MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 3860–3872 doi: 10.1111/mec.13291

Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*)

KATRINA M. MORRIS,* BELINDA WRIGHT,* CATHERINE E. GRUEBER,*† CAROLYN HOGG‡ and KATHERINE BELOV*

*University of Sydney, Faculty of Veterinary Science, Sydney, NSW 2006, Australia, †San Diego Zoo Global, San Diego, CA, USA, ‡Zoo and Aquarium Association, Mosman, NSW 2088, Australia

Abstract

The Tasmanian devil (*Sarcophilus harrisii*) is threatened with extinction due to the spread of devil facial tumour disease. Polymorphisms in immune genes can provide adaptive potential to resist diseases. Previous studies in diversity at immune loci in wild species have almost exclusively focused on genes of the major histocompatibility complex (MHC); however, these genes only account for a fraction of immune gene diversity. Devils lack diversity at functionally important immunity loci, including MHC and Toll-like receptor genes. Whether there are polymorphisms at devil immune genes outside these two families is unknown. Here, we identify polymorphisms in a wide range of key immune genes, and develop assays to type single nucleotide polymorphisms (SNPs) within a subset of these genes. A total of 167 immune genes were examined, including cytokines, chemokines and natural killer cell receptors. Using genome-level data from ten devils, SNPs within coding regions, introns and 10 kb flanking genes of interest were identified. We found low polymorphism across 167 immune genes examined bioinformatically using whole-genome data. From this data, we developed long amplicon assays to target nine genes. These amplicons were sequenced in 29–220 devils and found to contain 78 SNPs, including eight SNPS within exons. Despite the extreme paucity of genetic diversity within these genes, signatures of balancing selection were exhibited by one chemokine gene, suggesting that remaining diversity may hold adaptive potential. The low functional diversity may leave devils highly vulnerable to infectious disease, and therefore, monitoring and preserving remaining diversity will be critical for the long-term management of this species. Examining genetic variation in diverse immune genes should be a priority for threatened wildlife species. This study can act as a model for broad-scale immunogenetic diversity analysis in threatened species.

Keywords: diversity, immune, immunogenetics, *Sarcophilus harrisii*, Tasmanian devil

Received 23 March 2015; revision received 18 June 2015; accepted 24 June 2015

Introduction

Genetic diversity at immune loci is important for population-level resistance against infectious diseases. Immune genes include some of the most polymorphic genes in the genome (Hughes & Hughes 1995; Meyer & Thomson 2001), and diversity at many of these genes is thought to be maintained by pathogen-mediated balancing selection (Doherty & Zinkernagel 1975; Jeffery & Bangham 2000; Piertney & Oliver 2006; Fumagalli *et al*. 2009). Studies have demonstrated that the rate of adaptive evolution is higher at many immune genes compared with other classes of genes, in several vertebrate species (Huang

Correspondence: Katherine Belov, Fax: +61 2 93513957; E-mail: kathy.belov@sydney.edu.au

et al. 2004; Tonteri *et al.* 2010). Polymorphisms in immune genes have been shown to give resistance to diverse diseases in a range of species, both laboratory and wild (Else & Wakelin 1988; Paterson *et al.* 1998; Shiina *et al.* 2004; Piertney & Oliver 2006; Turner *et al.* 2011), and therefore may provide adaptive potential in wild populations.

The vast majority of studies that have examined diversity of immune genes in wild populations have focused on a single family: the major histocompatibility complex (MHC) genes, which form a key component of the adaptive immune system. There is a large body of research investigating MHC diversity, its implications for population fitness and association of these genes with diseases in wild populations (Reviewed in Sommer 2005; Ujvari & Belov 2011). To a lesser extent, diversity of Toll-like receptors (TLR), key receptors of the innate immune system, has been investigated in a growing number of wild species (Grueber *et al.* 2012; Abrantes *et al.* 2013; Tschirren *et al.* 2013). MHC, however, only accounts for a fraction of the genetic variability underlying resistance to infectious pathogens (Jepson *et al.* 1997; Acevedo-Whitehouse & Cunningham 2006). Studies that examine genetic diversity at immune genes outside these two families in wild populations are scarce. Additionally, studies of immune gene diversity in wild populations have been limited in scale, typically only examining one or a few genes. Several recent studies have examined a wider suite of genes such as Bollmer *et al.* (2011) which examined five non-MHC genes, and Turner *et al.* (2011, 2012) which studied diversity in 12 immune genes. However, to capture the genetic diversity of the immune system as a whole and understand its role in selection, a broader approach is required.

The Tasmanian devil (*Sarcophilus harrisii*) is being threatened with extinction due to the spread of the devil facial tumour disease (DFTD). This disease has wiped out over 85% of the devil population since it first emerged in 1996 (DPIPWE 2011). DFTD is highly unusual as it is a transmissible cancer where the tumour cells themselves are the infectious agent, transmitted from one devil to another through biting (Pearse & Swift 2006). The cancer cell is able to pass from devil to devil without triggering an immune response as it is able to downregulate cell surface MHC, and in such a way slip under the radar of the immune system (Siddle *et al.* 2013). However, these observations do not fully account for the ability of DFTD to avoid immune recognition, as a lack of MHC expression should still elicit a natural killer (NK) cell response (Whiteside & Herberman 1995). It has therefore been hypothesized that the lack of genetic diversity seen in the devil population may contribute to the tumour's ability to spread. Devils lack genetic diversity genome wide, as has been demonstrated in studies using microsatellites (Jones *et al.* 2004), single nucleotide polymorphisms (SNPs) and mtDNA (Miller *et al.* 2011). In addition, devils have critically low diversity at MHC loci (Siddle *et al.* 2007a,b; Cheng *et al.* 2012) and at TLR genes (Cui *et al.* 2015). However, it is not known whether the devil genome has additional diversity at immune genes outside these two gene families, particularly genes associated with the NK cell response.

In 2006, the Tasmanian devil insurance population was instigated by the Save the Tasmanian Devil Program, an initiative of the Tasmanian and Australian governments. Over 100 devils were recruited into the existing zoo-based breeding programme through three intakes between 2005 and 2008 (Hogg *et al.* in press). This programme aims to maintain 95% of the genetic diversity of the founders of the captive programme (CBSG 2008), so that captive devils can be used to repopulate Tasmania should extinction occur, or to repopulate locally extinct populations of devils. Achieving this goal will require maintenance of genetic diversity over many generations. As the wild devil population already has critically low genetic diversity, it is crucial that further losses, in particular diversity at functional loci, do not occur. To ensure maintenance of this diversity, genetic testing is required.

In this study, we examine SNP diversity at a broad range of immune genes including cytokines, chemokines, NK cell receptors, genes in the extended natural killer cell receptor complex (NKC) and other innate and adaptive immune genes found to be polymorphic in wild populations of other species, with 167 genes being examined in total. Of particular interest are SNPs that result in nonsynonymous changes as these are highly likely to have a functional effect, although synonymous SNPs, SNPs in introns and untranslated regions (UTRs), and SNPs up to 10 kb upstream/downstream were also investigated as these potentially can affect the regulation or expression of the gene, and thus also have a functional role (e.g. Bi *et al.* 2007; Pacheco *et al.* 2008; Liu *et al.* 2013). We then developed assays to target immune genes with nonsynonymous SNPs, in up to 220 samples of Tasmanian devils, to quantify levels of diversity. Genes amplified by these assays were examined using population-level diversity statistics and examined for signatures of selection.

Materials and methods

Identification of putative SNPs

A list of immune genes of interest was compiled from a review of literature to identify families and individual genes previously shown to have polymorphisms in wild populations. This included NK receptors (i.e. Futas &

Horin 2013), cytokines (i.e. Turner *et al.* 2012), chemokines (i.e. Gao *et al.* 2013) and a range of immune genes not belonging to one of these families (i.e. Xue *et al.* 2008; Wan *et al.* 2013; Meyer *et al.* 2014; Table S2 (Supporting information) for full details of all genes included). To identify SNPs, contigs containing genes of interest from the two devil genome sequence projects, which sequenced three devil genomes in total (DEVIL7.0, GenBank: GCA_000189315.1 (Murchison *et al.* 2012); GenBank: GCA_000219685.1 (Miller *et al.* 2011)), and seven resequenced devil genomes (European Nucleotide Archive: ERS682204-ERS682210) were examined in BIOEDIT. Genes on contigs <10 000 bp were not included in the analysis as the resequencing reads in these regions typically had low-quality scores. Single nucleotide sites that varied among the ten sequences were identified in BIOEDIT using the positional nucleotide summary tool. Positions containing SNPs were compared to previous devil immune gene predictions (K. M. Morris, Y. Cheng, W. Warren, A. T. Papenfuss, K. Belov, Submitted; van der Kraan *et al.* 2013) to determine whether SNPs were exonic (both synonymous and nonsynonymous), intronic, in a predicted UTR or within 10 kb upstream or downstream of the genes of interest. These previous gene predictions were based on a combination of genomic and transcriptomic data to accurately predict gene structure, although UTRs were not predicted for all genes where transcriptomic data were not available. These gene sequences can be found on the Immunome Database for Marsupials and Monotremes (http://hp580.angis.org.au/tagbase/gutentag/).

Samples

Samples were collected under University of Sydney animal ethics permit 5584. Samples (either ear biopsies or 2 mL blood) were obtained from 221 devils from the captive devil breeding programme from a range of Australian zoos and wildlife parks. These samples were collected as part of ongoing population management of the insurance population and were kindly shared with us by the Save the Tasmanian devil programme. The majority of these samples were from putatively unrelated animals obtained from the wild. These samples were presumed to be unrelated due to the distance between trap locations (Hogg *et al.* in press). Some related animals from the breeding programme were also included. DNA extractions were performed with the QIAGEN DNeasy Blood and Tissue kit following kit instructions.

Amplicon assay

Nine genes were selected as targets for long amplicon amplification based on the overall number of SNPs and the presence of nonsynonymous SNPs, targeting genes from a range of families. The targeted set of genes included members of the interleukin, chemokine and NK receptor families. The amplicons were designed to be between 2500 and 10 000 bp in length to maximize SNP coverage and encompass the entire coding sequence where possible. Primers were designed using the OLIGO6 software (Table 1). Between 29 and 220 devil samples were amplified with each primer set. Not all the samples used in this project were amplified with all the primer sets. Additional samples were amplified with some of the primer sets due to their use in concurrent projects, thus the number of samples per marker varied (Table S1, Supporting information). PCR amplification was performed using the Invitrogen SequalPrep Long PCR Kit. PCRs in a total volume of 20 μL contained 1x SequalPrep 10X Reaction Buffer, 0.4 μL DMSO, 1.8 U SequalPrep Long Polymerase, 0.5 μM each of the reverse and forward primer, *c.*50 ng of DNA and 0-1 X SequalPrep 10X Enhancer A and B (determined by prior primer optimization following kit instructions; Table 1). Amplification conditions were as follows. Reactions were heated at 90 °C followed by 10 cycles of 94 °C for 10 s, 55 °C for 30 s and 68 °C for 12 min. This was followed by 30 cycles of 94 °C for 10 s, 60 °C for 30 s and 68 °C for 12 min with the addition of 20 s per cycle. This was followed by a final extension step of 72 °C for 5 min. Amplification of the target products was confirmed by running 3.5 μL PCR product on a 1.5% agarose gel for 1 h at 100 V. Following this, the PCR products were simultaneously purified and normalized using Invitrogen SequalPrep Normalization Plate Kit following the kit guidelines. PCR products from all amplicons were pooled in equal concentrations by individuals, and sequencing libraries were then prepared using the Illumina Nextera DNA Sample Prep Kit following the kit guidelines in conjunction with the Nextera XT index kit (Illumina) allowing us to multiplex amplicons from 96 individuals in a single run. The pooled library samples were sent to the Ramaciotti Centre for Genomics for sequencing of 150-bp paired-end reads on an Illumina MiSeq. In total, data from this project were obtained via a total of four 96-sample sequencing runs, in which samples for multiple ongoing projects were combined. The average coverage per sample was *c.* 200× coverage.

Data analysis

Multiplexed samples were separated and aligned to the devil reference genome using Burrows-Wheeler Aligner (Li & Durbin 2009). Samtools (Li *et al.* 2009) was used to call SNPs with minimum base and mapping quality scores of at least 20. PLINK v1.07 (Purcell *et al.* 2007) was

Table 1 Primers (5′ to 3′) for amplicon markers; also provided are the concentrations of 'enhancer' reagents used as part of the SequalPrep Long Range PCR kit (Life Technologies)

Marker	Forward primer	Reverse primer	Enhancer A	Enhancer B	Product Length (bp)
CCLD5/6	GCAAAGTAGAAGAGATAACAGGGATAAG	CTATAAAAAGAGGGTTCAAGCATGAG	1×	1×	8209
CX3C	GGGGAATGAAAAATGTGTATG	GGGGAATGACTCTGAAGAAC	1×	1×	9425
TGFB1	CACAGACAACACAAAACACAC	GTGTACCAGTGGCTGAGTAG	0.5×	1×	2649
IL-22F1	ATATTGCTGACAGTGAACAACATC	CAGCATACTGACCAAAAAAAGAGTC	0.5×	0×	6531
IL-22	TTCACTTTGCTCCTCATATC	AATTGTCTGGACTTGAGAGAG	1×	1×	5275
IL-17B	GCTTGGAATTTAGAGTGTGTC	ACCACAAAGAGCTACAGATTAG	1×	1×	5717
DIG12	TTGCTAAAGGGTCTCGTGATCTGT	CGTTTCCTATGGATTAGTGATGACTC	1×	1×	6351
DIG24	GATAAACTCTAACTGCGTGTC	AACCCAGATATAAGTGCTAGAC	1×	1×	6809

used to identify SNPs with a minor allele frequency (MAF) < 0.03, which were removed from further analyses. PLINK was also used to determine heterozygosity and deviation from Hardy–Weinberg equilibrium (HWE). DNASP v5 (Librado & Rozas 2009) was used to infer haplotypes, calculate haplotype number (*h*), haplotype diversity (*hd*), average number of nucleotide differences (*k*) and nucleotide diversity per site (π). Tajima's D statistical test (Tajima 1989) was used to test the neutrality of the polymorphisms. The ratio of synonymous to nonsynonymous variation was examined using the codon-based *Z*-test of selection in MEGA6 (Tamura *et al.* 2013).

Results

SNP identification in devil immune genes

In total, 212 devil genes were identified as immune genes of interest likely to contain SNPs. Of these, 45 were not investigated further due to poor resequencing quality or missing data. The remaining 167 genes were examined in ten devil genomes for putative SNPs (Table 2, Table S1, Supporting information). The entire length of the genes including 10 kb upstream and downstream was examined for SNPs. A total of 5.3 MB of sequence was examined for SNPs, with a total of 3471 SNPs being identified within these genes or within 10 kb of these genes. All genes examined had at least one SNP either within the gene or within 10 kb of the gene. Of the genes examined, 92 had predicted UTRs and of these 28 had SNPs within the UTR. 101 genes had SNPs within introns totalling 1089 intronic SNPs. Forty-seven genes had SNPs within the exons and of these 39 genes had nonsynonymous changes. In total, 69 nonsynonymous SNPs were identified in this subset of genes. The number of nonsynonymous SNPs per gene ranged from one to nine. Overall, the 167 genes examined contained an average of 6.5 intronic SNPs, 0.35 nonsynonymous and 0.40 synonymous SNPs (Table 2).

The distribution of genes with nonsynonymous SNPs was not uniform between different families of immune genes (Table 2). Cytokines are a diverse group of secreted proteins produced by a broad range of cells that act as mediators of the immune system. Cytokine families include interleukins, chemokines, interferons and the TNF family. Within the cytokines, four of 41 interleukins, five of 37 chemokines, three of 17 TNF family members and three of 11 additional cytokines investigated contained nonsynonymous SNPs. The interferon family (six genes) had no SNPs within exons or introns. Interleukins that contained nonsynonymous SNPs were *IL-16*, *IL-17B*, *IL-1A*, *IL-22* and *IL-22F1*. Within the cytokines, the chemokine family showed the

Table 2 Summary of SNPs identified from resequencing data

Gene family	Number of genes	Length examined kb (Total (Mean))	No. of SNPs (Total (Mean))				
			Within 10 kb	UTR	Intron	Exon (s)	Exon (ns)
Interleukin	41	1396 (34)	439 (10.7)	4 (0.2)	221 (5.4)	9 (0.2)	12 (0.3)
Chemokine	37	887 (24)	700 (18.9)	23 (0.9)	58 (7.3)	6 (0.4)	5 (0.4)
Interferon	6	106 (18)	71 (11.8)	0 (0)	0 (0)	0 (0)	0 (0)
TNF	17	561 (33)	173 (10.2)	13 (1.3)	78 (4.6)	0 (0)	3 (0.2)
Additional cytokines	11	560 (51)	101 (9.2)	4 (0.5)	144 (13.1)	1 (0.1)	3 (0.3)
NK receptors (NKC)	6	190 (32)	18 (3)	2 (2)	11 (1.8)	0 (0)	0 (0)
NK receptors (LRC)	17	357 (21)	269 (15.8)	1 (0.2)	106 (6.2)	17 (1)	6 (0.4)
Extended LRC receptors	16	474 (30)	287 (17.9)	19 (3.2)	375 (23.4)	24 (1.5)	34 (2.1)
Additional immune genes	16	801 (50)	130 (8.1)	3 (0.3)	93 (5.8)	2 (0.1)	6 (0.4)
Total	167	5333 (32)	2188 (13.1)	69 (0.8)	1086 (6.5)	59 (0.35)	69 (0.4)

greatest SNP diversity with an average of 7.3 intronic and 0.8 exonic SNPs. Chemokines containing nonsynonymous SNPs were found in the CCL, CXCL and the CX3C families, though not the XCL family. The single CX3C family member had a single putative nonsynonymous SNP. Three CCL family genes contained nonsynonymous SNPs, *CCL26* and two devil-specific CCL genes, *CCLD6* and *CCLD7*. The latter two genes form part of a devil-specific chemokine gene expansion (K. M. Morris, Y. Cheng, W. Warren, A. T. Papenfuss, K. Belov, Submitted). A single CXCL family member, *CXCL9*, contained a nonsynonymous SNP. Within the TNF family, three members had nonsynonymous SNPs: *FASLG*, *TNFSF9* and *TNFSF10L*. A large number of SNPs (23) were also identified in the intron of the *FASLG* gene. Additional cytokines with putative coding variants were *LIF*, *TGFB1* and *TGFB3* (Table S2, Supporting information).

NK cells are critical components of the immune system, and their ability to recognize pathogenic or abnormal cells is mediated through cell surface NK receptors. These NK receptors are encoded by two families of genes: those of the leucocyte receptor complex (LRC) and the killer cell immunoglobulin-like receptors (KIRs). None of the six devil LRC NK receptors contained SNPs within the exons and only three had SNPs within introns. Six of the 17 genes investigated in the KIR family had putative nonsynonymous SNPs. A large number of genes with nonsynonymous SNPs were also seen in the extended LRC. These genes are linked to LRC NK receptors in the mammalian genome and have a wide variety of roles. Those with nonsynonymous SNPs include *MAG*, six *SIGLEC* genes and five *CEACAM* genes. In addition, genes within this group often had a large number of nonsynonymous SNPs, up to nine in one *CEACAM* gene.

Additional genes of the innate immune system which have been found to be polymorphic in other wild populations (Xue *et al.* 2008; Wan *et al.* 2013; Meyer *et al.* 2014) were investigated in the devil. Of these 15 genes, five were found to have nonsynonymous SNPs in the devil: *FLT3LG*, *CD209*, *CD11b* (*ITGAM*), *NOS2* (*iNOS*) and *LGP2*.

Amplicon SNPs

Nine genes were selected to be studied in greater depth as targets of an amplicon assay. These genes were selected from the 39 devil genes identified that contain putative nonsynonymous SNPs. Genes for this assay were selected to include a range of key immune gene families including chemokines, interleukins and NK receptors.

Nine genes were included in eight amplicons as one amplicon (CCLD5/6) amplified the region containing both *CCLD5* and *CCLD6*. In total, these eight amplicons cover 50 966 bp. These amplicons were sequenced in 29 to 220 devils per amplicon. Within these, 188 SNPs were predicted based on sequencing data in ten devils, but only 78 SNPs were confirmed experimentally (Table 3). This included 34 SNPs within gene introns and 8 SNPs within exons, including 4 nonsynonymous SNPs. The number of SNPs per marker varied from 1 to 22. The four nonsynonymous SNPs were found in *IL-17B*, *CCLD5* and *CX3C*. Although the overall number of SNPs within the amplicons was less than predicted from resequencing data, this varied across the markers; five markers had fewer SNPs confirmed in the amplicon sequencing, while two had a greater number of SNPs identified in the amplicon sequencing. In total, 30 new SNPs were discovered from the amplicon sequencing.

Conformation to Hardy–Weinberg and diversity statistics

Markers were examined for conformation to HWE (Table 4). Regions with a large number of SNPs that

Table 3 Amplicon SNP summary statistics (detailed SNP information is provided in Table 4)

Gene	Class	*N*.	Length	P SNPs	C SNPs	P non-synonymous SNPs	C non-synonymous SNPs
CCLD5/6	Chemokine	69	8209	29	16	1	1
CX3C	Chemokine	167	9425	6	3	1	1
TGFB1	Cytokine	93	2649	9	2	1	0
IL-22F1	Interleukin	29	6531	10	2	1	0
IL-22	Interleukin	196	5275	97	18	6	0
IL-17B	Interleukin	220	5717	17	20	2	2
DIG12	NK receptor	186	6351	15	16	1	0
DIG24	NK receptor	83	6809	5	1	1	0
Total			50 966	188	78	14	4

N, number of animals genotyped, P, Predicted from resequencing data, C, confirmed in amplicon assays.

deviate from HWE may indicate that selection is acting at that location. Of the 94 SNP markers identified in the eight amplicons, 30 SNPs were found to have significant deviation from HWE ($P < 0.05$). Of the four non-synonymous SNPs in these amplicons, the SNP in *CCLD5* and the two in the *IL-17B* amplicon did not deviate from HWE, but that in the *CX3C* amplicon showed highly significant deviation from HWE ($P = 2.2 \times 10^{-8}$). Those amplicons with the largest number of SNPs deviating from HWE were *IL-22* in which 16 of the 18 SNPs deviated from HWE, mostly (75%) with a heterozygote excess, and *CCLD5/6* in which 10 of 16 deviated from HWE, with the majority (80%) showing a heterozygote deficiency. The SNPs showing a heterozygote excess were concentrated at one end of the IL-22 amplicon (Table 4). Diversity statistics and tests of selection are shown in Table 5. *CCLD6* showed deviation from neutrality using the Tajima's D test, with a highly significant positive Tajima D value, which may arise from either a decreasing population size or balancing selection. Significant deviation from neutrality was not, however, detected with the Z-test. *CCLD6* also had the highest nucleotide diversity and highest mean nucleotide diversity of any of the genes. *IL-17B* and *CCLD5* also showed relatively high diversity, while *TGFB1*, *IL-22*, *IL-22F1* and *DIG24* had no SNPs within their coding sequence. No locus apart from *CCLD6* showed significant deviation from neutrality using Tajima's D or the *Z*-test.

Discussion

SNP diversity across devil immune genes

In total, 167 genes of the devil immune system were examined for the presence of SNPs in ten devil genomes. The entire coding region and 10 kb flanking these genes were searched for SNPs. A total of 5.3 MB of sequence was examined and a total of 3471 SNPs were identified. Nonsynonymous SNPs were of particular interest due to their effect on the protein sequence. Overall, *c*. 23% of the genes examined had at least one nonsynonymous SNP, but this varied greatly between families.

The interferon family showed no SNPs within exons or introns. SNPs within interferons do occur in other species and have been linked to infectious disease, indicating the potential adaptive role of these variants. For example, a SNP in the intron of *IFNG* has been associated with tuberculosis susceptibility in humans (Pacheco *et al.* 2008; Vallinoto *et al.* 2010). The lack of variation in devils, combined with the lower number of type I interferons present in the devil genome compared to both marsupials and eutherians (K. M. Morris, Y. Cheng, W. Warren, A. T. Papenfuss, K. Belov, Submitted), may indicate a greatly reduced genetic diversity and adaptive potential within this family in the Tasmanian devil. Within the cytokines, the chemokine family appeared to show the greatest SNP diversity. Two of the chemokine genes containing putative nonsynonymous SNPs were *CCLD6* and *CCLD7* which form part of a devil-specific chemokine gene expansion (K. M. Morris, Y. Cheng, W. Warren, A. T. Papenfuss, K. Belov, Submitted).

Within the NK receptors are two families. In marsupials, including the devil, genes of the LRC family are relatively conserved, and these species have much fewer genes in this family than eutherian and monotreme mammals (Belov *et al.* 2007; van der Kraan *et al.* 2013). None of the six genes within this family had any putative SNPs within their exons in devils. This suggests that these genes may have a more specific, conserved function in marsupials leading to selection against polymorphisms or gene duplications. However, devils do have a large number of NK receptors within the KIR family, and it has previously been speculated that marsupials may rely predominantly on this family as NK receptors (van der Kraan *et al.* 2013). In addition

Table 4 SNP marker statistics (sample sizes provided in Table 3)

SNP	Genomic position	Location in gene	H_O	H_E	P	MAF
CCLD5/6-1	GL857182:18313	Exon	0.313	0.459	$2.17 \times 10^{-5*}$	0.358
CCLD5/6-2	GL857182:19638	Intron	0.311	0.458	$1.97 \times 10^{-5*}$	0.356
CCLD5/6-3	GL857182:19990	Exon	0.811	0.500	$8.04 \times 10^{-5*}$	0.500
CCLD5/6-4	GL857182:19992	Exon	0.906	0.500	$1.37 \times 10^{-9*}$	0.491
CCLD5/6-5	GL857182:20241	Up/downstream	0.321	0.448	0.200	0.339
CCLD5/6-6	GL857182:20819	Up/downstream	0.310	0.414	0.190	0.293
CCLD5/6-7	GL857182:20831	Up/downstream	0.000	0.071	0.019*	0.037
CCLD5/6-8	GL857182:21677	Up/downstream	0.290	0.467	0.051	0.371
CCLD5/6-9	GL857182:22160	Up/downstream	0.000	0.083	0.022*	0.043
CCLD5/6-10	GL857182:24020	Up/downstream	0.000	0.059	0.015*	0.030
CCLD5/6-11	GL857182:24073	Up/downstream	0.000	0.069	0.018*	0.036
CCLD5/6-12	GL857182:24534	Exon	0.063	0.061	1.000	0.031
CCLD5/6-13	GL857182:24573	Exon	0.061	0.059	1.000	0.030
CCLD5/6-14	GL857182:24632	Intron	0.029	0.084	0.045*	0.044
CCLD5/6-15	GL857182:24648	Intron	0.000	0.062	0.016*	0.032
CCLD5/6-16§	GL857182:25136	Exon	0.286	0.245	1.000	0.143
CX3C-1	GL834650:1670393	Intron	0.095	0.090	1.000	0.047
CX3C-2	GL834650:1671527	Intron	0.394	0.468	0.048*	0.374
CX3C-3§	GL834650:1679470	Exon	0.101	0.213	$2.19 \times 10^{-8*}$	0.121
TGFB1-1	GL849759:218979	Downstream	0.424	0.436	0.812	0.321
TGFB1-2	GL849759:220632	Intron	0.065	0.063	1.000	0.033
IL-22F1-1	GL834513:1087266	Upstream	0.000	0.067	0.018*	0.034
IL-22F1-2	GL834513:1088712	Upstream	0.107	0.101	1.000	0.054
IL-22-1	GL861826:55229	Intron	0.559	0.445	0.005*	0.335
IL-22-2	GL861826:56392	Downstream	0.163	0.300	4.95×10^{-5}*	0.184
IL-22-3	GL861826:56394	Downstream	0.784	0.497	$8.63 \times 10^{-11*}$	0.464
IL-22-4	GL861826:56460	Downstream	0.556	0.499	0.7375	0.472
IL-22-5	GL861826:56483	Downstream	0.065	0.063	1	0.033
IL-22-6	GL861826:56832	Downstream	0.012	0.068	$1.48 \times 10^{-8*}$	0.035
IL-22-7	GL861826:57529	Downstream	0.012	0.130	$6.79 \times 10^{-16*}$	0.070
IL-22-8	GL861826:57560	Downstream	0.064	0.174	$4.06 \times 10^{-10*}$	0.096
IL-22-9	GL861826:57793	Downstream	0.262	0.354	0.001*	0.230
IL-22-10	GL861826:57962	Downstream	0.731	0.473	$2.95 \times 10^{-14*}$	0.383
IL-22-11	GL861826:58006	Downstream	0.712	0.459	$7.79 \times 10^{-16*}$	0.356
IL-22-12	GL861826:58060	Downstream	0.688	0.459	$2.23 \times 10^{-12*}$	0.356
IL-22-13	GL861826:58085	Downstream	0.692	0.459	$9.54 \times 10^{-13*}$	0.358
IL-22-14	GL861826:58120	Downstream	0.731	0.473	$2.95 \times 10^{-14*}$	0.383
IL-22-15	GL861826:58220	Downstream	0.655	0.448	$1.20 \times 10^{-10*}$	0.339
IL-22-16	GL861826:58226	Downstream	0.696	0.461	$8.15 \times 10^{-13*}$	0.360
IL-22-17	GL861826:58229	Downstream	0.696	0.461	$8.15 \times 10^{-13*}$	0.360
IL-22-18	GL861826:58425	Downstream	0.426	0.335	0.003*	0.213
IL-17B-1	GL834622:505431	Intron	0.183	0.183	1.000	0.102
IL-17B-2	GL834622:505749	Intron	0.183	0.183	1.000	0.102
IL-17B-3	GL834622:505871	Intron	0.187	0.187	1.000	0.104
IL-17B-4	GL834622:506399	Intron	0.185	0.185	1.000	0.103
IL-17B-5	GL834622:506405	Intron	0.185	0.185	1.000	0.103
IL-17B-6	GL834622:506477	Intron	0.186	0.186	1.000	0.104
IL-17B-7§	GL834622:506759	Exon	0.187	0.187	1.000	0.104
IL-17B-8§	GL834622:506769	Exon	0.187	0.187	1.000	0.104
IL-17B-9	GL834622:507289	Intron	0.187	0.187	1.000	0.104
IL-17B-10	GL834622:507488	Intron	0.187	0.187	1.000	0.104
IL-17B-11	GL834622:508357	Intron	0.172	0.192	0.236	0.108
IL-17B-12	GL834622:508358	Intron	0.177	0.188	0.428	0.105
IL-17B-13	GL834622:508460	Intron	0.186	0.186	1.000	0.104
IL-17B-14	GL834622:509409	Downstream	0.186	0.177	1.000	0.098
IL-17B-15	GL834622:509694	Downstream	0.187	0.178	1.000	0.099

Table 4 *Continued*

SNP	Genomic position	Location in gene	H_O	H_E	*P*	MAF
IL-17B-16	GL834622:510102	Downstream	0.186	0.186	1.000	0.104
IL-17B-17	GL834622:510358	Downstream	0.182	0.183	1.000	0.102
IL-17B-18	GL834622:510694	Downstream	0.185	0.185	1.000	0.103
IL-17B-19	GL834622:510709	Downstream	0.186	0.186	1.000	0.104
IL-17B-20	GL834622:510754	Downstream	0.187	0.187	1.000	0.104
DIG12-1	GL850167:29387	Upstream	0.189	0.200	0.443	0.112
DIG12-2	GL850167:29390	Upstream	0.180	0.210	0.069	0.119
DIG12-3	GL850167:29403	Upstream	0.227	0.236	0.535	0.136
DIG12-4	GL850167:29463	Upstream	0.230	0.246	0.366	0.143
DIG12-5	GL850167:29553	Upstream	0.225	0.233	0.539	0.135
DIG12-6	GL850167:29827	Upstream	0.233	0.239	0.754	0.139
DIG12-7	GL850167:30779	Upstream	0.233	0.239	0.754	0.139
DIG12-8	GL850167:30840	Upstream	0.233	0.239	0.754	0.139
DIG12-9	GL850167:30965	Upstream	0.232	0.238	0.753	0.138
DIG12-10	GL850167:31097	Upstream	0.232	0.238	0.753	0.138
DIG12-11	GL850167:31458	Upstream	0.233	0.239	0.754	0.139
DIG12-12	GL850167:31848	Upstream	0.232	0.238	0.753	0.138
DIG12-13	GL850167:32283	Exon	0.227	0.236	0.535	0.136
DIG12-14	GL850167:33128	Intron	0.229	0.236	0.749	0.137
DIG12-15	GL850167:33597	Exon	0.232	0.239	0.750	0.138
DIG12-16	GL850167:35520	Downstream	0.246	0.248	0.770	0.145
DIG24-1	GL849546:132660	Upstream	0.310	0.459	0.004*	0.357

*H*O, observed heterozygosity *H*E, expected heterozygosity, §Nonsynonymous SNP, **P* value for significant deviation from HWE heterozygosity deficiency, MAF, minimum allele frequency.

Table 5 Diversity statistics for genes

	n	SNPs	π	*h*	*hd*	*k*	Tajima's D	*Z* statistic
CCLD5	40	3	13.4	4	0.153	0.114	−1.342	−0.924
CCLD6	69	3	14	7	0.747	1.451	2.818 **	−0.992
CX3C	167	1	0.6	2	0.215	0.215	0.35	1.012
TGFB1	93	0	0	1	—	—	—	—
IL-22F1	29	0	0	1	—	—	—	—
IL-22	196	0	0	1	—	—	—	—
IL-17B	186	2	12.2	2	0.201	0.401	0.389	1.321
DIG12	186	2	2.7	3	0.268	0.531	0.902	−0.987
DIG24	83	0	0	1	—	—	—	—

n, number of individuals; π, nucleotide diversity ($\times 10^4$); *h*, number of inferred haplotypes; *hd* =haplotype diversity; *k*, average number of nucleotide differences; $^{**}P < 0.01$.

to having a large expansion of these genes, six of the 17 genes investigated in this family had putative nonsynonymous SNPs, suggesting that these genes may be under diversifying selection. Such selection is most likely pathogen-mediated as NK receptors are key components of the innate immune response and may evolve rapidly in response to viral pathogens (Kelley *et al.* 2005). A large number of genes with nonsynonymous SNPs were also seen in the extended LRC.

Several genes containing putative SNPs within the devil have been showed to be associated with disease resistance or susceptibility in other species. For example, a nonsynonymous variant in *IL-1A* has been associated with susceptibility to ovarian cancer (Charbonneau *et al.* 2014) and H1N1 influenza (Liu *et al.* 2013). In cattle, a SNP upstream of the *IL-8* was associated with resistance to mastitis (Chen *et al.* 2015). Two SNP variants in *IL-33* were associated with lower risk of Behcet's disease (Koca *et al.* 2015), while SNPs in the UTR of *IL-10* conferred resistance to psoriasis in humans (Al-Balbeesi *et al.* 2015). Also in humans, a nonsynonymous SNP in *FASLG* has been associated

with autoimmune lymphoproliferative syndrome (Bi *et al.* 2007) and SNPs in the intron of *FASLG* have been associated with increased risk of cutaneous malignant melanoma (Li *et al.* 2006) and HIV progression (Nasi *et al.* 2005). A SNP in a NK receptor gene *NKG2D* was associated with lower risk of recurrent miscarriages in women (Hizem *et al.* 2014). These studies demonstrate that SNPs within these genes can have an adaptive role in other mammals; therefore, these SNPs in the devil may also have adaptive potential.

Amplicon SNPs

Eight amplicons covering nine genes that contained putative nonsynonymous SNPs were designed and amplified in 29 to 220 devils per amplicon. Four nonsynonymous SNPs were found in *IL-17B* and two chemokines, *CX3C* and *CCLD5*. Polymorphism in these genes has not been studied in other wild species to our knowledge. Within the eight amplicons, 188 SNPs were predicted based on sequencing data in ten devils, but only 78 SNPs were confirmed experimentally. The number of SNPs within the amplicons was less than predicted from resequencing data, despite the greater sample numbers in the amplicon assays.

There are several potential reasons for the discrepancy between the number of predicted SNPs and the SNPs confirmed in amplicon sequencing. Most likely the discrepancy lies with the difference in filtering and error rate between the two methods; the predicted SNP rate may be elevated due to errors in the resequencing data. Some regions of the genome had lower read coverage from the resequencing data than others, particularly in regions where the genome assembly is highly fragmented. Although we attempted to place the amplicons in areas that had high-quality read coverage, these markers were selected based on high SNP density and the presence of nonsynonymous SNPs. One amplicon in particular, *IL-22*, was the source of the largest discrepancy between predicted SNPs and those confirmed through amplicon sequencing. This amplicon may have been placed in a region with higher frequency of errors.

Additionally, the amplicon data are filtered, removing SNPs that occur at a frequency of <3%, as these are more likely to be the results of errors. It is possible that this removes rare alleles, which were detected in the resequencing data, although this is unlikely to be a major factor because only ten individuals were included in the initial panel, while up to 220 devils were included in the amplicon sequencing. Thus, the amplicon sequencing data are more likely to reflect true SNP diversity due to the larger sample size and filtering. Another possible explanation for differences between predicted and observed SNP diversity is ascertainment bias; the samples used to predict SNPs were primarily from a single population in northwestern Tasmania, and these samples were not included in the amplicon phase of the study, while samples for the amplicon assays were sourced from diverse locations. Higher SNP diversity may be present in the samples from the resequenced population accounting for some of the difference in SNPs between the two approaches.

Amplicon diversity and selection

Evidence that selection is occurring was found for one of the amplicons, the *CCLD5/6* amplicon, which contained both the *CCLD5* and *CCLD6* genes. These two genes are part of a group of genes which appear to have recently duplicated within the devil lineage (K. M. Morris, Y. Cheng, W. Warren, A. T. Papenfuss, K. Belov, Submitted), suggesting that there may be diversifying selection acting on these genes. The *CCLD6* gene showed deviation from neutrality using the Tajima's D test, with a highly significant positive Tajima D value, which is a signature of balancing selection. While significant values in the Tajima's D test can also indicate demographic effects, demographic effects are expected to act upon genetic diversity at all loci (Beaumont & Balding 2004) which was not seen in this study. This indicates that balancing selection is acting on this locus. Additionally, *CCLD6* also had the highest diversity of any of the genes as estimated by nucleotide diversity, haplotype diversity and average number of nucleotide differences. This high diversity is consistent with the presence of diversifying selection. Apart from *CCLD6*, no other locus showed significant deviation from neutrality using Tajima's D or the *Z*-test.

IL-22 also had a large number of SNPs deviating from HWE with 16 of the 18 SNPs deviating, mostly with a heterozygote excess. The SNPs showing a heterozygote excess were concentrated at one end of the *IL-22* amplicon. This may indicate that selection is occurring in one of these SNPs, or at variation linked to these SNPs; however, no significant deviation from neutrality was seen in the Tajima's D test. Additionally, the nonsynonymous SNP within the *CX3C* gene showed deviation from HWE (heterozygote deficit). Genes that showed the highest diversity based on the nucleotide diversity and mean nucleotide diversity were *CCLD6*, *CCLD5* and *IL-17B,* while the genes with the lowest diversity were *TGFB1, IL-22, IL-22F1* and *DIG24* which had no SNPs within their coding sequence.

Low SNP diversity across immune genes

Overall, the number of SNPs identified across the immune genes studied was low compared with other

wild populations. Few studies have looked at SNP variation in wild populations in genes outside the MHC and TLR families. Where comparisons can be made, devils show lower or no diversity at genes that are polymorphic in other species.

In a study of field vole (*Microtus agrestis*) cytokines, Turner *et al.* (2011) found polymorphisms in seven of ten studied genes across twelve voles (totalling 15 SNPs). In comparison, the same ten cytokines in devil (comprising 13 devil sequences in total due to duplication of the *IL-18* locus) showed only a single putative SNP (in *TGFB1*). This SNP was not confirmed by amplicon sequencing and therefore may be the result of sequencing error. In a subsequent study, Turner *et al.* (2012) reported an average π of 15.1 (SD 7.9) in nine field vole cytokines, higher than observations in devils in the same cytokines (where average $\pi = 8.7$; SD 6.2). Average haplotype diversity was also lower in devils (0.32; SD 0.24) than in the field voles (0.50; SD 0.24). Overall therefore, the comparison between these two species highlights the severe lack of genetic diversity across Tasmanian devil immune genes.

Considering other gene families, in a study of *MIC* gene polymorphisms in primates, 18 alleles were observed in *MICA/B* in chimpanzee ($N = 63$), 4 alleles in gorilla *MICA* (N = 18) and 10–13 alleles at each in three macaque *MIC* genes (N = 72) (Meyer *et al.* 2014). In contrast, the two devil *MIC* genes were monomorphic across the 10 genomes we examined. In a study of *IL-8* in the large yellow croaker (*Larimichthys crocea*), five SNPs were observed across 20 individuals (Li & Yao 2013); we found no SNPs in the same regions for devil. Similarly, seven SNPs were identified in the large yellow croaker *MSTN* gene (N = 48) (Xue *et al.* 2008), but we found no variation in the orthologous region of devils. High diversity and an association with disease resistance were found in Atlantic Sea bass *IL-1B* (Chistiakov *et al.* 2008, 2010); again, no diversity was found in this gene in devils. The only possible exception to the general pattern of low diversity in devils is at *LGP2* where a study of grass carp observed 1 synonymous and 1 nonsynonymous SNP (N = 20) (Wan *et al.* 2013), similar to our observations for this gene in devils. Nevertheless, a general pattern of low devil immunogenetic diversity, compared to other wild populations, is striking.

The Tasmanian devil is an extremely vulnerable species due to its lack of genetic diversity. Low genetic diversity has previously been demonstrated in MHC loci (Siddle *et al.* 2007a,b; Cheng *et al.* 2012) and at TLR loci (Cui *et al.* 2015), and it has now been demonstrated that this lack of genetic diversity extends to a large number of immune genes. The history of the Tasmanian devils over the last 10 000 years has been one of population contractions, population crashes and local extinctions (Guiler 1992; Brown 2006). Studies have suggested that devil genetic diversity has been low since at least the mid-Holocene (Morris *et al.* 2012; Brüniche-Olsen *et al.* 2014). The lack of diversity exhibited across immune genes in the devil may represent loss of adaptive potential, explaining their historic population decline. As polymorphisms of diverse immune genes play a critical role in disease susceptibility, a lack of diversity across the genes of the immune system may have made the devil population extremely vulnerable to disease epidemics due to introduced diseases. Diseases may have been introduced by the arrival of aboriginal humans, dingos or European humans, events which correlate with devil population declines (Guiler 1970, 1992; Brown 2006; Brüniche-Olsen *et al.* 2014). Reports of a distemper-like disease in the devil population coincided with a population crash in the 19th century (Guiler 1970). Furthermore, the devil's lack of genetic diversity may have contributed to the ability of DFTD to spread between unrelated individuals. In particular, the lack of diversity in NK receptor genes may be involved in the ability of the disease to avoid a NK cell-mediated response. The role of these cells in DFTD should be investigated in further detail in the future. This demonstrates how critical immune gene diversity is for wild populations, and suggests that a lack of such diversity can result in rapid population collapse.

Conclusion

In this study, we have, for the first time, examined immune gene diversity broadly in a wild species. By utilizing genome sequencing data of ten devils, we have examined SNP diversity in 167 devil immune genes from a range of key immune families. Furthermore, we have developed an assay to examine diversity at multiple immune genes that allows for high-throughput analysis of sample sets. This is a highly novel approach for studying functional genetic diversity in a nonmodel species. This approach allowed us to examine diversity of nine immune genes in greater depth in a very large sample set of 221 devils.

Immune genes are some of the most polymorphic genes within the genome, and variation at many may be crucial for population-level disease resistance in the devil. In this study, both the SNP predictions based on resequencing data, and the data from amplicon sequencing confirm the extreme paucity of variation within the devil genome, even within key functional loci that may be important for disease resistance. SNPs that resulted in changes to the coding sequence were particularly rare.

While genetic diversity at functional loci in the devil appears to be critically low, some diversity remains at

MHC, TLR and at a range of genes identified in this study. SNP variation, including SNPs within possible regulatory regions, introns, exons and several nonsynonymous SNPs were identified in this study. Four nonsynonymous SNPs were confirmed by amplicon sequencing. Furthermore, signatures of selection were detected in one gene, suggesting that it may have adaptive potential. Maintaining any remaining functional diversity in the devil insurance population will be essential for maintaining the adaptive potential of this species and for maintenance of the species in the wild long term.

Over the next 5 years, the insurance population will provide source animals for localized site re-introductions, and is an insurance resource to repopulate Tasmania should complete extinction in the wild occur. Loss of functional variation at immune genes in the insurance population devils would further increase the devils susceptibility to disease outbreaks. Issues such as disproportionate contribution from founder devils to the insurance population, genetic drift and relaxed selection in captivity due to lower parasite and disease loads could lead to changes in allele frequencies and potential loss of remaining polymorphisms within the immune genes of devils. Therefore, the amplicon markers developed in this study, in combination with MHC typing and use of neutral markers, should be used in the insurance population to monitor genetic diversity across generations, to ensure that genetic diversity, and most critically variation at functional loci, is not lost from the insurance population. These markers also provide a resource for examining functional genetic diversity between wild populations and for examining changes over time. Additionally, these assays could be used to examine the relationship between genetic variants in key genes in the immune system, and resistance to disease, including, but not limited to, DFTD. This resource may be vital for the long-term conservation of this iconic Australian species.

This study has greatly expanded the scope of immune gene diversity investigation in wild populations. While MHC diversity has been the focus of the majority of previous studies examining functional diversity in wild species, diversity in hundreds of immune genes that comprise the immunome, which account for a large proportion of the variability underlying disease resistance, have been largely overlooked. By examining diversity in a large number of diverse immune genes in wild species, we can gain a greater understanding of the role of immune gene diversity in disease resistance, population fitness and selective processes, and such an approach may play a key role in the management of threatened species in the future.

Acknowledgments

This work was funded by the Australian Research Council and the Save the Tasmanian Devil Appeal. KB is supported by an ARC Future Fellowship. We thank Menna Jones and Rodrigo Hamede from the School of Zoology at the University of Tasmania for West Pencil Pine samples. We thank the Zoo and Aquarium Association and all the zoos participating in the insurance programme (www.zooaquarium.org.au/devils) for insurance population samples. These include Devil Ark; Taronga Zoo; Healesville Sanctuary; Tasmanian Government Department of Primary Industries, Parks, Water and Environment; Australia Walkabout Wildlife Park; Wild Life Sydney Zoo; National Zoo & Aquarium; Perth Zoo; Monarto Zoo; Moonlit Sanctuary Wildlife Conservation Park; Kyabram Fauna Park; Halls Gap Zoo; Lone Pine Koala Sanctuary; Currumbin Wildlife Sanctuary; and Bonorong Wildlife Sanctuary.

References

Abrantes J, Areal H, Esteves PJ (2013) Insights into the European rabbit (*Oryctolagus cuniculus*) innate immune system: genetic diversity of the toll-like receptor 3 (TLR3) in wild populations and domestic breeds. *BMC Genetics*, **14**, 73.

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology and Evolution*, **21**, 433–438.

Al-Balbeesi AO, Halwani M, Alanazi M *et al.* (2015) Novel mutations in IL-10 promoter region -377 (C>T), -150 (C>A) and their association with psoriasis in the Saudi population. *Asian Pacific Journal of Cancer Prevention: APJCP*, **16**, 1247–1250.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Belov K, Sanderson CE, Deakin JE *et al.* (2007) Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Research*, **17**, 982–991.

Bi LL, Pan G, Atkinson TP *et al.* (2007) Dominant inhibition of Fas ligand-mediated apoptosis due to a heterozygous mutation associated with autoimmune lymphoproliferative syndrome (ALPS) Type Ib. *BMC Medical Genetics*, **8**, 41.

Bollmer JL, Ruder EA, Johnson JA, Eimes JA, Dunn PO (2011) Drift and selection influence geographic variation at immune loci of prairie-chickens. *Molecular Ecology*, **20**, 4695–4706.

Brown O (2006) Tasmanian devil (*Sarcophilus harrisii*) extinction on the Australian mainland in the Mid-Holocene: multicausality and ENSO intensification. *Alcheringa*, **30**, 49–57.

Brüniche-Olsen A, Jones ME, Austin JJ, Burridge CP, Holland BR (2014) Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. *Biology Letters*, **10**, 20140619.

CBSG (2008) *Tasmanian Devil PHVA Final Report*. IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, MN.

Charbonneau B, Block MS, Bamlet WR *et al.* (2014) Risk of ovarian cancer and the NF-κB pathway: genetic association with IL1A and TNFSF10. *Cancer Research*, **74**, 852–861.

Chen RJ, Wang ZZ, Yang ZP *et al.* (2015) Association of IL8 -105G/A with mastitis somatic cell score in Chinese Holstein dairy cows. *Animal Biotechnology*, **26**, 143–147.

Cheng Y, Sanderson C, Jones M, Belov K (2012) Low MHC class II diversity in the Tasmanian devil (*Sarcophilus harrisii*). *Immunogenetics*, **64**, 525–533.

Chistiakov DA, Tsigenopoulos CS, Lagnel J *et al.* (2008) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Animal Genetics*, **39**, 623–634.

Chistiakov DA, Kabanov FV, Troepolskaya OD, Tischenko MM (2010) A variant of the interleukin-1beta gene in European sea bass, *Dicentrarchus labrax* L., is associated with increased resistance against *Vibrio anguillarum*. *Journal of Fish Diseases*, **33**, 759–767.

Cui J, Cheng Y, Belov K (2015) Diversity in the Toll-like receptor genes of the Tasmanian devil (*Sarcophilus harrisii*). *Immunogenetics*, **67**, 195–201.

Doherty PC, Zinkernagel RM (1975) A biological role for the major histocompatibility antigens. *Lancet*, **28**, 1406–1409.

DPIPWE (2011) *Save the Tasmanian Devil Program: 2010/11 Annual Program Report*. Department of Primary Industries, Parks, Water and Environment, Hobart, Tasmania.

Else K, Wakelin D (1988) The effects of H-2 and non-H-2 genes on the expulsion of the nematode *Trichuris muris* from inbred and congenic mice. *Parasitology*, **96**(Pt 3), 543–550.

Fumagalli M, Pozzoli U, Cagliani R *et al.* (2009) Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *Journal of Experimental Medicine*, **206**, 1395–1408.

Futas J, Horin P (2013) Natural killer cell receptor genes in the family Equidae: not only Ly49. *PLoS ONE*, **8**, e64736.

Gao L, Du X, Su H *et al.* (2013) The polymorphisms of chemokine gene in channel catfish (*Ictalurus punctatus*) and the associations with susceptibility/resistance to *Edwardsiella ictaluri*. *HOAJ Biology*, **2**, 3.

Grueber CE, Wallis GP, King TM, Jamieson IG (2012) Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE*, **7**, e45011.

Guiler ER (1970) Tasmanian devils and agriculture. *Tasmanian Journal of Agriculture*, **41**, 134–137.

Guiler ER (1992) *The Tasmanian Devil*. St David's Park Publishing, Hobart, Australia.

Hizem S, Mtiraoui N, Massaoudi S *et al.* (2014) Polymorphisms in genes coding for the NK-cell receptor NKG2D and its ligand MICA in recurrent miscarriage. *American Journal of Reproductive Immunology*, **71**, 577–585.

Hogg CJ, Ivy JA, Srbet C *et al.* (in press) Influence of genetic provenance and birth origin on productivity of the Tasmanian devil insurance population. *Conservation Genetics*.

Huang H, Winter EE, Wang H *et al.* (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biology*, **5**, R47.

Hughes AL, Hughes MK (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics*, **42**, 233–243.

Jeffery KJ, Bangham CR (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.

Jepson A, Banya W, Sisay-Joof F *et al.* (1997) Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infection and Immunity*, **65**, 872–876.

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

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genetics*, **1**, 129–139.

Koca SS, Kara M, Deniz F *et al.* (2015) Serum IL-33 level and IL-33 gene polymorphisms in Behçet's disease. *Rheumatology International*, **35**, 471–477.

van der Kraan LE, Wong ES, Lo N, Ujvari B, Belov K (2013) Identification of natural killer cell receptor genes in the genome of the marsupial Tasmanian devil (*Sarcophilus harrisii*). *Immunogenetics*, **65**, 25–35.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li C, Yao CL (2013) Molecular and expression characterizations of interleukin-8 gene in large yellow croaker (*Larimichthys crocea*). *Fish and Shellfish Immunology*, **34**, 799–809.

Li C, Larson D, Zhang Z *et al.* (2006) Polymorphisms of the FAS and FAS ligand genes associated with risk of cutaneous malignant melanoma. *Pharmacogenetics and Genomics*, **16**, 253–263.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liu Y, Li S, Zhang G *et al.* (2013) Genetic variants in IL1A and IL1B contribute to the susceptibility to 2009 pandemic H1N1 influenza A virus. *BMC Immunology*, **14**, 37.

Meyer D, Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1–26.

Meyer A, Carapito R, Ott L *et al.* (2014) High diversity of MIC genes in non-human primates. *Immunogenetics*, **66**, 581–587.

Miller W, Hayes VM, Ratan A *et al.* (2011a) Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences, USA*, **108**, 12348–12353.

Morris K, Austin JJ, Belov K (2012) Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. *Biology Letters*, **9**, 20120900.

Murchison EP, Schulz-Trieglaff OB, Ning Z *et al.* (2012) Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell*, **148**, 780–791.

Nasi M, Pinti M, Bugarini R *et al.* (2005) Genetic polymorphisms of Fas (CD95) and Fas ligand (CD178) influence the rise in CD4 + T cell count after antiretroviral therapy in drug-naïve HIV-positive patients. *Immunogenetics*, **57**, 628–635.

Pacheco AG, Cardoso CC, Moraes MO (2008) IFNG +874T/A, IL10 -1082G/A and TNF -308G/A polymorphisms in association with tuberculosis susceptibility: a meta-analysis study. *Human Genetics*, **123**, 477–484.

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proceedings of the National Academy of Sciences, USA*, **95**, 3714–3719.

Pearse AM, Swift K (2006) Allograft theory: transmission of devil facial-tumour disease. *Nature*, **439**, 549.

Piertney S, Oliver M (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Shiina T, Inoko H, Kulski JK (2004) An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens*, **64**, 631–649.

Siddle H, Kriess A, Eldridge M *et al.* (2007a) Transmission of a fatal clonal tumour by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences, USA*, **104**, 16221–16226.

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.

Siddle HV, Kreiss A, Tovar C *et al.* (2013) Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proceedings of the National Academy of Sciences, USA*, **110**, 5103–5108.

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

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Tonteri A, Vasemägi A, Lumme J, Primmer CR (2010) Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci. *Molecular Ecology*, **19**, 1273–1282.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PR, Råberg L (2013) Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130364.

Turner AK, Begon M, Jackson JA, Bradely JE, Paterson S (2011) Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. *PLoS Genetics*, **7**, e1002343.

Turner AK, Begon M, Jackson JA, Paterson S (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology*, **21**, 1632–1646.

Ujvari B, Belov K (2011) Major Histocompatibility Complex (MHC) markers in conservation biology. *International Journal of Molecular Sciences*, **12**, 5168–5186.

Vallinoto AC, Graça ES, Araújo MS *et al.* (2010) IFNG +874T/A polymorphism and cytokine plasma levels are associated with susceptibility to Mycobacterium tuberculosis infection and clinical manifestation of tuberculosis. *Human Immunology*, **71**, 692–696.

Wan Q, Wang L, Su J *et al.* (2013) Genetic structure, polymorphism identification of LGP2 gene and their relationship with the resistance/susceptibility to GCRV in grass carp, *Ctenopharyngodon idella*. *Gene*, **521**, 166–175.

Whiteside TL, Herberman RB (1995) The role of natural killer cells in immune surveillance of cancer. *Current Opinion in Immunology*, **7**, 704–710.

Xue L, Li T, Sun S *et al.* (2008) Genetic polymorphism analysis of exon I in myostatin gene from large yellow croaker *Pseudosciaena crocea*. *Fisheries Science*, **27**, 503–506.

All authors contributed to the design of the study. K.M.M. performed bioinformatic SNP prediction, selected gene targets and designed primers. K.M.M. optimized primers and K.M.M., B.W. and C.E.G. performed PCRs, gel analysis, normalization and library preparation. B.W. performed analysis of Illumina data, and B.W. and K.M.M. performed analysis of SNP genotypes. K.M.M. wrote the manuscript with contributions and editing from B.W., C.E.G., C.H. and K.B. All authors read and approved the final manuscript.

Data accessibility

Miseq reads, tped and tfam files for PLINK, and final SNP genotyping data: DRYAD entry doi:10.5061/dryad.m8297.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Sample details including location of the devil and the amplicons at which the samples were type.

Table S2. SNPs identified from resequencing data.