sahal*39	Saha-UB
	Lascruces	✓					Sahal*27		Sahal*32	Saha-UB
	Lautaro	χ	Sahal*29	Sahal*35	Sahal*47		Sahal*27	Sahal*28		Saha-UD
	Limache	✓	Sahal*35		Sahal*49		Sahal*27	Sahal*91	Sahal*32	
	Mapuche	χ	Sahal*29	Sahal*35	Sahal*47		Sahal*27		Sahal*32	
	Melipilla	χ	Sahal*35		Sahal*49		Sahal*53	Sahal*96		Saha-UD
	Mirasol	χ			Sahal*37					Saha-UA, UC, UD
	Okapi	✓					Sahal*27		Sahal*32	Saha-UB
	Pomaire	✓	Sahal*29				Sahal*27	Sahal*28	Sahal*32	Saha-UB
	Puyehue	χ	Sahal*35		Sahal*49					Saha-UC, UD
	Racelette	✓			Sahal*36	Sahal*94	Sahal*27		Sahal*32	
	Sanantonio	✓	Sahal*35		Sahal*46		Sahal*28		Sahal*32	
	Sanvicente	χ	Sahal*29	Sahal*35			Sahal*28		Sahal*32	
	Segundo	χ	Sahal*29	Sahal*35	Sahal*49	Sahal*95	Sahal*27	Sahal*28	Sahal*32	
	Timoteo	✓			Sahal*36	Sahal*37	Sahal*27		Sahal*32	
	Trancura	✓	Sahal*35		Sahal*49		Sahal*27		Sahal*32 Sahal*75	
	Veronika	χ	Sahal*38		Sahal*49	Sahal*80	Sahal*27	Sahal*28	Sahal*32	
	Vieiochoco	χ	Sahal*35		Sahal*49		Sahal*27			Saha-UD
Diseased	Aconcagua	χ	Sahal*29	Sahal*35	Sahal*49		Sahal*27			Saha-UD
	Amor	χ	Sahal*35		Sahal*49		Sahal*27		Sahal*32	
	Arica	✓	Sahal*35				Sahal*28		Sahal*32	Saha-UB
	Calafquen	χ	Sahal*35	Sahal*92	Sahal*49		Sahal*27	Sahal*28	Sahal*32	
	Cassolette	✓	Sahal*35		Sahal*46		Sahal*28		Sahal*32 Sahal*67	
	Concon	✓					Sahal*27		Sahal*32	Saha-UB
	Copiapo	χ			Sahal*46	Sahal*49	Sahal*27	Sahal*28		Saha-UA, UD
	Curanto	✓	Sahal*35		Sahal*47					Saha-UC, UD
	Fabulosa	✓	Sahal*29		Sahal*80		Sahal*27	Sahal*28		Saha-UD
	Infima	✓			Sahal*49		Sahal*27	Sahal*28		Saha-UD
	Loncoche	✓	Sahal*35		Sahal*49	Sahal*95				Saha-UC, UD
	Malagente	✓	Sahal*29				Sahal*27	Sahal*28		Saha-UB, UD
	Marquez	✓					Sahal*27		Sahal*32	Saha-UB
	Mistral	✓	Sahal*29				Sahal*27	Sahal*28	Sahal*32	Saha-UB
	Negrita	χ	Sahal*29	Sahal*35	Sahal*49		Sahal*30			Saha-UD
	Olmue	χ	Sahal*35		Sahal*33		Sahal*28	Sahal*91		Saha-UD
	Princessa	χ	Sahal*35	Sahal*92	Sahal*49		Sahal*27	Sahal*28	Sahal*32	
	Renaca	✓	Sahal*35		Sahal*46		Sahal*28			Saha-UD
	Ruudgullit	χ	Sahal*29	Sahal*79	Sahal*36	Sahal*94	Sahal*27		Sahal*32	

Table 1. Cont.

Disease status	Name	Deletion at Saha-UA locus	Alleles at Saha-UA locus		Alleles at Saha-UB locus		Alleles at Saha-UC locus		Alleles at Saha-UD locus	Missing data
	Saucisette	χ	Sahal*35	Sahal*92	Sahal*48	Sahal*49	Sahal*74			Saha-UD
	Tartiflette	χ	Sahal*29	Sahal*35	Sahal*49		Sahal*27	Sahal*28	Sahal*32	
	Tranamil	χ			Sahal*46		Sahal*28			Saha-UA, UD
	Summary	**45.01% deleted**	**9.80% missing**	**25.49% missing**	**9.80% missing**	**43.14% missing**				

In each individual devil the loci which can be confirmed as failing to amplify are identified (as indicated by no alleles present). No significant differences were found between healthy and diseased animals in the prevalence of the Saha-UA deletion or in the frequency with which alleles failed to amplify at any loci.
doi:10.1371/journal.pone.0036955.t001

numbers of alleles/genes are real or simply the result of primer inefficiency. Due to the high sequence similarity of Saha-UA, Saha-UB and Saha-UC it is impossible to develop locus specific primers for these loci. Previous sequencing of the devil MHC provided no evidence of a deletion at either Saha-UB or Saha-UC [21] and further work is underway to determine if copy number variations observed at these loci are real, or the result of poor sequence amplification. No significant differences were found between healthy and diseased devils in the frequency with which alleles failed to amplify at any of the four loci (Saha-UA p = 1.0; Saha-UB p = 0.76; Saha-UC p = 1.0; Saha-UD p = 0.17).

A specific PCR test was used to determine the frequency of copy number variation of Saha-UA as a result of a deletion at this locus. A deletion at Saha-UA occurred in 45.01% of all devils (Table 1). There was no significant difference between the two groups (healthy and diseased) in the frequency of this deletion (p = 0.58). It should be noted that the PCR test used to detect the deletion cannot resolve whether the deletion occurs in one or both haplotypes in the individual. Therefore, the association between missing one or two Saha-UA in a devil and the extent of DFTD susceptibility remains to be determined.

Given the difficulties in differentiating real copy number variations from artifactual variations, we chose to employ MHC-linked microsatellite markers to assist in MHC typing healthy and diseased devils. These microsatellites are linked to various genes within the MHC region, including markers located close to each of the four MHC class I loci (Saha-UA, Saha-UB, Saha-UC, Saha-UD: Fig. 1) [30].

Population Genetics Using Neutral and MHC-linked Microsatellite Markers

Nine neutral microsatellite markers provided a background for assessing selection on the MHC [31]. All neutral markers conform to Hardy-Weinberg expectations and display similar numbers of alleles, allele frequencies and levels of heterozygosity for diseased and healthy animals (Table 2). Population pairwise

$F_{ST} = 0.005$ (p = 0.21±0.004) indicates no significant difference between diseased and healthy animals at these loci. As no demographic factors such as migration or assortative mating are evident in the neutral markers, any deviations from Hardy-Weinberg equilibrium observed at MHC-linked loci may be attributed to selection.

The eight MHC-linked microsatellite markers also display similar levels of heterozygosity, allele frequencies and identical numbers of alleles for diseased and healthy devils (Fig. 2, Table 2). Linkage disequilibrium was high for almost all MHC-linked loci, which is expected as they all map to chromosome 4 q [30] (Fig. 1). While heterozygosity was relatively high for some loci (Sh-I01, Sh-I07, Sh-I08), Hardy-Weinberg expectations held for all but one locus (Sh-I07) for healthy devils only (p = 0.003, Table 2). This remained significant after sequential Bonferroni correction. In sum, these results suggest no strong heterozygote advantage nor the presence of DFTD-resistance alleles. Further, population pairwise $F_{ST} = 0.0007$ (p = 0.41±0.006) indicates no differentiation between diseased and healthy devils.

The frequency of genotypes in the population was also assessed. This measure is particularly sensitive to changes over short time scales (allele frequencies take longer to change in a population than genotype frequencies), which is relevant considering the recent arrival of DFTD in the region. Microsatellite markers located closest to antigen presenting MHC class I loci (Sh-I01, Sh-I02, Sh-I05, Sh-I06, Sh-I10, Sh-I11) had low allelic (Fig. 2) and genotype diversity (Fig. 3a), possibly the result of a selective sweep due to a prior disease epidemic [32,33]. Two markers (Sh-I07 and Sh-I08) are located within the MHC but are closest to non-antigen presenting genes, MTCH1 and FGD2 (Fig. 3). Both loci have considerably higher polymorphism with six alleles each and 10–15 genotypes (mean 12.5), compared to 2–4 alleles (mean 2.83) and 3–7 genotypes (mean 4.67) in the six remaining markers which are located closest to genes involved in antigen presentation.

Figure 1. Location of eight MHC-linked microsatellites on devil chromosome four, associated with the MHC region. Six of these (Sh-I01, Sh-I02, Sh-I05, Sh-I06, Sh-I10 and Sh-I11) are located close to the four MHC class I loci (Saha-UA, Saha-UB, Saha-UC and Saha-UD) and several other genes involved in antigen presentation (TAP1, TAP2, PSMB8, PSMB9). The two remaining markers (Sh-I07, Sh-I08) are more closely linked with genes within the MHC that do not play a direct role in antigen presentation (MTCH1, FGD2).
doi:10.1371/journal.pone.0036955.g001

Table 2. Neutral and MHC-linked microsatellite loci summary statistics for diseased (shaded rows) and healthy devils.

Neutral loci					MHC-linked loci				
Locus	FIS	A	Ho/He	HWE p value (±s.d.)	Locus	FIS	A	Ho/He	HWE p value (±s.d.)
Sh2i	0.215	3	0.44/0.56	0.388 (±0.001)	ShI01	−0.240	3	0.82/0.67	0.333 (±0.001)
	0.274	3	0.28/0.38	0.059 (±0.001)		−0.018	3	0.68/0.67	0.134 (±0.001)
Sh2g	−0.048	3	0.50/0.48	0.723 (±0.001)	ShI02	0.051	2	0.41/0.43	1.000 (±0.000)
	0.091	3	0.45/0.50	0.137 (±0.001)		0.171	2	0.39/0.47	0.435 (±0.002)
Sh2v	0.046	5	0.68/0.71	0.111 (±0.001)	ShI05	0.102	2	0.38/0.42	0.660 (±0.002)
	0.017	4	0.67/0.68	0.916 (±0.001)		0.277	2	0.32/0.44	0.209 (±0.001)
Sh5c	−0.099	3	0.59/0.54	0.893 (±0.001)	ShI06	−0.051	2	0.48/0.46	1.000 (±0.000)
	0.194	3	0.40/0.49	0.323 (±0.001)		0.030	2	0.45/0.47	1.000 (±0.000)
Sh6e	0.010	2	0.33/0.34	1.000 (±0.000)	ShI07	−0.107	5	0.79/0.71	0.753 (±0.001)
	−0.191	2	0.34/0.29	0.556 (0.001)		−0.092	5	0.76/0.70	**0.003 (±0.000)***
Sh6L	0.315	2	0.25/0.36	0.119 (±0.001)	ShI08	−0.053	6	0.76/0.72	0.940 (±0.001)
	0.142	2	0.36/0.42	0.640 (±0.002)		−0.040	6	0.81/0.78	0.169 (±0.001)
Sh2L	−0.059	2	0.14/0.14	1.000 (±0.000)	ShI10	0.097	3	0.45/0.50	0.465 (±0.002)
	−0.019	2	0.07/0.07	1.000 (±0.000)		−0.120	4	0.31/0.28	1.000 (±0.000)
Sh2p	−0.104	4	0.68/0.62	0.889 (±0.001)	ShI11	−0.007	4	0.64/0.64	0.465 (±0.002)
	−0.014	3	0.46/0.46	1.000 (±0.000)		0.290	3	0.45/0.63	0.153 (±0.001)
Sh3a	−0.027	2	0.44/0.43	1.000 (±0.000)					
	0.006	2	0.48/0.48	1.000 (±0.000)					

A single locus (Sh-I07) displays a departure from Hardy-Weinberg equilibrium for healthy devils only. Numbers of alleles and levels of heterozygosity are very similar for healthy and DFTD infected devils though with a slight trend for higher heterozygosity in infected devils.
*denotes statically significant departure from Hardy-Weinberg equilibrium (HWE). FIS indicates homozygote (+ve) or heterozygote (−ve) excess. A is the number of alleles. Ho/He is observed and expected heterozygosities.
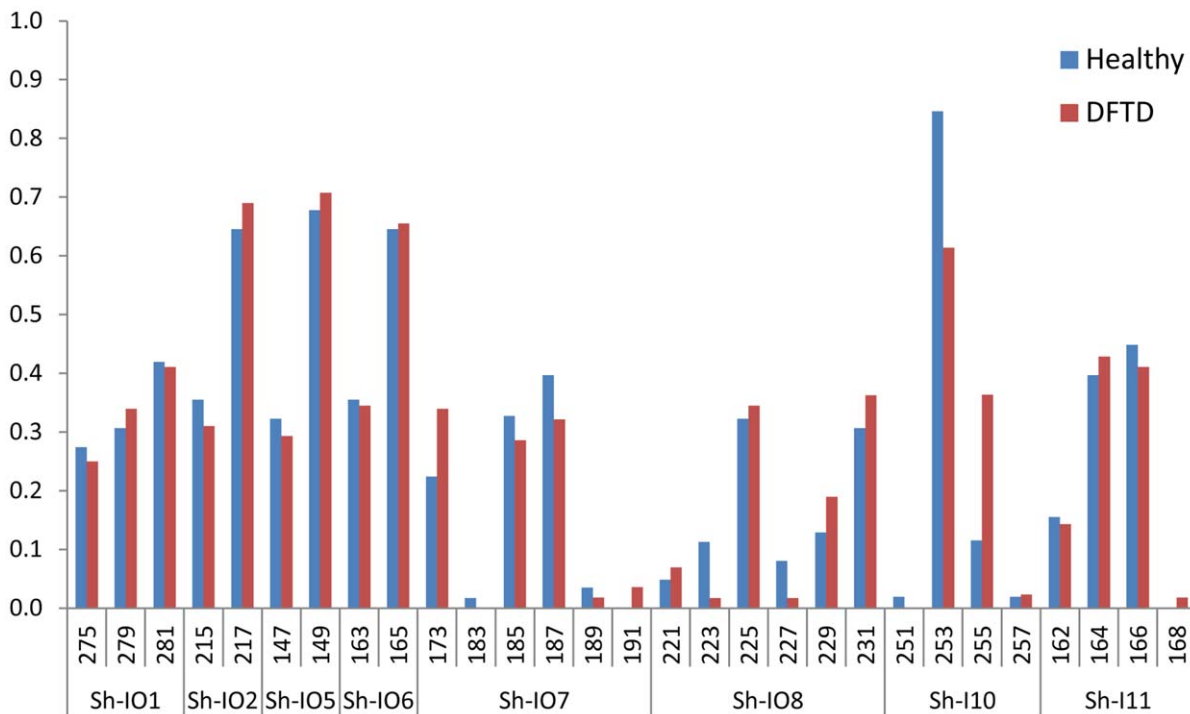doi:10.1371/journal.pone.0036955.t002

Figure 2. MHC-linked microsatellite loci allele frequencies showing little variation between healthy and DFTD infected devils. A single locus (Sh-I07) does not conform to Hardy-Weinberg expectations for healthy devils only (p = 0.003).
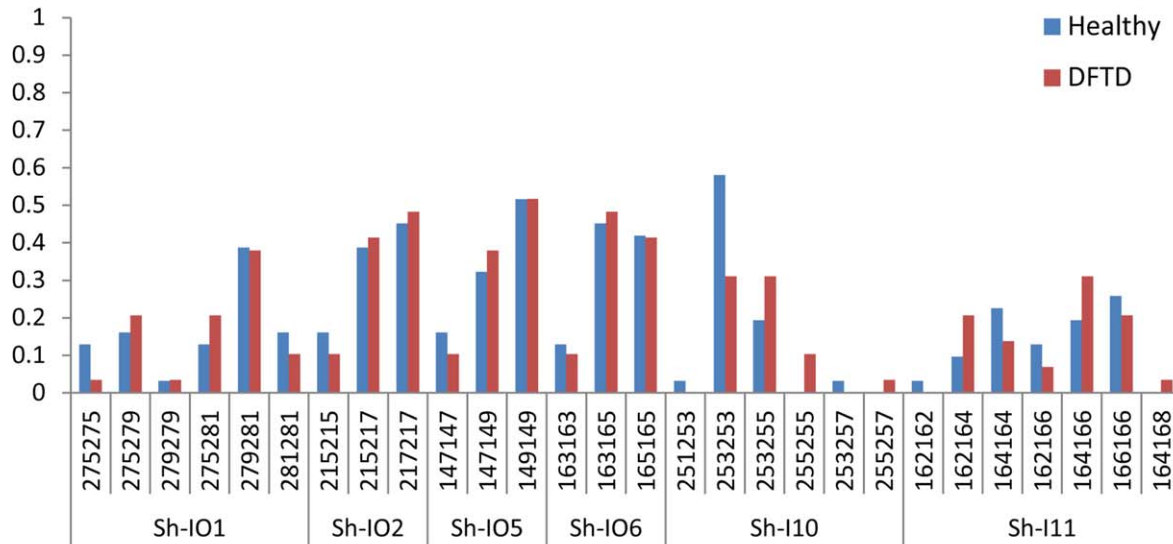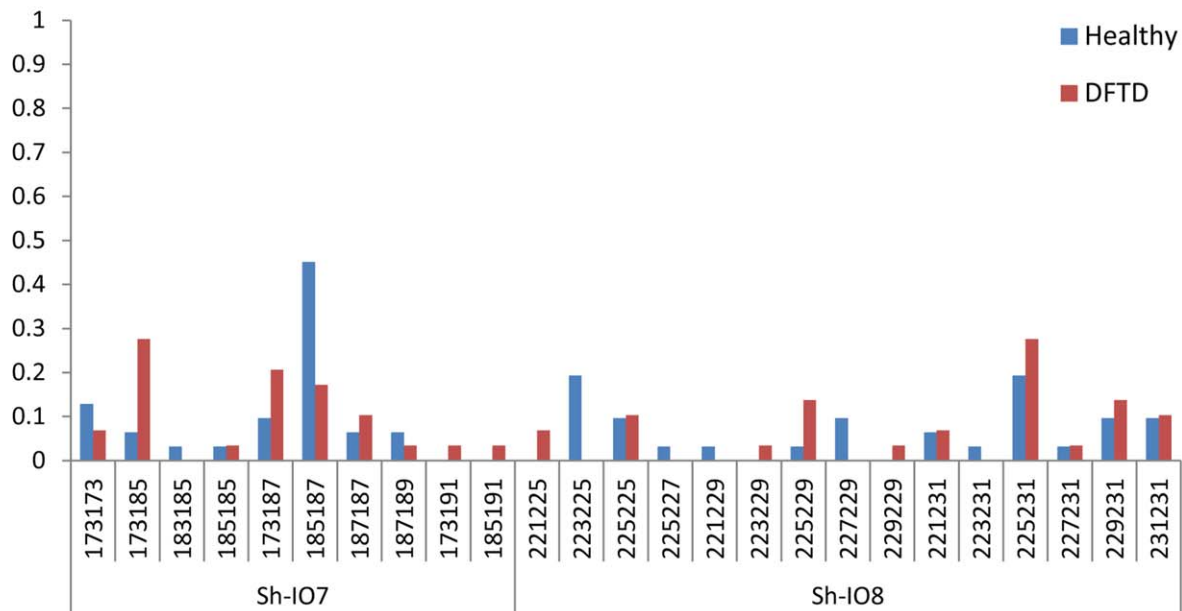doi:10.1371/journal.pone.0036955.g002

Figure 3. Genotype frequencies for healthy and DFTD infected devils. (**A**) At the six microsatellite loci associated with antigen-presenting genes within the MHC (Sh-I01, Sh-I02, Sh-I05, Sh-I06, Sh-I10 and Sh-I11). No deviations from Hardy-Weinberg equilibrium are observed for either healthy or DFTD infected devils. (**B**) For two microsatellite markers more closely associated with non-antigen presenting genes within the MHC region (Sh-I07 and Sh-I08). The Sh-I07 locus is out of Hardy-Weinberg equilibrium for healthy devils only (p = 0.029) and the Sh-I08 locus does not conform to Hardy-Weinberg equilibrium at the 0.1 significance level (p = 0.076). Three differences in genotype frequencies were significant before Bonferroni correction (Sh-I07:173/185, p = 0.041 and 185/187, p = 0.024; Sh-I08:223/225, p = 0.024).
doi:10.1371/journal.pone.0036955.g003

Monte Carlo analysis of genotype distributions showed that the actual distribution of genotypes, when compared across the two subgroups (healthy and diseased), differed significantly from 10,000 randomly generated distributions for one locus (Sh-I07 p = 0.04). Fishers exact test was then used to investigate in more detail which particular genotypes differed in frequency between healthy and diseased devils. Three differences were identified in genotype frequency at two separate loci, which were significant before sequential Bonferroni correction (Sh-I07 genotype 173/185, p = 0.041 and 185/187, p = 0.024; Sh-I08 genotype 223/225, p = 0.024: Fig. 3b), though not afterwards due to the large number of comparisons. These differences warrant further investigation,

and future work will focus on increasing the number of markers in these regions.

As mentioned above, both Sh-I07 and Sh-I08 are located within the MHC, but are physically located closer to non-antigen presenting genes than to antigen presenting genes. Sh-108 is located 14.4 kb upstream of MTCH1 (also referred to as PSAP), an evolutionarily conserved gene important in mitochondrial transport [34]. Interestingly, this gene plays a key role in apoptosis – a form of programmed cell death which is crucial in maintaining health as it eliminates old and unhealthy cells [35–37]. Sh-I07 is found 5.5 kb away from FGD2, a gene primarily responsible for the formation of microskeletal structures [38]. This gene also plays a role in immunity via leukocyte signaling and it is known to be expressed in B lymphocytes, macrophages and dendritic cells [39]. In all cases it is possible that the signals we are detecting are simply due to a hitchhiking effect and that use of additional markers within the region may help pinpoint stronger associations.

Conclusions and Implications for Tasmanian Devil Management

The reduced effects of DFTD at the north-western site of WPP, including low disease prevalence and no population decline [23] are not due to copy number variation of MHC class I 'group 2' (Saha-UD) alleles, as previously proposed [20]. We show that perceived differences at this locus are due to inconsistent primer amplification as all individuals are confirmed to have alleles present. No differences are detected between healthy and diseased devils in either the frequency of MHC-linked microsatellite alleles (indicating no 'resistance/susceptibility' alleles) or in the prevalence of a deletion of locus Saha-UA, signifying no influence of copy number variation at this locus. The results of genotype frequency analysis are equivocal with some evidence that certain genotypes are more frequently encountered in healthy devils. As the significance of differences in individual genotype frequencies disappeared after Bonferroni correction, we cannot rule out the possibility that these differences are due to chance, though they do present these genes as being of interest for further study. Future work should aim to increase sample sizes and the number of markers in the region to determine whether these results are real or the result of chance alone.

The different epidemiology of DFTD in WPP is currently unexplained. The possibility remains that the uninfected devils used in our analysis may have the disease but not yet developed tumours or that they have simply not yet come into contact with the disease. Both of these explanations are plausible but unlikely. The latency state, between infection and the presentation of tumours, is estimated to average 6 months but could be at least 10 months [1, R.H. unpublished data]. Low disease prevalence among adults (13%) observed in WPP at the time of this study, relative to a large devil population (300 individuals trapped between 2006 and 2011), means that the number of 'latent' animals would be low. Older devils greater than three years of age were specifically selected for our analysis as they have had maximum opportunity for exposure to DFTD including social and reproductive interactions which involve biting. By three years of age virtually all devils in eastern Tasmania have succumbed to the disease and the remainder die within months [1,10]. While there is no current evidence that animals at WPP differ from those in the east in behaviours linked to disease transmission, we suggest modeling the disease risks of each population as a research priority.WPP provides our best hopes of determining whether genetic resistance to DFTD exists. Future studies will continue to target the MHC, as well as expanding to look genome-wide, using techniques such as SNPs (single nucleotide polymorphisms) and next generation sequencing to identify the impact that immune gene variants have on disease susceptibility.

Methods

Ethics Statement

The field research was carried out with approval from the University of Tasmania's Animal Ethics Committee (A0010296) and from the Tasmanian Department of Primary Industries and Water (TFA 08211).

Study Site and Sample Collection

West Pencil Pine (41° 31′ S, 145° 46′ E) is a 25 km^2 area situated on private production forestry land to the west of Cradle Mountain in northwest Tasmania. This study site was established in May 2006 (when the disease was first detected in the region), following three exploratory expeditions aimed at locating the epidemic front. We subsequently sampled this population four times per year, from August 2006 until May 2010, at three month intervals. Forty carnivore traps were set for 10 nights each trip. All individuals caught were individually marked using implantable microchip transponders and a 3 mm biopsy of ear tissue was taken using a sterile biopsy punch. Due to the regularity with which this site has been monitored, most of the individuals in our data set were originally captured as sub-adults and are therefore of known age. We aged devils first captured as adults using a combination of molar eruption, molar tooth wear and canine over eruption. This method is considered precise for ageing devils up to three years of age (M.J., unpublished data). Disease status was assessed by histopathological examination of biopsies from tumours [8], or when this was not possible, by visual inspection and identification of tumours [9].

The healthy and diseased devils were trapped and genetically sampled between May 2006 and May 2010. All the healthy, tumour-free devils used in this study were greater than three years old at the time of capture and had co-existed with DFTD infected individuals at least since sexual maturity for males, and for most females, for their entire lives (devils have male-biased natal dispersal) [12].

DNA Extraction and MHC Class I Sequencing

Total genomic DNA was extracted from ear biopsies using the Hot-SHOT extraction method [40]. A fragment that targeted the MHC class I α 1 domains (exon 2), was amplified in polymerase chain reactions (PCR) utilizing primers developed by Siddle et al. [41]. Amplification reactions were 25 μL volume and comprised; 1 x High Fidelity Buffer (Invitrogen) consisting of 60 mM Tris-HCl (pH 8.9) and 18 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.5 μM forward and reverse primers, 1.5 U of Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and approximately 10 ng of template DNA. PCR amplification of double-stranded product was performed with a MJ Mini Personal Thermocycler (Biorad) using a cycling profile consisting of an initial denaturing step of 94C for 3 mins followed by 35 cycles of 94C for 30 s, 60C for 30 s and 72C for 30 s. This was followed by a final extension at 72C for 20 mins. PCR products were checked for quality on a 1% agarose gel and DNA concentration measured using a Nano Photometer (Implen). Bands of approximately 300 bp in size were extracted from agarose gels using an Ultra Clean DNA Purification Kit (MoBio) and cloned into pGEM-T easy vector (Promega) according to the manufacturer's protocol. Colonies with inserts of the predicted size were inoculated into LB broth, grown overnight, and DNA was extracted using a QIAprep

spin miniprep kit (Qiagen). In total 10 clones each from two separate PCRs were sequenced per individual, resulting in a total of 1060 sequences (sequencing performed by Australian Genome Research Facility, Queensland).

Sequences were aligned and quality-checked using Bioedit v7.0.9 [42]. All sequences containing ambiguous nucleotides were removed from analysis. To confirm the validity of new sequence variants and establish that they were not the result of PCR error, new sequences had to occur in a minimum of two separate PCR amplifications (ie., either two separate PCRs of the same individual or PCRs of two different devils). All new sequences were translated into amino acids using the standard genetic code and checked for the presence of stop codons in MEGA v 5 [43]. Novel nucleotide sequences were submitted to Genbank. They were placed in the context of other, previously identified devil MHC sequences via phylogenetic analysis. MEGA v 5 [43] was used to build a neighbor joining tree using Jukes-Cantor distance with 1000 bootstraps, consistent with the methods used to build the phylogeny of marsupial class I sequences [44].

Two PCR tests were performed to confirm the existence of *Saha-UA* and *Saha-UD* in each individual. The copy number variation at *Saha-UA* was detected following the protocol described previously [21]. A new pair of primers (forward 5′- ATGGATA-GAGAAGATGGAGAAT -3′ and reverse 5′- CTGGTTGTAG-TAGCCGTGTA -3′) were designed to specifically amplify a 121 bp segment within the α1 domain of *Saha-UD* at the following conditions: 94°C initial denaturation for 3 min; 32 cycles of 94°C denaturation for 30 s, 56°C annealing for 30 s, 72°C extension for 30 s; and 72°C final extension for 10 min.

Microsatellite Genotyping

Eight molecular markers linked to MHC-loci [30] were used to genotype 29 diseased and 31 healthy Tasmanian devils. Note that sample number includes seven additional diseased animals than were used in sequencing and two additional healthy animals. Other than the inclusion of these extra nine devils, the same individuals were used for both sequencing and microsatellite analyses. Microsatellite loci were amplified in 15 µL reactions using approximately 10 ng of genomic DNA with 0.2 mM dNTPs, 1 x PCR buffer (containing 15 mM MgCl$_2$), additional 2 mM MgCl$_2$, 0.5 mM unlabelled primer, 0.05 mM primer labeled with M13 (−21) tail, 0.5 mM universal fluorescent-labelled M13(−21) primer, 25 ng/µL BSA (Bovine Albumin Serum) and 0.05 U of *Taq* DNA polymerase (Qiagen). Forward and reverse primers were labeled as in Cheng et al [30]. PCR reactions were performed with a MJ Mini Personal Thermocycler (Biorad) under the following conditions: 94°C for 3 mins, 6 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s using a touchdown program with the annealing temperature decreasing by 1°C per cycle for a final temperature of 54°C; 30 cycles of 94°C for 30 s, 54°C for 30 s, 72°C for 30 s and a final extension of 72°C for 10 mins.

Nine neutral microsatellite markers [45] were also used to genotype the same 29 DFTD-infected and 31 healthy animals. Neutral loci were amplified in 15 µL reactions using approximately 10 ng of genomic DNA, 0.2 mM dNTPs, 1 x PCR buffer (containing 15 mM MgCl$_2$), additional 0.5 mM MgCl$_2$, 0.2 mM forward and reverse primers labeled as in Jones et al [45] (fluorescent labels: Applied Biosytems) 0.025 U of *Taq* DNA polymerase (Qiagen). Reactions were performed with a MJ Mini Personal Thermocycler (Biorad) under the following conditions: 94°C for 1.45 mins, 6 cycles of 94°C for 15 s, 60°C for 20 s and 72°C for 15 s using a touchdown program with the annealing temperature decreasing by 1°C per cycle for a final temperature of 54°C; 30 cycles of 94°C for 15 s, 54°C for 20 s, 72°C for 10 s and final extension of 72°C for 2 mins.

The amplified products were separated by electrophoresis on an ABI 3130XL sequencer (Applied Biosystems) and scored against the size marker LIZ 500 using Genemarker v 1.95 (Soft Genetics LLC).

Analysis

The frequency of a deletion at the *Saha-UA* locus was compared between healthy and diseased devils using two-tailed Fisher's Exact Test in SPSS version 18 (SPSS Inc). This same method was used to compare the likelihood of the primers failing to amplify alleles a given locus in either healthy or diseased animals. Cases where only one allele was detected may be the result of a failure to amplify, but conversely the individual may be homozygous. For this reason only when a locus was missing all alleles did we consider this conclusive evidence of non-amplification.

Population genetic (microsatellite) analysis was performed using Arlequin: version 3.11 [46] and FSTAT: version 2.9.3.2 [47]. Hardy-Weinberg equilibrium was calculated in Arlequin via an exact test using a Markov chain (forecasted chain length: 100,000, dememorisation steps: 5000) [48]. Population structure was analysed using Arlequin to calculate F_{ST} values (distance method, 10,100 permutations) as well as F_{IS} for each category- diseased and healthy. FSTAT was used to calculate linkage disequilibrium. Allele and genotype frequencies were calculated using GenAlEx [49] and compared between subgroups using two-tailed Fishers Exact Test in SPSS version 18 (SPSS Inc). Allelic diversity and observed and expected heterozygosities were calculated with Arlequin and results were corrected by sequential Bonferroni analysis [50]. A Monte Carlo test of randomization was used to assess whether the observed distribution of genotypes at each locus differed significantly from 10,000 randomly generated distributions. Compared to parametric tests, Monte Carlo procedures are more robust to small and unbalanced datasets [51]. The program PopTools v 3.2 [52] provided the shuffle algorithm for generating random distributions as well as the difference statistic used for the Monte Carlo simulation. P values were calculated according to the number of times the actual distribution of genotypes across the two disease statuses (diseased and non-diseased) fell outside the confidence limits (lower percentile = 0.025, upper percentile = 0.975) of 10,000 randomly generated distributions.

Supporting Information

Table S1 Amino acid sequences of all Tasmanian devil MHC class I alleles amplified in this study. The consensus sequence is SahaI*27 and all polymorphic residues are shown. (DOCX)

Author Contributions

Conceived and designed the experiments: KB. Performed the experiments: AL BW YC LL. Analyzed the data: AL YC BW BU. Contributed reagents/materials/analysis tools: RH MJ. Wrote the paper: AL KB. Read and approved drafts of the manuscripts: AL YC BW RH LL MJ BU KB.

References

1. McCallum H, Jones M, Hawkins C, Hamede R, Lachish S, et al. (2009) Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology 90: 3379–3392.

2. The IUCN Red List of Threatened Species. Available: http://www.iucnredlist.org/apps/redlist/details/40540/0. Accessed 2012 May, 15.

3. Pearse A-M, Swift K (2006) Transmission of devil facial-tumour disease. Nature 439: 549.

4. Pyecroft SB, Pearse A-M, Loh R, Swift K, Belov K, et al. (2007) Towards a case definition for Devil Facial Tumour Disease: What is it? EcoHealth 4: 346–351.

5. Siddle HV, Kriess A, Eldridge MD, Noonan E, Clarke CJ, et al. (2007a) Transmission of a fatal clonal tumour by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proc Natl Acad Sci 104: 16221–16226.

6. Pemberton D, Renouf D (1993) A field study of communication and social behaviour of the Tasmanlan devil at feeding sites. Aust J Zool 41: 507–526.

7. Hamede RK, McCallum H, Jones M (2008) Seasonal, demographic and density-related patterns of contact between Tasmanian devils (Sarcophilus harrisii): Implications for transmission of devil facial tumour disease. Austral Ecol 33: 614–622.

8. Loh R, Bergfeld J, Hayes D, O'Hara A, Pyecroft S, et al. (2006) The pathology of Devil Facial Tumour Disease (DFTD) in Tasmanian devils (Sarcophilus harrisii). Vet Pathol 43: 890–895.

9. Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, et al. (2006) Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. Biol Conserv 131: 307–324.

10. Lachish S, Jones M, McCallum H (2007) The impact of disease on the survival and population growth rate of the Tasmanian devil. J Anim Ecol 76: 926–936.

11. Lachish S, McCallum H, Jones M (2009) Demography, disease and the devil: life-history changes in a disease-affected population of Tasmanian devils (Sarcophilus harrisii). J Anim Ecol 78: 427–436.

12. Lachish S, Miller KJ, Storfer A, Goldizen AW, Jones M, et al. (2011) Evidence that disease-induced population decline changes genetic structure and alters dispersal patterns in the Tasmanian devil. Heredity 106: 172–182.

13. Jones M, Cockburn A, Hamede RK, Hawkins CE, Hesterman H, et al. (2008) Life-history change in disease-ravaged Tasmanian devil populations. Proc Natl Acad Sci 105: 1–5.

14. Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, et al. (2010) The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science 327: 84–87.

15. Jones ME, Paetkau D, Geffen E, Moritz C (2004) Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. Mol Ecol 13: 2197–2209.

16. Miller W, Hayes VM, Ratan A, Petersen, DC, Wittekindt NE, et al. (2011) Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). Proc Natl Acad Sci 108: 12348–53.

17. Woods GM, Kriess A, Belov K, Siddle HV, Obendorf DL, et al. (2007) The immune response of the Tasmanian devil (Sarcophilus harrisii) and Devil Facial Tumour Disease. Ecohealth 4: 338–345.

18. Belov K (2011) The role of the Major Histocompatibility Complex in the spread of contagious cancers. Mamm Genome 22: 83–90.

19. Kreiss A, Cheng Y, Kimble F, Wells B, Donovan S, et al. (2011) Allorecognition in the Tasmanian devil (Sarcophilus harrisii), an endangered marsupial species with limited genetic diversity. PLoS one 6: e22402.

20. Siddle HV, Marzec J, Cheng Y, Jones M, Belov K (2010) MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. P Roy Soc B-Biol Sci 277: 2001–6.

21. Cheng Y, Stuart A, Morris K, Taylor R, Siddle H, et al. (2012) BAC-contig-based characterization of MHC Class I and Class II regions in the Tasmanian devil. BMC Genomics. 13: 87.

22. Jones M, Barmuta L (2000) Niche differentiation among sympatric Australian dasyurid carnivores. J Mammal 81: 434–447.

23. Hamede R, Lachish S, Belov K, Woods G, Kriess A, et al. (2012) Reduced impact of Tasmanian Devil Facial Tumour Disease at the current disease front. Conserv Biol. 26: 124–134.

24. Hamede RK, Bashford J, McCallum H, Jones M (2009) Contact networks in a wild Tasmanian devil (Sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of a devil facial tumour disease. Ecol Lett 12: 1147–1157.

25. Schwaiger FW, Buitkamp J, Weyers E, Epplen JT (1993) Typing of artiodactyl Mhc-Drb genes with the help of intronic simple repeated Dna-sequences. Mol Ecol 2: 55–59.

26. Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, et al. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci 101: 3490–3494.

27. Doxiadis GGM, de Groot N, Claas FHJ, Doxiadis IIN, van Rood JJ, et al. (2007) A highly divergent microsatellite facilitating fast and accurate DRB haplotyping in humans and rhesus macaques. Proc Natl Acad Sci 104: 8907–8912.

28. Prasad VK, Yang SY (1996) Allele assignment for HLA-A, -B, and –C genes to the Tenth International Histocompatibility Workshop cell lines. Tissue Antigens 47: 538–546.

29. Santamaria P, Lindstrom AL, Boyce-Jacino MT, Myster SH, Barbosa JJ, et al. (1993) HLA class I sequence-based typing. Hum Immunol 37: 39–50.

30. Cheng Y, Belov K (2011) Isolation and characterisation of 11 MHC-linked microsatellite loci in the Tasmanian devil (Sarcophilus harrisii). Conserv Genet Res DOI 10.1007/s12686–011–9575–4.

31. Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex (MHC), Heredity 96: 7–21.

32. Paddle R (2012) The thylacine's last straw: epidemic disease in a recent mammalian extinction. Australian Zoologist 35: to be published March/April 2012.

33. Cheng Y, Sanderson C, Jones M, Belov K (2012) Low MHC Class II diversity in the Tasmanian devil (Sarcophilus harrisii). Immunogenetics, in press (PMID. 22460528).

34. Lai CH, Chou CY, Ch'ang LY, Liu CS, Lin WC (2000) Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. Genome Res 10: 703–713.

35. Lamarca V, Sanz-Clemente A, Perez-Pe R, Martinez-Lorenzo MJ, Halaihel N, et al. (2007) Two isoforms of PSAP/MTCH1 share two proapoptotic domains and multiple internal signals for import into the mitochondrial outer membrane. Am J Physiol Cell Physiol 293: C1347–C1361.

36. Xu X, Shi YC, Gao W, Mao G, Zhao G, et al. (2002) The novel presenilin-1-associated protein is a proapoptotic mitochondrial protein. J Biol Chem 277: 48913–22.

37. Meijer D, Jansen MPHM, Look MP, Ruigrok-Ritstier K, van Staveren IL, et al. (2009) TSC22D1 and PSAP predict clinical outcome of tamoxifen treatment in patients with recurrent breast cancer. Breast Cancer Res Treat 113: 253–260.

38. Pasteris NG, Gorski JL (1999) Isolation, characterization, and mapping of the mouse and human Fgd2 genes, faciogenital dysplasia (Fgd1: Aaeskog syndrome) gene homologues. Genomics. pp 57–66.

39. Huber C, Martensson A, Bokoch GM, Nemazee D, Gavin AL (2008) FGD2, a CDC42-specific exchange factor expressed y antigen-presenting cells, localizes to early endosomes and active membrane ruffles. J Biol Chem 283: 34002–34012.

40. Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, et al. (2000) Preparation of the PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). Biotechniques 29: 52–54.

41. Siddle HV, Sanderson C, Belov K (2007b) Characterization of major histocompatibility complex Class I and Class II genes from the Tasmanian devil (Sarcophilus harrisii). Immunogenetics 59: 753–760.

42. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acid S 41: 95–98.

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

44. Siddle HV, Deakin JE, Coggill P, Hart E, Cheng Y, et al. (2009) MHC-linked and un-linked class I genes in the wallaby. BMC Genomics 10: 310.

45. Jones ME, Paetkau D, Geffen E, Moritz C (2003) Microsatellites for the Tasmanian devil (Sarcophilus laniarius). Mol Ecol Notes 3: 277–279.

46. Excoffier L, Laval G, Schneider S (2005) Arelequin (version 3.0): An integrated software package for population genetic data analysis. Evol Bioinform Online 1: 47–50.

47. Goudet F (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. J Hered 86: 485–486.

48. Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361–372.

49. Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288–295.

50. Rice WR (1989) Analyzing tables of statistical tests. Evolution 43: 223–225.

51. Manly BFJ (1997) Randomization, Bootstrap and Monte Carlo Methods in Biology, 2nd edition. Taylor and Francis Inc., London.

52. Hood GM PopTools version 3.2. Available: http://www.poptools.org/.