

# **Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations**

Alexandra K. Fraik<sup>1</sup>, Mark J. Margres<sup>1</sup>, Brendan Epstein<sup>1,2</sup>, Soraia Barbosa<sup>3</sup>, Menna Jones<sup>4</sup>, Sarah Hendricks<sup>3</sup>, Barbara Schönfeld<sup>4</sup>, Amanda R. Stahlke<sup>3</sup>, Anne Veillet<sup>3</sup>, Rodrigo Hamede<sup>4</sup>, Hamish McCallum<sup>5</sup>, Elisa Lopez-Contreras<sup>1</sup>, Samantha J. Kallinen<sup>1</sup>, Paul A. Hohenlohe<sup>3</sup>, Joanna L. Kelley<sup>1</sup>, Andrew Storfer<sup>1</sup>

**Affiliations:** <sup>1</sup>*School of Biological Sciences, Washington State University Pullman, WA 99164, USA;* <sup>2</sup>*Plant Biology, University of Minnesota, Minneapolis, MN 55455, USA;* <sup>3</sup>*Department of Biological Sciences, University of Idaho, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive, Moscow, Idaho 83844, USA;* <sup>4</sup>*School of Biological Sciences, Hobart, TAS 7004, Australia;* <sup>5</sup>*School of Environment, Griffith University Nathan, QLD 4111, Australia*

**Emails for correspondence:** Alexandra Fraik, School of Biological Sciences, Washington State University, Pullman, WA, [alexandra.fraik@wsu.edu](mailto:alexandra.fraik@wsu.edu); Andrew Storfer, School of Biological Sciences, Washington State University, Pullman, WA, [astorfer@wsu.edu](mailto:astorfer@wsu.edu)

**Keywords:** landscape genomics, genetic-environmental associations, Tasmanian devils, devil facial tumor disease, population genetics, local adaptation

## Abstract

Landscape genomics studies focus on identifying candidate genes under selection via spatial variation in abiotic environmental variables, but rarely by biotic factors such as disease. The Tasmanian devil (*Sarcophilus harrisii*) is found only on the environmentally heterogeneous island of Tasmania and is threatened with extinction by a nearly 100% fatal, transmissible cancer, devil facial tumor disease (DFTD). Devils persist in regions of long-term infection despite epidemiological model predictions of species' extinction, suggesting possible adaptation to DFTD. Here, we test the extent to which spatial variation and genetic diversity are associated with the abiotic environment and/or by DFTD. We employ genetic-environment association (GEAs) analyses using a RAD-capture panel consisting of 6,886 SNPs from 3,286 individuals sampled pre- and post-disease arrival. Pre-disease, we find significant correlations of allele frequencies with environmental variables, including 349 unique loci linked to 64 genes, suggesting local adaptation to abiotic environment. Following DFTD arrival, however, we detected few of the pre-disease candidate loci, but instead frequencies of candidate loci linked to 14 genes correlated with disease prevalence. Loss of apparent signal of abiotic local adaptation following disease arrival suggests swamping by the strong selection imposed by DFTD. Further support for this result comes from the fact that post-disease candidate loci are in linkage disequilibrium with genes putatively involved in immune response, tumor suppression and apoptosis. This suggests the strength GEA associations of loci with the abiotic environment are swamped resulting from the rapid onset of a biotic selective pressure.

## Introduction

A central goal of molecular ecology is understanding how ecological processes generate and maintain the geographic distribution of adaptive genetic variation. Landscape genomics has emerged as a popular framework for identifying candidate loci that underlie local adaptation (Manel et al. 2010; Rellstab et al. 2015; Hoban et al. 2016; Lowry et al 2016; Storfer et al. 2017). Using genomic-scale data sets, researchers screen for loci that exhibit patterns of selection across heterogeneous environments (Haas et al. 2016). One widely-used method for testing for statistical associations of allele frequencies of marker loci across the genome with environmental variables is genetic-environmental associations (GEAs) (Rellstab et al. 2015; Francois et al. 2016; Whitlock & Lotterhos 2015; Hoban et al. 2016).

GEAs identify significant correlations of allele frequencies at candidate loci with abiotic environmental variables such as altitude, rainfall, and temperature. These efforts have been successful, for example, by identifying hypoxia adaptation in high-elevation human populations (Beall 2007, Peng et al. 2010), stress response in lichen populations along altitudinal gradients (Dal Grande 2017), and leaf longevity and morphogenesis in response to aridity (Steane et al. 2014). Climatic, geographic, and fine-scale remote sensing data that explain large amounts of heterogeneity in the environment are often easily obtained (Rellstab et al. 2015). However, few landscape genomics studies have tested for the influence of biotic variables on the spatial distribution of adaptive genetic variation.

Infectious diseases often impose strong selective pressures on their host and thereby represent key biotic variables increasingly recognized for their severe impacts on natural populations (summarized in Kozakiewicz et al. 2018, e.g. Biek et al. 2010; Wenzel et al 2015;

Leo et al. 2016; Mackinnon et al. 2016; Eoche-Bosy et al. 2017). However, data on biotic factors such as life history traits (Sun et al. 2015), community composition (Harrison et al. 2017), or disease prevalence often involve extensive field work and are far more difficult and labor-intensive to collect than abiotic variables. Landscape genomic studies of disease can help guide management programs, such as captive breeding designs and reintroductions (Hoban et al. 2016; Hohenlohe et al. 2018). Additionally, landscape genomics studies have been used elucidate how the landscape influences the distribution of both the host and vector (Blanchong et al. 2008; Robinson et al. 2013; Schwabl et al. 2017). However, studies disentangling the influence of pathogen dynamics from that of other abiotic landscape features have had limited statistical power to date. For example, Wenzel et al. (2015) tested for correlations between parasite burden and genetic differentiation across the landscape but did not detect any statistically significant associations due to noise created by the genomic background. Disease may play an important role in the distribution of adaptive genetic variation across the landscape, potentially swamping signatures of local adaptation to the abiotic environment.

Tasmanian devils (*Sarcophilus harrisii*) and their transmissible cancer, devil facial tumor disease (DFTD), offer such an opportunity. This species is isolated to the island of Tasmania and has been sampled across its entire geographic range. Intense mark-recapture studies and collection of thousands of genetic samples has been conducted over the past 20 years both before and after DFTD arrival across multiple populations. This sampling effort, in addition to resources such as an assembled genome (Murchison et al. 2012), provide extensive data to employ GEAs to test for the relative effects of abiotic environmental factors versus DFTD.

In 1996, the first evidence of DFTD was documented in Mt. William/Wukalina National Park (McCallum 2008). In just over two decades, DFTD has spread across the majority of Tasmania, with nearly a 100% case fatality rate (Hawkins et al. 2006; McCallum et al. 2009). Transmission appears to be largely frequency-dependent (McCallum et al. 2009; McCallum 2012), with tumors originating on the face or in the oral cavity and being transferred as allografts (Pearse et al. 2006) through biting during social interactions (Hamede et al. 2009). DFTD has a single, clonal Schwann cell origin (Siddle et al. 2010) and evades host immune system detection via down-regulation of host major histocompatibility complex (Siddle et al. 2013). Low overall genetic variation in devils resulting from population bottlenecks (Miller et al. 2011) that occurred during the last glacial maximum, as well as during extreme El Niño events 3,000-5,000 years ago (Brüniche-Olsen et al. 2014) has also been attributed to high susceptibility to this emerging infectious disease. DFTD likely imposes extremely strong selection as it threatens Tasmanian devils with extinction with resulting local population declines exceeding 90%, and an overall species-wide decline of at least 80% (Jones et al. 2004; Lachish et al. 2009). However, small numbers of devils persist in areas with long-term infection likely resulting from evolutionary responses in the Tasmanian devil (Epstein et al. 2016; Wright et al. 2017). Indeed, recent work has shown: 1) rapid evolution in genes associated with immune-related functions across multiple populations (Epstein et al. 2016), 2) sex-biased response in a few large-effect loci in survival following infection (Margres et al. 2018a), 3) evidence of effective immune response (Pye et al. 2016) and 4) cases of spontaneous tumor regression (Pye et al. 2016; Wright et al. 2017; Margres et al. 2018b). The latter two studies, however, focused on small geographic areas.

Here, we estimated allele frequencies in 6,886 SNPs both randomly selected and previously shown to be associated with DFTD (Epstein et al. 2016) in 3,286 devils from seven localities throughout Tasmania sampled both prior to and following disease arrival. Our study had five objectives to test the relative effects of abiotic environmental variables and DFTD on spatial genetic variation pre- and post-disease arrival: 1) to estimate the number of genetic clusters to determine whether population declines caused by disease altered population structure; 2) to identify genetic-environmental associations of devil populations to abiotic variables prior to disease arrival; 3) to identify genetic-environmental associations to both abiotic variables and DFTD prevalence post-disease arrival; 4) to determine whether signals of local adaptation to abiotic variables pre-disease were weakened by the strong selection imposed by DFTD; 5) to assess whether changes in genetic variation pre- and post-disease arrival were the result of non-neutral processes, as opposed to genetic drift due to small surviving population sizes. We predict that DFTD epidemics would result in swamping of prior genetic-environmental association with the abiotic environment.

## Methods

### *Trapping and sampling data*

Field data and ear biopsy samples from 3,286 Tasmanian devils were collected from seven different locations across their geographic range in Tasmania (Figure 1), pre and post-DFTD, between 1999 to 2014 (Table 1). Five of these locations became infected with DFTD during the study; one location remained disease-free, and one location was already infected at the beginning of the study (Figure 1, Table 1). We considered the first year of disease arrival as pre-disease in

our analyses, as the generation time for Tasmanian devils is approximately 2 years. Tasmanian devils were trapped using standard protocols (Hamede et al. 2015) involving custom-built polypropylene pipe traps 30 cm in diameter (Hawkins et al. 2006). These traps were set and baited with meat for ten consecutive trapping nights. In each 25 km<sup>2</sup> trapping site, at least 40 traps were set. Traps were checked daily, commencing at dawn. Each individual was permanently identified upon first capture with a microchip transponder (Allflex NZ Ltd, Palmerstone North, New Zealand). Additional specifics regarding field protocols, samples taken, and phenotypic data recorded can be found in Hamede et al. (2015).

#### *Overview of RAD-capture array development*

We used 90,000 Restriction-site Associated DNA sequencing (RAD-seq) loci generated from 430 individuals sampled across 39 sampling localities (Epstein et al. 2016; Hendricks et al. 2017) to develop a RAD-capture probe set (Ali et al. 2015). The details of the sample collection, preparation, and data processing of these original RAD loci is described in Epstein et al. (2016). Using these data, baits were designed to target a total of 15,898 RAD loci. This array included three categories of loci (with some overlap among categories): i) a set of 7,108 loci spread widely across the genome that were genotyped in more than half of the individuals, had  $\leq 3$  non-singleton SNPs, and had a minor allele frequency (MAF)  $\geq 0.05$ ; ii) 6,315 candidate loci based on immune function, restricted to non-singleton SNPs genotyped in  $\geq 1/3$  of the individuals, in and within 50 kb of an immune related gene; iii) 3,316 loci showing preliminary evidence of association with DFTD susceptibility with  $\leq 5$  non-singleton SNPs. Each RAD-capture locus was  $\geq 20$  kb away from other targeted loci to minimize potentially confounding effects of linkage

disequilibrium. Additional details of the creation of the RAD-capture probe set have been summarized in Margres et al. (2018a).

### *Data quality and filtering*

Libraries produced from the RAD-capture arrays were constructed using the *PstI* restriction enzyme and included 3,568 individuals from the seven distinct geographical locations across Tasmania (Figure 1). Libraries were then sequenced on a total of twelve lanes of an Illumina platform (5 lanes on NextSeq at the University of Oregon Genomics & Cell Characterization Core Facility; 7 lanes on HiSeq 4000 at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley). We de-multiplexed paired-end 150 bp reads, removed low quality reads and removed potential PCR/optical duplicates using the clone\_filter program (Stacks v1.21; Catchen et al. 2013). We then used Bowtie2 (Langmead & Salzberg 2012) to align reads to the reference genome (Murchison et al. 2012; downloaded from Ensembl May 2017) requiring the entire read align from one end to the other without any trimming (--end-to-end) with sensitive and -X 900 mapping options. With the resulting bam files, we created individual GVCF files using the option to emit all sites in the aligned regions with HaplotypeCaller from GATK (McKenna et al. 2010). All GVCF files were analyzed together using the option GenotypeGVCFs which re-genotyped and re-annotated the merged records. We selected SNPs using the SelectVariants option and filtered using the following parameters: QD < 2.0 (variant quality score normalized by allele depth), FS > 60.0 (estimated strand bias using Fisher's exact test), MQ < 40.0 (root mean square of the mapping quality of reads across all samples), MQRankSum < -12.5 (rank sum test for mapping qualities of reference versus



alternative reads) and  $\text{ReadPosRankSum} < -8$  (rank sum test for relative positioning of reference versus alternative alleles within reads). Using VCFtools (Danecek et al. 2011), additional filtering was performed to remove SNPs on the X chromosome, non-biallelic SNPs, indels, those with a minor allele frequency  $< 0.05$ , and those missing data at more than 50% of the individuals genotyped and 40% of the sites across all individuals. To parse out the genetic-environmental associations of abiotic factors from disease, samples were divided into pre and post-disease subsets prior to analyses. After filtering, we retained 3,286 Tasmanian devils (1521 before and 1765 after DFTD) and 6,886 SNPs (Table 1).

### *Analyses of population structure*

To examine the underlying population structure in our dataset, both pre- and post-disease, we used fastSTRUCTURE (Alexander 2009). fastSTRUCTURE is an algorithm that estimates ancestry proportions using a variational Bayesian framework to infer population structure and assign individuals to genetic clusters. We ran both the pre- and post-disease data sets with the number of genetic clusters (K) set from 1 to 18. Genetic relationships amongst sampled populations were also evaluated using discriminant analysis of principal components (DAPC) analysis (Jombart et al. 2010) implemented in the Adegenet package in R (Jombart et al. 2008). This method transforms genotypic data into principal components and then uses discriminant analysis to maximize between group genetic variation and minimize within group variation. To determine the optimal number of genetic clusters, we used the k-means clustering algorithm for increasing values of K from 1 to 20. We selected the optimal K value by selecting the visual “elbow” in the Bayesian information criterion (BIC) score, which is the lowest BIC value that also minimizes the number of components or genetic clusters retained.

193

# 194 *Genetic diversity*

195 To identify signals of a potential genetic bottleneck, we tested for changes in genetic diversity  
 196 pre- and post-disease emergence. We used VCFtools (Danecek et al. 2011) to calculate Weir and  
 197 Cockerham's estimator of  $F_{ST}$  ( $\Theta$ ; Weir and Cockerham 1984) for all pairwise comparisons and  
 198 all pre-post disease population comparisons. To gauge whether populations had significantly  
 199 different  $F_{ST}$  values, we bootstrapped 95% confidence intervals for 10,000 iterations and tested  
 200 whether the confidence intervals bounded zero. Additionally, we calculated  $\pi$  (Nei and Li 1979)  
 201 and Tajima's D (Tajima 1989) for each population before and after disease arrival to estimate  
 202 standing levels of genetic variation (Danecek et al. 2011). Each metric was calculated by taking  
 203 the average of each non-overlapping windows of 10 kilobases (kb) per population pre and post-  
 204 disease. Finally, we tested for changes in inbreeding ( $F_{IS}$ ) using the R package Adegnet  
 205 (Jombart et al. 2008) for each population pre and post-disease.

206

# 207 *Abiotic environmental variables*

208 We used ArcGIS 10 (ESRI 2011) to plot the location of every sampled Tasmanian devil. Over the  
 209 course of the 15 years (1999-2014; Table 1), there was some variation in the shape of the  
 210 trapping grids at the sampling sites due to changes in land use and permissions. To account for  
 211 this variation, we located the centroid of each sampling area using the calculate geometry tool in  
 212 ArcGIS 10 (ESRI 2011). We then drew a 25 km<sup>2</sup> ellipse around each centroid to represent the  
 213 total trapping area for each sampling location and extracted environmental data for each location.  
 214 Although the Freycinet and Forestier sites were larger than other sampling sites, the centroids

were in the middle of the peninsulas likely making them representatives of those sites. The data layers containing these environmental data were obtained from [www.ga.gov.au](http://www.ga.gov.au), [www.worldclim.org](http://www.worldclim.org), and [www.thelist.tas.gov.au](http://www.thelist.tas.gov.au) (see Supplementary Table 1 for a list of the 18 abiotic environmental variables analyzed).

We extracted the values for each environmental variable at five randomly chosen locations within the ellipse generated for each trapping area to test whether there was significant heterogeneity for any of the environmental variables across the trapping area of any particular site. Using the calculate statistics tools in ArcGIS 10, we calculated the mean, standard deviation, and 95% confidence intervals for each of the environmental variables within each site. If the centroid fell within the 95% confidence interval of the randomly selected points at each site, we deemed the centroid as an adequate representation for that environmental variable site-wide (Supplementary Table 2).

To estimate collinearity among environmental variables (Supplementary Table 1), we conducted a principal component analysis (PCA) using the prcomp package in R (R Core Team 2013) including the centroid values for each of our seven sampling localities. The top two abiotic environmental variables that explained the greatest proportion of the variance in each of the first six principal components in the PCA were used as explanatory variables in the subsequent GEA analyses. We did not include the seventh principal component as the amount of variation in the data explained by this component was negligible ( $< 0.001$ ). To ensure that environmental variables were not correlated, we conducted paired correlative tests using the Spearman's  $\rho$  statistic.

### *Biotic environmental variables*

To estimate disease prevalence within populations, we used trapping records compiled in a database provided by collaborators at the University of Tasmania and the Department of Primary Industries, Water and Environment from 1999-2014 from the six infected populations (McCallum et al. 2009; Hamede et al. 2009, 2012, 2013, 2015, unpubl; Lachish et al. 2010). Field records contained data regarding every unique trapping encounter as well as phenotypic data for each devil. As there is currently no early diagnostic test for DFTD, likelihood of a single devil being infected with DFTD is recorded for devils in the field using a scale of 1-5 (Hawkins et al. 2006). We calculated disease prevalence using the total number of unique devils trapped with a visual DFTD score  $\geq 3$  (indicative of a growth with high probability of being a DFTD tumor) divided by the number of unique devils captured alive that year. In our GEAs, we averaged the annual disease prevalence values across years 2-4 after DFTD was confirmed at a particular sampling location. We selected these years as the trapping and monitoring efforts across these field sites instead of using the first year of infection of DFTD at a particular location as there was variation in estimates of prevalence during the first year of DFTD detection and no guarantee that this truly was the first year of infection.

### *Genetic-environmental association (GEA) analyses*

We tested correlations of allele frequencies with abiotic and biotic environmental variables across the seven study locations using latent factor mixed models (LEA; Frichot et al. 2015) and Bayenv2 (Gunther and Coop 2013). The R package for landscape and ecological association studies (LEAs) uses latent factors, analogous to principal components in a PCA, to account for

background population structure (Frichot et al 2013). These latent factors serve as random effects in a linear model that tests for correlations between environmental variables and genetic variation (Frichot et al 2013). The number of populations sampled was used for the number of latent factors in LEA. Bayenv2 uses allelic data to generate a variance-covariance matrix, or kinship matrix, to estimate the neutral or null model for underlying demographic structure (Coop et al. 2010; Gunther and Coop 2013). Spearman's  $\rho$  statistics were then calculated to provide non-parametric rankings of the strength of the association of each genotype with each environmental variable compared to the null distribution described by the kinship matrix alone (Gunther and Coop 2013). The non-parametric Spearman's rank correlation coefficients have been shown to be more powerful in describing genetic-environmental relationships if there are extreme outliers (Whitlock and Lotterhos 2015; Rellstab et al. 2015).

### *Candidate gene identification*

Although each of the landscape genomic methods listed above detect GEAs, each program has been shown to vary in true-positive detection rate given sampling scheme, underlying demography, and population structure (Hoban et al. 2016; Lotterhos and Whitlock 2014; 2015; Rellstab et al. 2015). To reduce the false-positive rate without sacrificing statistical power, we used MINOTAUR (Lotterhos et al., 2017; Verity et al. 2017) to identify putative loci under selection. MINOTAUR takes the test statistic output from each of the GEAs and identifies the top outliers in multi-dimensional space; here, we used the Mahalanobis distance metric to ordinate our outliers. We selected this distance metric as it has been shown to have high statistical power on simulated genomic data sets (Verity et al. 2017) and because our data followed a parametric

distribution. We chose our final set of candidate SNPs by selecting the top 1% of loci that had the largest Mahalanobis distance for each environmental variable. Thus, we performed seventeen separate MINOTAUR runs: eight for the eight abiotic environmental variables tested pre-DFTD arrival and nine for the eight abiotic and single biotic environmental variable we tested post-DFTD arrival. We performed seventeen MINOTAUR runs to select the top 1% of loci that had the greatest support for associations with each environmental variable across both methods pre and post-disease.

We then used bedtools (Quinlan and Hall 2010) to identify genes within 100 kb of each candidate SNP in the devil reference genome (Murchison et al. 2010). We chose 100 kb as it has been shown to be the conservative estimate of the size of linkage groups in Tasmanian devils (Epstein et al. 2016). Protein-coding genes within these windows were included in the list of candidate genes. Gene annotations were retrieved from the ENSEMBL database (Akey et al. 2016), gene IDs were derived from the NCBI GenBank database (Wheeler et al. 2017), and descriptions of putative function and gene ontologies were gathered from [www.genecards.org](http://www.genecards.org) (Fishilevich et al. 2017) and the Gene Ontology (GO) Consortium (Ashburner et al. 2000). We then tested for enrichment of GO terms in our candidate gene list using the R package SNP2GO (Szkiba et al. 2014). SNP2GO tests for over-representation of GO terms associated with genes within a specified region using Fisher's exact test and corrects for multiple testing using the Benjamini-Hochberg (1995) false-discovery rate (FDR). We ran this program on the pre- and post-disease candidate sets separately and used all genes within 100 kb of the original RAD-capture data set as our reference set. Enriched terms were those with an  $FDR \leq 0.05$ .

### *Fisher's exact test*

To determine if GEAs detected pre-DFTD remained post-DFTD arrival, we calculated the change in the MINOTAUR rank of each candidate locus following disease arrival. We ranked loci by Mahalanobis distance and compared the rank of the pre-DFTD candidate loci to post-disease arrival for each of the eight abiotic environmental variables. To determine whether a greater proportion of loci detected pre-disease had reduced MINOTAUR rankings post-disease arrival than we would expect by random chance, we conducted Fisher's exact tests using an  $\alpha=0.05$ . If we detected a significantly higher proportion of models in which the MINOTAUR rankings were lower post-disease than pre-disease, then we considered the molecular signal of the genetic-environmental to be swamped by disease.

## **Results**

### *Population structure and genetic diversity*

Our analyses generally supported  $K=6$  pre and post-DFTD arrival reflecting the six populations sampled during these time periods. However, there was uncertainty of the optimal  $K$ -value in fastSTRUCTURE. Using the "chooseK" python script recommended in fastSTRUCTURE,  $K=9$  (Supplementary Figure 1a) received the greatest support pre-DFTD and  $K=8$  post-DFTD (Supplementary Figure 1b). However, in both cases, the difference in the marginal likelihood between  $K=4$  and the selected  $K$  was very small (pre-DFTD  $\Delta$ Marginal Likelihood=0.003, post-DFTD  $\Delta$ Marginal Likelihood=0.0032; Supplementary Figure 1c,d). For visual comparison, we also plotted  $K=6$  pre-DFTD (Figure 2a) and post-DFTD (Figure 2b), which reflects the number of populations sampled. The  $K$  value for DAPC that minimized the BIC, as well as the number of

components included, was  $K=6$  both pre-DFTD (Figure 2c) and post-DFTD (Figure 2d). Pairwise  $F_{ST}$  values produced using Weir and Cockerham's estimator would also support  $K=6$  as all bootstrapped confidence intervals for population comparisons did not bound zero (Table 2). We found that there were no significant differences between  $\pi$  (Supplementary Table 3), Tajima's  $D$  (Supplementary Table 3), Nei's  $F_{IS}$  (Supplementary Table 3) or Weir and Cockerham's  $F_{ST}$  values for any of the populations after DFTD arrived, suggesting there were no significant changes in genetic diversity following disease arrival.

### *Environmental variables*

We initially considered 18 abiotic environmental variables (Supplementary Table 1) that may be relevant to the distribution of genetic variation across the devil's geographic range. Eight of these 18 variables were not correlated (Table 3) and explained a significant proportion ( $> 0.99$ ) of the variance in the environmental data as summarized in the top principal components (Supplementary Table 4). There were no statistically significant differences between the centroid values or any of the five randomly selected points for any of the environmental variables within each of the sampling locations (Supplementary Table 2), indicating a lack of significant within-site heterogeneity. These eight abiotic environmental variables were subsequently used in the GEAs.

### *Landscape genomics analyses*

Combining GEAs across all abiotic variables, we identified 349 unique loci pre-DFTD and 397 unique loci post-DFTD that received the greatest support from MINOTAUR as candidate loci.



We identified candidate genes based on linkage disequilibrium (within 100 kb) to the unique loci, which resulted in 64 genes pre-disease (Supplementary Table 5) and 83 genes post-disease (Supplementary Table 6); seven of those genes overlapped temporally (Supplementary Figure 2, Supplementary Table 5). There were 58 additional unique loci detected by MINOTAUR post-DFTD arrival as related to disease prevalence in the GEAs. Within 100 kb of these 58 SNPs, there were seven annotated genes, none of which overlapped temporally with the pre-disease candidate set (Supplementary Table 6). For candidate genes in which there was more than a single association with a particular environmental variable, the variable that had the greatest absolute Pearson's correlate value was selected to be used in these candidate gene-environment associations.

#### *Candidate gene identification for selection by the abiotic environment*

Mean annual temperature and annual temperature range were the two most common variables significantly correlated with the allele frequencies of candidate SNPs amongst devil populations pre-disease arrival (Supplementary Table 7). Twenty-nine of the top 64 candidate genes pre-disease were associated with at least one of these two environmental variables and broadly associated with response to external stress and ion transport (Supplementary Tables 5, 7). Specifically, genes associated with annual temperature range had GOs including RNA transport and processing, and gene pathway regulation. Mean annual temperature, in contrast, was correlated with genes with intracellular protein transport, protein and lipid binding, and cell adhesion GOs. Additionally, isothermality was frequently correlated with significant differences in allele frequency of candidate loci among populations and have similar GOs as annual

temperature range. Elevation and vegetation index explained a large proportion of the observed variance in the abiotic environment across our seven sampling sites (Supplementary Table 2). Elevation was correlated with seven of the candidate genes with ion binding and oxidoreductase activity functions (Supplementary Tables 5, 7). Vegetation index was correlated with allele frequencies of thirteen of the candidate genes that had GOs including transcription and cellular protein binding.

In our abiotic GEAs post-disease arrival annual temperature range, mean annual temperature and vegetation index were the top environmental variables most frequently associated with candidate SNPs (Supplementary Tables 6, 8). Similar to pre-DFTD, GOs for genes associated with these two temperature variables included cellular response to stimulus, RNA processing and regulation, and ion binding. In contrast to pre-DFTD, mean annual temperature and annual temperature range were associated with genes with cell signaling and apoptotic processes. Vegetation index was associated with fourteen of the candidate genes post-DFTD with GOs including transcription and cell signaling pathways (Supplementary Tables 6, 8).

There was no statistically significant enrichment of any GO category of our candidates (pre-disease  $p=0.344$ ; post-disease  $p=0.297$ ), there were commonalities in the putative functions of candidate genes pre- versus post-disease including regulation of transcription and translation and cellular response to external stress (Supplementary Tables 5, 6).

*Candidate gene identification for selection by the biotic environment*

Disease prevalence was significantly correlated with divergent allele frequencies among sampled devil populations (Supplementary Table 8, 9). Fourteen of the 64 candidate genes detected post-DFTD arrival were associated with disease-prevalence (Supplementary Tables 8, 9). The associated GOs for these fourteen genes included tumor regression, apoptosis, and immune response (Supplementary Table 6). Of these fourteen genes, six genes were also associated with abiotic environmental variables. Genes *TBXAS1* and *FGGY* were found in both the pre and post-DFTD candidate gene lists and associated with GOs including oxidoreductase activity and metabolism (Supplementary Table 5).

#### *Fisher's exact test for disease swamping*

Putative GEAs found pre-DFTD were consistently undetected post-DFTD arrival (Supplementary Tables 5,6). Using Fisher's exact tests, fewer pre-DFTD candidate loci were detected post-DFTD arrival than expected by random chance for each MINOTAUR analysis for each of the eight abiotic environmental variables (Figure 3, Supplementary Figure 3). For example, we only detected three of the putative pre-DFTD candidate loci using MINOTAUR post-DFTD arrival for genetic associations with mean annual temperature across sampled populations (Fisher's Exact Test Odd's ratio=23.211,  $p < 0.001$ ; Figure 3).

Seven of the original 64 candidate genes detected pre-disease overlapped with the 83 candidate genes post-disease arrival (Supplementary Table 5). Four were associated with at least one of the same environmental variables post-disease arrival as pre-disease, and three were correlated with different variables (Supplementary Table 9). Two of the variants linked with *TBXAS1* and *FGGY* were found to be associated with disease prevalence post-disease arrival. For

example, the *FGGY* gene was originally associated with mean annual temperature prior to disease arrival, but disease prevalence post-arrival. Putative functions of *FGGY* include carbohydrate phosphorylation and neural cell homeostasis (Singh et al. 2017; Dunckley et al. 2007; Supplementary Table 5). Similarly, *TBXAS1*, which is putatively involved in oxidoreductase activity (Ullrich et al. 1994), was associated with mean annual temperature pre-disease arrival, but was more strongly correlated with disease prevalence following disease arrival (Supplementary Tables 5, 9).

## Discussion

Here, we showed Tasmanian devils are more genetically structured across Tasmania than previously documented (Bruniche-Olsen et al. 2014; Hendricks et al. 2017; Miller et al. 2011) thereby creating conditions that favor local adaptation. Indeed, GEAs showed signatures of selection across populations, suggestive of adaptation to local abiotic factors prior to disease arrival. Mean annual temperature and annual temperature range were abiotic variables most frequently associated with candidate genes (Loyau et al. 2016; Tseng et al. 2011; Xu et al. 2017), and these abiotic factors have been established as important for devil habitat use (Jones and Barmuta 2000; Jones and Rose 1996). Isothermality and vegetation index were also associated with candidate genes and correlated with the largest amount of geographic heterogeneity across the landscape (Schweizer et al. 2016; Mukherjee et al. 2018). Nonetheless, most of the candidate genes detected prior to disease arrival were not detected post-DFTD emergence reflecting the strong selection imposed by DFTD. Instead, GEAs showed evidence of significant correlations of disease prevalence with allele frequencies at several SNPs after disease arrival. Taken together,

these results suggest that DFTD swamps molecular signatures of local adaptation to abiotic variables. Finally, despite large declines in census population size across all diseased populations, no significant changes in genetic diversity were detected, suggesting detected patterns were not likely the result of genetic drift.

### *Genetic population structure*

We found conflicting results among our demographic analyses of our sampled populations. Similar to previous studies, we detected admixture between some of the seven sampling locations using fastSTRUCTURE (Figure 2a), which showed an optimal K value between 4-9. However, using DAPC (Figure 2b), we detected distinct genetic clusters among the sampling locations, irrespective of disease presence. Pairwise population  $F_{ST}$  calculations were all significantly greater than 0, indicating significant population differentiation between all populations sampled (Table 2). Regardless of the analysis, we identified greater amounts of genetic structure than previously detected. Two previous studies found high levels of admixture between two genetic clusters separating eastern and western Tasmania (Brüniche-Olsen 2014; Hendricks et al. 2017), while another, based on mtDNA genomes, suggested three genetic clusters (Northwest, East and Central; Miller et al. 2011). Our results differed from the previous work because we included much larger sample sizes and numbers of loci, which likely increased our power to detect population structure.

### *Genetic diversity*

Despite extensive, range-wide population declines, we found no evidence for loss of genetic diversity following disease arrival. Devil populations had low levels of  $\pi$  across the species' geographic range, regardless of whether DFTD was present (Supplementary Table 6), consistent with findings from previous studies (Bruinche-Olsen et al. 2014; Hendricks et al. 2017; Jones et al. 2004). While there was population genetic differentiation ( $F_{ST}$ ) between sampled populations, there was no significant difference in inbreeding coefficient ( $F_{IS}$ ) among populations pre- versus post-disease arrival. Additionally, there was no significant change in genetic differentiation ( $F_{ST}$ ) temporally among populations post-DFTD arrival. Lack of a significant change in the positive Tajima's D values across populations following disease arrival (Supplementary Table 3) also suggests possible maintenance of some levels of standing genetic variation even after substantial population declines. One possibility for the lack of detectable changes in genetic diversity would be that simply not enough time has passed since the bottleneck. If we take account of the increased precocial breeding in diseased populations, which reduces the generation time of females from 2-3 years prior to disease outbreak to 1.5-2 years (Jones et al. 2008) the maximum number of generations in our data set since disease arrival would be in Freycinet for 8-9 generations (number of generations passed for all populations in Table 1). Detection of significant changes in genetic diversity requires substantial reduction in effective population size for several generations (Luikart et al. 2010). Following DFTD outbreak, devil populations typically decline by 90% after 5-6 years. This time lag, coupled with the short timescale, perhaps resulted in low power to detect changes in genetic diversity.

*Landscape genomics pre-disease*

The abiotic environmental variables used in our GEAs have been shown to be important determinants of devil distribution (Jones and Barmuta 2000; Jones and Rose 1996; Rounsevell et al. 1991). Devils are distributed throughout the diverse habitats of Tasmania, but their core distribution comprises areas of low to medium rainfall which are dominated by dry, open forest, woodland and coastal shrubland. They use varying vegetation types across the environment for different functions. Devils prefer a clear understory for movement, forest-grassland edges for hunting, and generally avoid structurally complex vegetation and landscape features such as rocky areas and steep slopes (Jones and Barmuta 2000; Jones and Rose 1996). Candidate loci linked to the CUX1 gene were strongly associated with vegetation index in devils, especially in heavily forested populations such as WPP (Supplementary Table 7). CUX1, which is putatively involved in transcription regulation (Sansregret 2008) and limb development in morphogenesis (Lizarraga et al. 2002), has been detected as a candidate in the mammalian systems for adaptation to inhabiting environments with complex understory (Schweizer et al. 2016; Mukherjee et al. 2018). Landscape genomic analyses of grey wolves found evidence for the involvement of this gene, and others related to auditory function and morphogenesis, in navigating between dense and more open landscape habitat types (Schweizer et al. 2016). Differential expression of this gene in muscle tissue of the bovine species *Bos frontalis* relative to domestic cattle was suggested to assist in reducing energetics required for navigating complex, hilly environments in India (Mukherjee et al. 2018). Strong genetic-environmental association across devil populations throughout their heterogeneous geographic range provide evidence for local adaptive genetic variation.

Although we found no significant enrichment for any GO category in the 64 genes identified in strong LD with the top 1% of SNPs pre-DFTD, 29 of the genes were putatively involved in stress response and ion binding. Climatic variables including mean annual temperature, annual temperature range, and isothermality were all important in describing observed genetic variation across the devil geographic range. The *RPL7L1* gene, for example, which functions in mRNA transcription and translation during mammalian cellular development and differentiation (Maserati et al. 2012), was strongly associated with annual temperature range. In Chinese Sanhe cattle, *RPL7L1* was found to be down-regulated when exposed to severe cold, reflective of an adaptive response to rapid temperature variation (Xu et al. 2017). In devils the allele frequencies of candidate loci linked to this gene appear to be strongly associated with populations occupying the narrowest temperature ranges in devils' geographic range including NAR, WPP and FRY (Supplementary Table 1). Expressional changes of this gene may therefore be a crucial local adaptation to devils living in these habitats when temperature fluctuates outside the normal range. Additionally, mean annual temperature was significantly associated with genes such as *CYP39A1*, which putatively functions in oxidoreductase activity of steroids and lipid metabolism (Li-Hawkins et al. 2000). This gene is differentially expressed in response to long-term temperature variation, for example, it is downregulated in chicken embryos following extended heat exposure (Loyau et al. 2016) and upregulated in zebrafish brains in response to severe cold (Tseng et al. 2011). Overall, the functions of genes associated with the abiotic variables would suggest specific, local responses to some key environmental conditions that explain devil geographic distributions.



## *Landscape genomics pre- vs post-disease*

Following the arrival of DFTD, both the number of loci and proportion of variation in the allele frequencies of the pre-disease candidate loci explained by the abiotic environment were significantly reduced. The *TMCC2* gene, for example, was strongly correlated with mean annual temperature prior to disease arrival (pre-disease, Pearson's correlate=0.915, MINOTAUR rank=39), but the strength of the genetic-environmental association significantly decreased post-disease arrival and it was no longer detected as a candidate locus (post-disease, Pearson's correlate=-0.3, MINOTAUR rank=6071). Only seven of the 64 candidate genes detected pre-disease were included in the 83 genes detected post-disease set, with only one gene, *ARPC2* (Actin related protein 2/3 Complex Subunit 2), correlated with the same variable prior to disease arrival. *FGGY*, for example, was strongly correlated with mean annual temperature prior to disease arrival (pre-disease, Pearson's correlate=-0.745, MINOTAUR rank=018). Post-disease arrival, the strength of this relationship significantly decreased (post-disease, Pearson's correlate=-0.082, MINOTAUR rank=6199), and the SNP linked to this gene was more strongly correlated with disease prevalence (post-disease, Pearson's correlate=-0.082, MINOTAUR rank=6199). These discordant patterns of association of candidate loci with environmental variables pre compared to post-disease arrival may possibly be explained by pleiotropy or loss of association due to DFTD.

We detected 83 candidate genes post-disease arrival associated with both abiotic environmental variables and/or disease prevalence. Although there was no significant enrichment of any GO category, most genes associated with abiotic variables had stress response GOs, whereas post-disease many genes had GOs involved in cellular processes including apoptosis,

cell differentiation and cell development, similar to what has been previously detected (Epstein et al. 2016; Margres et al. 2018a; Frampton et al. 2018). Genes involved in apoptosis detected in our GEAs, including *DNAJA3* in the heat shock protein family, have also been previously identified in DFTD literature as candidates for anti-cancer vaccines (Tovar et al. 2018) for their putative involvement in immunogenicity (Graner et al. 2000) and tumor suppression (Shinagawa et al. 2008). Functionally similar to *PAX3* a gene associated with apoptosis and angiogenesis (Asher et al. 1996) and devil tumor regression detected in previous studies (Wright et al. 2017), we detected genes including *FLCN* and *BMPER* post-DFTD arrival (Qi et al. 2009). We also identified *MGLL* and *TLR6* in our post-DFTD candidate list which are involved in inflammatory response (Epstein et al 2016; Margres et al. 2018a).

The identification of candidate genes putatively involved in immune response (e.g. *RFN126* and *IL9R*), tumor suppression (e.g. *DMBT1*, *FLCN* and *ITFG1*), signal transduction (e.g. *ARHGEF37* and *PPP1R12B*) and cell-cycle regulation (e.g. *TK2*) associated with disease prevalence post-DFTD arrival provide evidence that disease may have had a strong influence on standing genetic variation among devil populations (Tsai et al. 2014; Purwar et al. 2012; Mollenhauer et al. 1997; Hasumi et al. 2015; Leushacke et al. 2011; Tsapogas et al. 2003; Bannert et al. 2003; Sun, Eriksson & Wang et al. 2014).

The lack of GO enrichment may be an artifact of the RAD-capture panel. Loci targeted herein were linked to or found in coding regions of both putatively neutral genes as well as those involved in immune response, potentially biasing our GO enrichment analysis. Another explanation for the lack of enrichment could stem from our conservative approach for identifying outliers. Using MINOTAUR, we took the top 1% of loci from each GEA as our putative

candidates. Although this approach reduces false positives, it also can reduce true positives and imposes an upper bound on the number of possible associations we can ascertain between our candidate loci and disease.

### *Support for disease swamping abiotic genetic-environmental associations*

Intense, long-term monitoring of Tasmanian devils coupled with an expansive temporal and geographic dataset provided the unique opportunity to test for changes in statistical signatures of GEAs. We hypothesized that DFTD would serve as an extreme selective event as it is nearly 100% lethal (Hamede et al. 2015) and produces a rapid adaptive response (Epstein et al 2016; Wright et al. 2017; Margres et al. 2018a,b) that could swamp signals of genetic-environmental associations with the abiotic environment found pre-disease. Our comparisons of candidate loci detected pre- and post-disease support this hypothesis. Although annual temperature range and vegetation index continued to be important drivers of selection in devil populations regardless of disease presence, most candidate genes identified pre-DFTD were not correlated with these variables after DFTD arrived. (Figure 3; Supplementary Figure 3).

The observed loss of pre-DFTD GEAs following disease arrival could also be the result of stochastic processes operating on our populations following a significant demographic event (Lande 1976; Lande 1993; Shaffer 1981). Following large catastrophes, locally-adapted genotypes may be displaced or swamped out randomly by new genetic variation from neighboring populations via genetic rescue (Waddington 1974; Brown and Kodric-Brown 1977). If this were the case, our observation of loss of pre-DFTD candidate genes may be due to drift resulting from demographic change induced by DFTD versus the selection that DFTD imposed.

However, numerous studies showing evidence of the rapid evolutionary response to DFTD across devil populations suggest that DFTD is selecting for particular genetic variants post-DFTD arrival (summarized in Russell et al. 2018 & Storfer et al. 2018). Additionally, in spite of large decreases in population size, there have been no detectable significant changes to genetic diversity following disease arrival.

Traditional landscape genomics studies incorporating biotic variables have successfully identified signatures of adaptation to biotic variables, such as predators (Ritcher-Boix et al. 2011), parasites (Orsini et al. 2013; Wenzel et al. 2016) and symbionts (Harrison et al. 2017). However, few have successfully identified GEAs with infectious disease using the landscape genomics framework (reviewed in Kozakiewicz et al. 2018). In this study we identified genes putatively associated with an emerging infectious disease and quantified how the onset of disease reduced evidence of GEAs with the abiotic environment among populations of Tasmanian devils across their geographic range. Here, we document a novel, biotic variable swamping out molecular signals of association to the abiotic environment. The findings of this study demonstrate the utility of landscape genomics as a tool to explicitly test the influence of biotic factors, such as disease, on the spatial distribution of genetic variation.

## Conclusion

Emerging infectious diseases are increasingly recognized as significant threats to biodiversity and in extreme cases, species' range contractions and extinctions (Smith et al. 2006). Yet, few landscape genomics studies have tested for statistical correlations of allele frequencies with biotic variables such as infectious diseases. Here, we show that a lethal disease of Tasmanian devils

608    apparently swamps associations of abiotic variables with allele frequencies in devil populations  
609    detected before disease arrival. This finding is consistent with previous work that showed rapid  
610    evolution of Tasmanian devils in response to disease (Epstein et al. 2016; Wright et al. 2017;  
611    Margres et al. 2018a,b). Additionally, no appreciable declines in genetic diversity were detected  
612    across multiple analysis methods. Taken together, these results suggest that observed patterns of  
613    allele frequency correlations with disease prevalence are more likely attributable to selection than  
614    genetic drift. Our approach and results suggest that future landscape genomics studies may prove  
615    valuable for identifying loci under selection pre- and post-disease.

## Literature Cited

- Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*, 12(12), 1805-1814. doi:10.1101/gr.631202
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655-1664. doi: 10.1101/gr.094052.109
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*, 202(2), 389-400. doi:10.1534/genetics.115.183665
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Eppig, J. T. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25. doi: 10.1038/75556
- Asher Jr, J. H., Sommer, A., Morell, R., & Friedman, T. B. (1996). Missense mutation in the paired domain of PAX3 causes craniofacial-deafness-hand syndrome. *Human mutation*, 7(1), 30-35. doi: 10.1002/(SICI)1098-1004(1996)7:1<30::AID-HUMU4>3.0.CO;2-T
- Bannert, N., Vollhardt, K., Asomuddinov, B., Haag, M., König, H., Norley, S., & Kurth, R. (2003). PDZ Domain-mediated interaction of interleukin-16 precursor proteins with myosin phosphatase targeting subunits. *Journal of Biological Chemistry*, 278(43), 42190-42199. doi: 10.1074/jbc.M306669200
- Beall, C. M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci U S A*, 104 Suppl 1, 8655-8660. doi:10.1073/pnas.0701985104

637 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and  
638 powerful approach to multiple testing. *Journal of the Royal statistical society: series B*  
639 (Methodological), 57(1), 289-300. doi: 10.1111/j.2517-6161.1995.tb02031.x

640 Biek, R., & Real, L. A. (2010). The landscape genetics of infectious disease emergence and  
641 spread. *Mol Ecol*, 19(17), 3515-3531. doi:10.1111/j.1365-294X.2010.04679.x

642 Blanchong, J. A., Samuel, M. D., Scribner, K. T., Weckworth, B. V., Langenberg, J. A., &  
643 Filcek, K. B. (2007). Landscape genetics and the spatial distribution of chronic wasting disease.  
644 *Biology Letters*, 4(1), 130-133. doi: 10.1098/rsbl.2007.0523

645 Brown, J. H., & Kodric-Brown, A. (1977). Turnover rates in insular biogeography: effect of  
646 immigration on extinction. *Ecology*, 58(2), 445-449. doi: 10.2307/1935620

647 Brüniche-Olsen, A., Jones, M. E., Austin, J. J., Burrridge, C. P., & Holland, B. R. (2014).  
648 Extensive population decline in the Tasmanian devil predates European settlement and devil  
649 facial tumour disease. *Biology Letters*, 10(11), 20140619. doi: 10.1098/rsbl.2014.0619

650 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an  
651 analysis tool set for population genomics. *Molecular ecology*, 22(11), 3124-3140. doi:  
652 10.1111/mec.12354

653 Chattopadhyay, S., Whitehurst, C. E., & Chen, J. (1998). A nuclear matrix attachment region  
654 upstream of the T cell receptor  $\beta$  gene enhancer binds Cux/CDP and SATB1 and modulates  
655 enhancer-dependent reporter gene expression but not endogenous gene expression. *Journal of*  
656 *Biological Chemistry*, 273(45), 29838-29846. doi: 10.1074/jbc.273.45.29838

657 Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental  
658 correlations to identify loci underlying local adaptation. *Genetics*, 185(4), 1411-1423.  
659 doi:10.1534/genetics.110.114819

660 Dal Grande, F., Rolshausen, G., Divakar, P. K., Crespo, A., Otte, J., Schleuning, M., & Schmitt,  
661 I. (2018). Environment and host identity structure communities of green algal symbionts in  
662 lichens. *New Phytologist*, 217(1), 277-289. doi: 10.1111/nph.14770

663 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T.  
664 (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. doi:  
665 10.1093/bioinformatics/btr330

666 Dunckley, T., Huentelman, M. J., Craig, D. W., Pearson, J. V., Szelinger, S., Joshipura, K., . . .  
667 Letizia, D. (2007). Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *New*  
668 *England journal of medicine*, 357(8), 775-788. doi: 10.1056/NEJMoa070174

669 Eoche-Bosy, D., Gautier, M., Esquibet, M., Legeai, F., Bretaudeau, A., Bouchez, O., . . .  
670 Montarry, J. (2017). Genome scans on experimentally evolved populations reveal candidate  
671 regions for adaptation to plant resistance in the potato cyst nematode *Globodera pallida*.  
672 *Molecular ecology*, 26(18), 4700-4711. doi: 10.1111/mec.14240

673 Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E. P., . . . Storfer,  
674 A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nat*  
675 *Commun*, 7, 12684. doi:10.1038/ncomms12684

676 Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., . . . Safran, M.  
677 (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards.  
678 Database, 2017. doi: 10.1093/database/bax028



679 Frampton, D., Schwenzer, H., Marino, G., Butcher, L. M., Pollara, G., Kriston-Vizi, J., . . .  
680 Fassati, A. (2018). Molecular Signatures of Regression of the Canine Transmissible Venereal  
681 Tumor. *Cancer Cell*, 33(4), 620-633 e626. doi:10.1016/j.ccell.2018.03.003  
682 Francois, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in  
683 genome scans for selection. *Mol Ecol*, 25(2), 454-469. doi:10.1111/mec.13513  
684 Frichot, E., Schoville, S. D., Bouchard, G., & Francois, O. (2013). Testing for associations  
685 between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*,  
686 30(7), 1687-1699. doi:10.1093/molbev/mst063  
687 Frichot, E., Schoville, S. D., de Villemereuil, P., Gaggiotti, O. E., & Francois, O. (2015).  
688 Detecting adaptive evolution based on association with ecological gradients: orientation matters!  
689 *Heredity (Edinb)*, 115(1), 22-28. doi:10.1038/hdy.2015.7  
690 Graner, M., Raymond, A., Akporiaye, E., & Katsanis, E. (2000). Tumor-derived multiple  
691 chaperone enrichment by free-solution isoelectric focusing yields potent antitumor vaccines.  
692 *Cancer Immunology, Immunotherapy*, 49(9), 476-484. doi: 10.1007/s002620000138  
693 Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies.  
694 *Genetics*, 195(1), 205-220. doi:10.1534/genetics.113.152462  
695 Haasl, R. J., & Payseur, B. A. (2016). Fifteen years of genomewide scans for selection: trends,  
696 lessons and unaddressed genetic sources of complication. *Mol Ecol*, 25(1), 5-23.  
697 doi:10.1111/mec.13339  
698 Hamede, R. K., Mccallum, H., & Jones, M. (2008). Seasonal, demographic and density-related  
699 patterns of contact between Tasmanian devils (*Sarcophilus harrisii*): Implications for

transmission of devil facial tumour disease. *Austral Ecology*, 33(5), 614-622. doi:  
10.1111/j.1442-9993.2007.01827.x

Hamede, R. K., 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 devil facial  
tumour disease. *Ecology letters*, 12(11), 1147-1157. doi: 10.1111/j.1461-0248.2009.01370.x

Hamede, R. K., McCallum, H., & Jones, M. (2013). Biting injuries and transmission of  
Tasmanian devil facial tumour disease. *Journal of Animal Ecology*, 82(1), 182-190. doi:  
10.1111/j.1365-2656.2012.02025.x

Hamede, R. K., Pearse, A.-M., Swift, K., Barmuta, L. A., Murchison, E. P., & Jones, M. E.  
(2015). Transmissible cancer in Tasmanian devils: localized lineage replacement and host  
population response. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814),  
20151468. doi: 10.1098/rspb.2015.1468

Hamede, R., Lachish, S., Belov, K., Woods, G., Kreiss, A., PEARSE, A. M., . . . McCALLUM,  
H. (2012). Reduced effect of Tasmanian devil facial tumor disease at the disease front.  
*Conservation Biology*, 26(1), 124-134. doi: 10.1111/j.1523-1739.2011.01747.x

Harrison, T. L., Wood, C. W., Borges, I. L., & Stinchcombe, J. R. (2017). No evidence for  
adaptation to local rhizobial mutualists in the legume *Medicago lupulina*. *Ecol Evol*, 7(12), 4367-  
4376. doi:10.1002/ece3.3012

Hasumi, H., Baba, M., Hasumi, Y., Lang, M., Huang, Y., Oh, H. F., ... & Furuya, M. (2015).  
Folliculin-interacting proteins Fnip1 and Fnip2 play critical roles in kidney tumor suppression in

cooperation with Flcn. *Proceedings of the National Academy of Sciences*, 112(13), E1624-  
E1631. doi: 10.1073/pnas.1419502112

Hawkins CE, Baars C, Hesterman H, et al (2006) Emerging disease and population decline of an  
island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation*, **131**, 307-324.  
doi: 10.1016/j.biocon.2006.04.010

Hendricks S, Epstein B, Schönfeld B, et al (2017) Conservation implications of limited genetic  
diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). *Conservation  
Genetics*, 1-6. doi: 10.1007/s10592-017-0939-5

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . .  
Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical  
Solutions, and Future Directions. *Am Nat*, 188(4), 379-397. doi:10.1086/688018

Hohenlohe, P. A., McCallum, H. I., Jones, M. E., Lawrance, M. F., Hamede, R. K., & Storfer, A.  
(2019). Conserving adaptive potential: lessons from Tasmanian devils and their transmissible  
cancer. *Conservation Genetics*, 20(1), 81-87. doi:10.1007/s10592-019-01157-5

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.  
*Bioinformatics*, 24(11), 1403-1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components:  
a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1), 94.

Jones, M. E., Paetkau, D., Geffen, E., & Moritz, C. (2004). Genetic diversity and population  
structure of Tasmanian devils, the largest marsupial carnivore. *Mol Ecol*, 13(8), 2197-2209.  
doi:10.1111/j.1365-294X.2004.02239.x

742 Jones, M. E., Cockburn, A., Hamede, R., Hawkins, C., Hesterman, H., Lachish, S., . . .  
743 Pemberton, D. (2008). Life-history change in disease-ravaged Tasmanian devil populations. *Proc*  
744 *Natl Acad Sci U S A*, 105(29), 10023-10027. doi:10.1073/pnas.0711236105  
745 Jones, M. E., & Barmuta, L. A. (2000). Niche differentiation among sympatric Australian  
746 dasyurid carnivores. *Journal of Mammalogy*, 81(2), 434-447. doi: 10.1644/1545-  
747 1542(2000)081<0434:NDASAD>2.0.CO;2  
748 Jones, M. E., & Rose, R. K. (1996). Preliminary assessment of distribution and habitat  
749 associations of the spotted-tailed quoll (*Dasyurus maculatus maculatus*) and eastern quoll (*D.*  
750 *viverrinus*) in Tasmania to determine conservation and reservation status: Tasmanian Public Land  
751 Use Commission.  
752 Jones, M. E., Jarman, P. J., Lees, C. M., Hesterman, H., Hamede, R. K., Mooney, N. J., . . .  
753 McCallum, H. (2007). Conservation management of Tasmanian devils in the context of an  
754 emerging, extinction-threatening disease: devil facial tumor disease. *EcoHealth*, 4(3), 326-337.  
755 doi: 10.1007/s10393-007-0120-6  
756 Kozakiewicz, C. P., Burrridge, C. P., Funk, W. C., VandeWoude, S., Craft, M. E., Crooks, K. R.,  
757 . . . Carver, S. (2018). Pathogens in space: advancing understanding of pathogen dynamics and  
758 disease ecology through landscape genetics. *Evolutionary Applications*, 11(10), 1763-1778. doi:  
759 10.1111/eva.12678  
760 Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution.  
761 *Evolution*, 30(2), 314-334. doi:10.1111/j.1558-5646.1976.tb00911.x  
762 Lande, R. (1988). Genetics and demography in biological conservation. *Science*, 241(4872),  
763 1455-1460. doi: 10.1126/science.3420403

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, 142(6), 911-927. doi: 10.1086/285580

Leo, S. S., Gonzalez, A., & Millien, V. (2016). Multi-taxa integrated landscape genetics for zoonotic infectious diseases: deciphering variables influencing disease emergence. *Genome*, 59(5), 349-361. doi:10.1139/gen-2016-0039

Leushacke, M., Sporle, R., Bernemann, C., Brouwer-Lehmitz, A., Fritzmann, J., Theis, M., . . . Morkel, M. (2011). An RNA interference phenotypic screen identifies a role for FGF signals in colon cancer progression. *PLoS One*, 6(8), e23381. doi:10.1371/journal.pone.0023381

Li-Hawkins, J., Lund, E. G., Bronson, A. D., & Russell, D. W. (2000). Expression cloning of an oxysterol 7 $\alpha$ -hydroxylase selective for 24-hydroxycholesterol. *J Biol Chem*, 275(22), 16543-16549. doi:10.1074/jbc.M001810200

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., Verity, B., & Kelley, J. (2017). Composite measures of selection can improve the signal-to-noise ratio in genome scans. *Methods in Ecology and Evolution*, 8(6), 717-727. doi:10.1111/2041-210x.12774

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol*, 23(9), 2178-2192. doi:10.1111/mec.12725

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol*, 24(5), 1031-1046. doi:10.1111/mec.13100

785 Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer,  
786 A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA  
787 sequencing for genome scans of adaptation. *Molecular ecology resources*, 17(2), 142-152. doi:  
788 10.1111/1755-0998.12635

789 Loyau, T., Hennequet-Antier, C., Coustham, V., Berri, C., Leduc, M., Crochet, S., ... & Brionne,  
790 A. (2016). Thermal manipulation of the chicken embryo triggers differential gene expression in  
791 response to a later heat challenge. *BMC genomics*, 17(1), 329. doi: 10.1186/s12864-016-2661-y

792 Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation  
793 of census and effective population sizes: the increasing usefulness of DNA-based approaches.  
794 *Conservation Genetics*, 11(2), 355-373. doi: 10.1007/s10592-010-0050-7

795 Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of  
796 different selection signature statistics and a new strategy for combining them. *Heredity (Edinb)*,  
797 115(5), 426-436. doi:10.1038/hdy.2015.42

798 Mackinnon, M. J., Ndila, C., Uyoga, S., Macharia, A., Snow, R. W., Band, G., . . . Williams, T.  
799 N. (2016). Environmental Correlation Analysis for Genes Associated with Protection against  
800 Malaria. *Mol Biol Evol*, 33(5), 1188-1204. doi:10.1093/molbev/msw004

801 Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., . . .  
802 FORTIN, M. J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive  
803 variation in the field. *Molecular ecology*, 19(17), 3760-3772. doi: 10.1111/j.1365-  
804 294X.2010.04717.x

805 Margres, M. J., Ruiz-Aravena, M., Hamede, R., Jones, M. E., Lawrance, M. F., Hendricks, S. A.,  
806 . . . Storfer, A. (2018). The Genomic Basis of Tumor Regression in Tasmanian Devils  
807 (*Sarcophilus harrisii*). *Genome Biol Evol*, 10(11), 3012-3025. doi:10.1093/gbe/evy229  
808 Margres, M. J., Jones, M. E., Epstein, B., Kerlin, D. H., Comte, S., Fox, S., ... & Lazenby, B.  
809 (2018). Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a  
810 transmissible cancer. *Molecular ecology*, 27(21), 4189-4199. doi: 10.1111/mec.14853  
811 Maserati, M., Dai, X., Walentuk, M., & Mager, J. (2014). Identification of four genes required for  
812 mammalian blastocyst formation. *Zygote*, 22(3), 331-339. doi:10.1017/S0967199412000561  
813 McCallum, H. (2008). Tasmanian devil facial tumour disease: lessons for conservation biology.  
814 *Trends Ecol Evol*, 23(11), 631-637. doi:10.1016/j.tree.2008.07.001  
815 McCallum, H. (2012). Disease and the dynamics of extinction. *Philos Trans R Soc Lond B Biol*  
816 *Sci*, 367(1604), 2828-2839. doi:10.1098/rstb.2012.0224  
817 McCallum, H., Jones, M., Hawkins, C., Hamede, R., Lachish, S., Sinn, D., . . . Lazenby, B.  
818 (2009). Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-  
819 induced extinction. *Ecology*, 90(12), 3379-3392. doi: 10.1890/08-1763.1  
820 McCallum, H., Tompkins, D. M., Jones, M., Lachish, S., Marvanek, S., Lazenby, B., . . .  
821 Hawkins, C. E. (2007). Distribution and Impacts of Tasmanian Devil Facial Tumor Disease.  
822 *EcoHealth*, 4(3), 318-325. doi:10.1007/s10393-007-0118-0  
823 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., ... &  
824 DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing  
825 next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303. doi:  
826 10.1101/gr.107524.110

827 Miller, W., Hayes, V. M., Ratan, A., Petersen, D. C., Wittekindt, N. E., Miller, J., . . . Zhao, F.  
828 (2011). Genetic diversity and population structure of the endangered marsupial *Sarcophilus*  
829 *harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences*, 108(30), 12348-  
830 12353. doi: 10.1073/pnas.1102838108

831 Mollenhauer, J., Deichmann, M., Helmke, B., Müller, H., Kollender, G., Holmskov, U., . . .  
832 Bantel-Schaal, U. (2003). Frequent downregulation of DMBT1 and galectin-3 in epithelial skin  
833 cancer. *International journal of cancer*, 105(2), 149-157. doi: 10.1002/ijc.11072

834 Mukherjee, S., Mukherjee, A., Jasrotia, R. S., Jaiswal, S., Iquebal, M. A., Longkumer, I., . . .  
835 Kumar, D. (2019). Muscle transcriptome signature and gene regulatory network analysis in two  
836 divergent lines of a hilly bovine species *Mithun* (*Bos frontalis*). *Genomics*.  
837 doi:10.1016/j.ygeno.2019.02.004

838 Murchison, E. P., Tovar, C., Hsu, A., Bender, H. S., Kheradpour, P., Rebbeck, C. A., . . .  
839 Blizzard, C. A. (2010). The Tasmanian devil transcriptome reveals Schwann cell origins of a  
840 clonally transmissible cancer. *Science*, 327(5961), 84-87. doi: 10.1016/j.cell.2011.11.065

841 Murchison, E. P., Schulz-Trieglaff, O. B., Ning, Z., Alexandrov, L. B., Bauer, M. J., Fu, B., . . .  
842 Stratton, M. R. (2012). Genome sequencing and analysis of the Tasmanian devil and its  
843 transmissible cancer. *Cell*, 148(4), 780-791. doi:10.1016/j.cell.2011.11.065

844 National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of  
845 Medicine (US), National Center for Biotechnology Information; [1988]. Available from:  
846 <https://www.ncbi.nlm.nih.gov/>



847 Nei, M., & Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of  
848 restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10), 5269-5273.  
849 doi: 10.1073/pnas.76.10.5269

850 Orsini, L., Mergeay, J., Vanoverbeke, J., & De Meester, L. (2013). The role of selection in  
851 driving landscape genomic structure of the waterflea *Daphnia magna*. *Molecular ecology*, 22(3),  
852 583-601. doi: 10.1111/mec.12117

853 Ostrander, E. A., Davis, B. W., & Ostrander, G. K. (2016). Transmissible Tumors: Breaking the  
854 Cancer Paradigm. *Trends Genet*, 32(1), 1-15. doi:10.1016/j.tig.2015.10.001

855 Pearse, A. M., Swift, K., Hodson, P., Hua, B., McCallum, H., Pyecroft, S., . . . Belov, K. (2012).  
856 Evolution in a transmissible cancer: a study of the chromosomal changes in devil facial tumor  
857 (DFT) as it spreads through the wild Tasmanian devil population. *Cancer Genet*, 205(3), 101-112.  
858 doi:10.1016/j.cancergen.2011.12.001

859 Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., . . . Shi, H. (2010). Genetic variations in  
860 Tibetan populations and high-altitude adaptation at the Himalayas. *Molecular biology and*  
861 *evolution*, 28(2), 1075-1081. doi: 10.1093/molbev/msq290

862 Purwar, R., Schlapbach, C., Xiao, S., Kang, H. S., Elyaman, W., Jiang, X., . . . Kuchroo, V. K.  
863 (2012). Robust tumor immunity to melanoma mediated by interleukin-9-producing T cells.  
864 *Nature medicine*, 18(8), 1248. doi: 10.1038/nm.2856

865 Pye, R., Hamede, R., Siddle, H. V., Caldwell, A., Knowles, G. W., Swift, K., . . . Woods, G. M.  
866 (2016). Demonstration of immune responses against devil facial tumour disease in wild  
867 Tasmanian devils. *Biology Letters*, 12(10), 20160553. doi: 10.1098/rsbl.2016.0553

868 Qi, C., Zhu, Y. T., Hu, L., & Zhu, Y. J. (2009). Identification of Fat4 as a candidate tumor  
869 suppressor gene in breast cancers. *International journal of cancer*, 124(4), 793-798. doi:  
870 10.1002/ijc.23775

871 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing  
872 genomic features. *Bioinformatics*, 26(6), 841-842. doi: 10.1002/ijc.23775

873 Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical  
874 guide to environmental association analysis in landscape genomics. *Mol Ecol*, 24(17), 4348-  
875 4370. doi:10.1111/mec.13322

876 RICHTER-BOIX, A. L. E. X., Quintela, M., Segelbacher, G., & Laurila, A. (2011). Genetic  
877 analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in  
878 *Rana arvalis*. *Molecular ecology*, 20(8), 1582-1600. doi: 10.1002/ijc.23775

879 Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., ... &  
880 Beltran, H. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, 161(5), 1215-  
881 1228. doi: 10.1016/j.cell.2015.05.001

882 Rounsevell, D., Taylor, R., & Hocking, G. (1991). Distribution records of native terrestrial  
883 mammals in Tasmania. *Wildlife Research*, 18(6), 699-717. doi: 10.1071/WR9910699

884 Ruiz-Aravena, M., Jones, M. E., Carver, S., Estay, S., Espejo, C., Storfer, A., & Hamede, R. K.  
885 (2018). Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease.  
886 *Proceedings of the Royal Society B*, 285(1891), 20182239. doi: 10.1098/rspb.2018.2239

887 Russell, T., Madsen, T., Thomas, F., Raven, N., Hamede, R., & Ujvari, B. (2018). Oncogenesis  
888 as a Selective Force: Adaptive Evolution in the Face of a Transmissible Cancer. *Bioessays*, 40(3).  
889 doi:10.1002/bies.201700146

890 Schwabl, P., Llewellyn, M. S., Landguth, E. L., Andersson, B., Kitron, U., Costales, J. A., . . .  
891 Grijalva, M. J. (2017). Prediction and Prevention of Parasitic Diseases Using a Landscape  
892 Genomics Framework. *Trends Parasitol*, 33(4), 264-275. doi:10.1016/j.pt.2016.10.008  
893 Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., . . .  
894 Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North  
895 American grey wolves. *Mol Ecol*, 25(1), 380-402. doi:10.1111/mec.13364  
896 Shaffer, M. L. (1981). Minimum population sizes for species conservation. *BioScience*, 31(2),  
897 131-134. doi: 10.2307/1308256  
898 Shinagawa, N., Yamazaki, K., Tamura, Y., Imai, A., Kikuchi, E., Yokouchi, H., . . . Nishimura,  
899 M. (2008). Immunotherapy with dendritic cells pulsed with tumor-derived gp96 against murine  
900 lung cancer is effective through immune response of CD8+ cytotoxic T lymphocytes and natural  
901 killer cells. *Cancer Immunol Immunother*, 57(2), 165-174. doi:10.1007/s00262-007-0359-3  
902 Siddle, H. V., Marzec, J., Cheng, Y., Jones, M., & Belov, K. (2010). MHC gene copy number  
903 variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proc Biol Sci*,  
904 277(1690), 2001-2006. doi:10.1098/rspb.2009.2362  
905 Siddle, H. V., Kreiss, A., Tovar, C., Yuen, C. K., Cheng, Y., Belov, K., . . . Jones, M. E. (2013).  
906 Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease  
907 illustrates immune escape by a contagious cancer. *Proceedings of the National Academy of*  
908 *Sciences*, 110(13), 5103-5108. doi: 10.1073/pnas.1219920110  
909 Singh, C., Glaab, E., & Linster, C. L. (2017). Molecular Identification of d-Ribulokinase in  
910 Budding Yeast and Mammals. *J Biol Chem*, 292(3), 1005-1028. doi:10.1074/jbc.M116.760744

911 Smith, K. F., Sax, D. F., & Lafferty, K. D. (2006). Evidence for the role of infectious disease in  
 912 species extinction and endangerment. *Conservation biology*, 20(5), 1349-1357. doi:  
 913 10.1111/j.1523-1739.2006.00524.x

914 Stammnitz, M. R., Coorens, T. H. H., Gori, K. C., Hayes, D., Fu, B., Wang, J., . . . Murchison, E.  
 915 P. (2018). The Origins and Vulnerabilities of Two Transmissible Cancers in Tasmanian Devils.  
 916 *Cancer Cell*, 33(4), 607-619 e615. doi:10.1016/j.ccell.2018.03.013

917 Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., &  
 918 Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree  
 919 species. *Mol Ecol*, 23(10), 2500-2513. doi:10.1111/mec.12751

920 Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape  
 921 genetics and landscape genomics. *Frontiers in genetics*, 9, 68. doi: 10.3389/fgene.2018.00068

922 Storfer, A., Hohenlohe, P. A., Margres, M. J., Patton, A., Fraik, A. K., Lawrance, M., . . . Jones,  
 923 M. E. (2018). The devil is in the details: Genomics of transmissible cancers in Tasmanian devils.  
 924 *PLoS pathogens*, 14(8), e1007098. doi: 10.1371/journal.ppat.1007098

925 Sun, Z. X., Zhai, Y. F., Zhang, J. Q., Kang, K., Cai, J. H., Fu, Y., . . . Zhang, W. Q. (2015). The  
 926 genetic basis of population fecundity prediction across multiple field populations of *Nilaparvata*  
 927 *lugens*. *Mol Ecol*, 24(4), 771-784. doi:10.1111/mec.13069

928 Sun, R., Eriksson, S., & Wang, L. (2014). Down-regulation of mitochondrial thymidine kinase 2  
 929 and deoxyguanosine kinase by didanosine: implication for mitochondrial toxicities of anti-HIV  
 930 nucleoside analogs. *Biochemical and biophysical research communications*, 450(2), 1021-1026.  
 931 doi: 10.1016/j.bbrc.2014.06.098

932 Szkiba, D., Kapun, M., von Haeseler, A., & Gallach, M. (2014). SNP2GO: functional analysis of  
933 genome-wide association studies. *Genetics*, 197(1), 285-289. doi: 10.1534/genetics.113.160341  
934 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA  
935 polymorphism. *Genetics*, 123(3), 585-595.  
936 Tsai, C. T., Yang, P. M., Chern, T. R., Chuang, S. H., Lin, J. H., Klemm, L., ... & Chen, C. C.  
937 (2014). AID downregulation is a novel function of the DNMT inhibitor 5-aza-deoxycytidine.  
938 *Oncotarget*, 5(1), 211. doi: 10.18632/oncotarget.1319  
939 Tovar, C., Patchett, A. L., Kim, V., Wilson, R., Darby, J., Lyons, A. B., & Woods, G. M. (2018).  
940 Heat shock proteins expressed in the marsupial Tasmanian devil are potential antigenic  
941 candidates in a vaccine against devil facial tumour disease. *PLoS One*, 13(4), e0196469.  
942 doi:10.1371/journal.pone.0196469  
943 Tsapogas, P., Breslin, T., Bilke, S., Lagergren, A., Månsson, R., Liberg, D., ... & Sigvardsson, M.  
944 (2003). RNA analysis of B cell lines arrested at defined stages of differentiation allows for an  
945 approximation of gene expression patterns during B cell development. *Journal of leukocyte*  
946 *biology*, 74(1), 102-110. doi: 10.1189/jlb.0103008  
947 Tseng, Y. C., Chen, R. D., Lucassen, M., Schmidt, M. M., Dringen, R., Abele, D., & Hwang, P.  
948 P. (2011). Exploring uncoupling proteins and antioxidant mechanisms under acute cold exposure  
949 in brains of fish. *PLoS One*, 6(3), e18180. doi: 10.1371/journal.pone.0018180  
950 Ullrich, V., & Brugger, R. (1994). Prostacyclin and thromboxane synthase: new aspects of  
951 hemethiolate catalysis. *Angewandte Chemie International Edition in English*, 33(19), 1911-1919.  
952 doi: 10.1002/anie.199419111

953 Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L., & Lotterhos, K. E. (2017).  
954 minotaur: A platform for the analysis and visualization of multivariate results from genome scans  
955 with R Shiny. *Molecular ecology resources*, 17(1), 33-43. doi: 10.1111/1755-0998.12579  
956 Waddington, C. H. (1974). A catastrophe theory of evolution. *Annals of the New York Academy*  
957 *of Sciences*, 231(1), 32-41. doi: 10.1111/j.1749-6632.1974.tb20551.x  
958 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population  
959 structure. *Evolution*, 38(6), 1358-1370. doi: 10.1111/j.1749-6632.1974.tb20551.x  
960 Wenzel, M. A., Douglas, A., James, M. C., Redpath, S. M., & Piernney, S. B. (2016). The role of  
961 parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus*  
962 *scoticus*). *Mol Ecol*, 25(1), 324-341. doi:10.1111/mec.13473  
963 Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., . . .  
964 Federhen, S. (2006). Database resources of the national center for biotechnology information.  
965 *Nucleic acids research*, 35(suppl\_1), D5-D12. doi: 10.1093/nar/gkj158  
966 Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local  
967 adaptation: Inference of a null model through trimming the distribution of  $F_{ST}$ . *The American*  
968 *Naturalist*, 186(S1), S24-S36. doi: 10.1086/682949  
969 Wright, B., Willet, C. E., Hamede, R., Jones, M., Belov, K., & Wade, C. M. (2017). Variants in  
970 the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide  
971 association. *Sci Rep*, 7(1), 423. doi:10.1038/s41598-017-00439-7  
972 Xu, Q., Wang, Y. C., Liu, R., Brito, L. F., Kang, L., Yu, Y., . . . Liu, A. (2017). Differential gene  
973 expression in the peripheral blood of Chinese Sanhe cattle exposed to severe cold stress. *Genet*  
974 *Mol Res*, 16(2). doi:10.4238/gmr16029593

## Acknowledgements

We thank Omar Cornejo, and his lab, David Crowder, and Joanna Kelley's lab for their helpful review and insightful discussion. We are grateful to the anonymous reviewers for feedback on our manuscript. This work was funded by NSF grant DEB-1316549 and NIH grant R01-GM126563 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program (AS, PAH, MJ and HM).

## Author Contributions

AKF conducted the analyses and wrote the paper; MJM, SB and BE assisted with the analyses and helped write the paper; BS, SH, AV, and AS assisted with the sample preparation and helped write the paper; RH, and MJ collected the samples, contributed to the study design, managed field surveys and helped write the paper; HM helped write the paper; EL-C, and SJK assisted with the analyses; PH, JLK, and AS directed the project, supervised this work and helped write the paper.

## Ethics

Animal use was approved under IACUC protocol ASAF#04392 at Washington State University.

## Data Accessibility

The sequence data has been deposited at NCBI under BioProject PRJNA306495 (<http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA306495>).

Code is available at [github.com/jokelley/devil-landscape-genomics](https://github.com/jokelley/devil-landscape-genomics).

## Tables and Figures

**Table 1:** Information for samples collected from each population (the acronyms for the population included in parentheses). Values listed include the number of ear biopsy samples collected per population prior to and post-disease arrival, the years the populations were sampled, and the date that DFTD was first detected in that population.

Populations	Samples Pre-Disease	Samples Post-Disease	Years Samples Collected	Date of DFTD Infection
Fentonbury (FEN)	97	157	2004-2009	2005
Forestier (FOR)	137	382	2004-2010, 2012-2013	2004
Freycinet (FRY)	511	562	1999-2014	2001
Mt William (MTW)	N/A	155	1999-2014	1996
Narawntapu (NAR)	258	153	1999-2000, 2003-2012	2007
West Pencil Pine (WPP)	55	357	2006-2014	2006
Woolnorth (WOO)	491	N/A	2006-2010, 2011-12	N/A



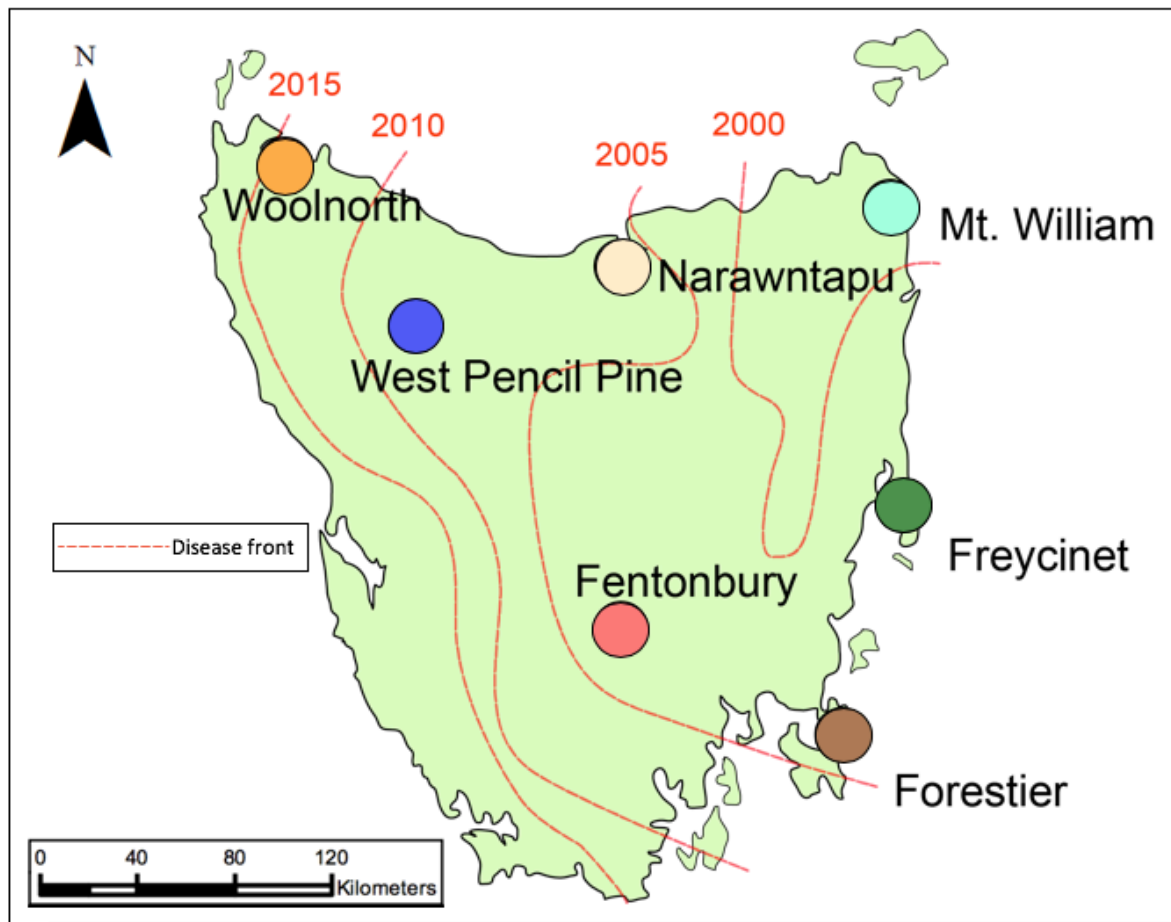
**Table 2:** The mean pairwise  $F_{ST}$  values  $\pm$  95% confidence intervals generated by resampling by bootstrapping 10,000 iterations. The values below the diagonal are pairwise comparisons for samples collected pre-disease and those above the diagonal are the same for post-disease. All pairwise comparisons were statistically significant.

		Post-DFTD						
		FEN	FOR	FRY	MTW	NAR	WOO	WPP
Pre-DFTD	FEN		0.049 ± 0.028	0.051 ± 0.026	0.062 ± 0.035	0.026 ± 0.015	N/A	0.066 ± 0.037
	FOR	0.057 ± 0.033		0.030 ± 0.017	0.033 ± 0.019	0.0445 ± 0.025	N/A	0.094 ± 0.052
	FRY	0.046 ± 0.029	0.036 ± 0.021		0.0212 ± 0.012	0.038 ± 0.022	N/A	0.091 ± 0.051
	MTW	N/A	N/A	N/A		0.050 ± 0.028	N/A	0.060
	NAR	0.022 ± 0.013	0.050 ± 0.028	0.039 ± 0.021	N/A		N/A	0.106 ± 0.057
	WOO	0.156 ± 0.087	0.176 ± 0.098	0.150 ± 0.080	N/A	0.129 ± 0.072		N/A
	WPP	0.063 ± 0.036	0.110 ± 0.059	0.104 ± 0.059	N/A	0.061 ± 0.035	0.077 ± 0.044	

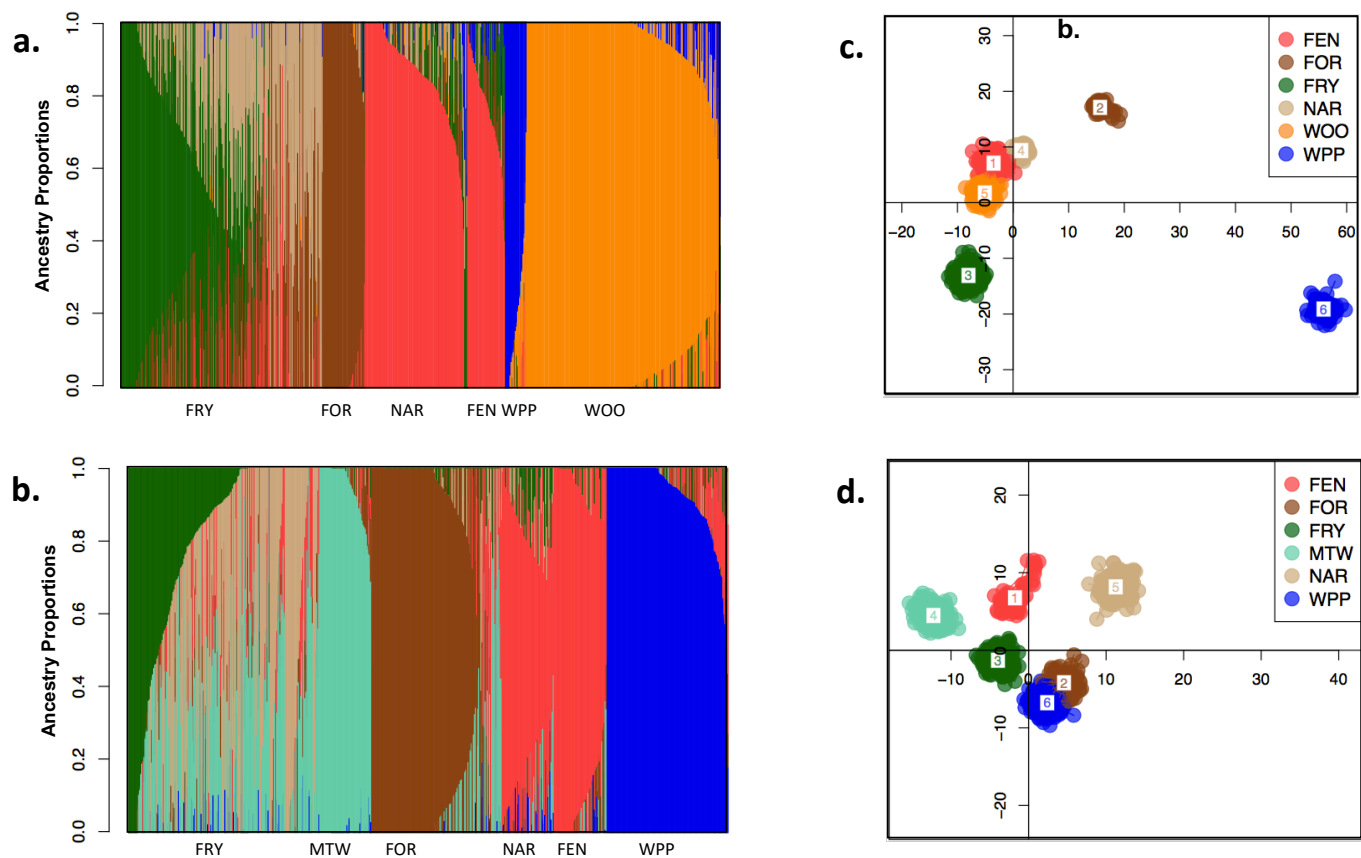
1002 **Table 3:** The centroid values for the eight abiotic environmental variables and the biotic variable utilized in the GEA  
1003 associations for each population.

	Mean Annual Temperature (°C)	Precipitation seasonality (mm)	Isothermality	Annual Temperature Range (°C)	Enhanced Vegetation Index (EVI)	Length of Sealed Roads (km)	Elevation (m)	Surface Area of Water (m <sup>2</sup> )	Average Disease Prevalence
<b>FEN</b>	10.51	18.24	49.87	20.07	34.87	314.16	232.15	455.62	0.27
<b>FOR</b>	12.11	15.71	51.40	15.99	52.51	263.31	88.78	29627.00	0.22
<b>FRY</b>	12.67	14.51	54.09	17.24	74.01	61.65	57.17	124.44	0.11
<b>MTW</b>	13.12	21.61	49.88	16.4	52.15	420.78	43.75	16695.96	0.19
<b>NAR</b>	12.48	31.57	48.07	18.09	76.58	671.02	74.31	316.11	0.02
<b>WPP</b>	7.98	31.17	47.98	17.11	61.24	245.23	703.74	494.04	0.07
<b>WOO</b>	12.83	33.56	49.92	14.79	51.42	53.83	12.83	1133.77	0.00

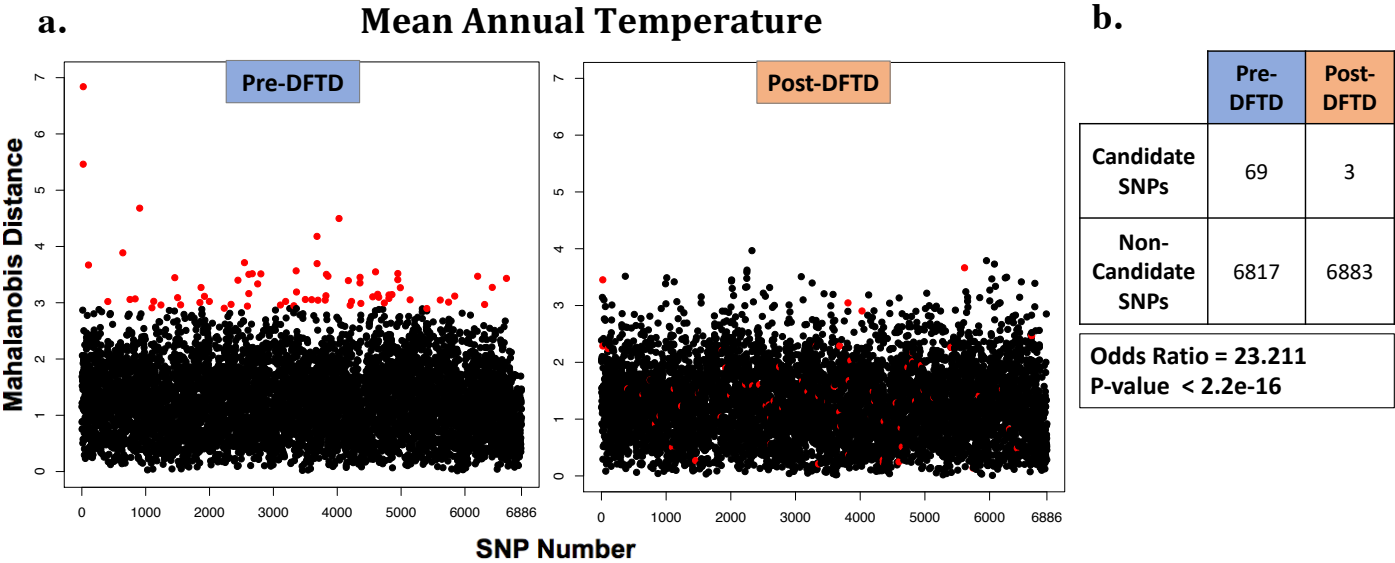
1004



**Figure 1:** Map of Tasmania with each sampling location. The red lines and the corresponding years indicate the first year that disease was detected in these locations.



**Figure 2:** Population assignments computed by fastSTRUCTURE and DAPC for samples collected prior to (a,c) and post (b,d) DFTD arrival. (a,b) Each vertical bar in the fastSTRUCTURE plots represents a single individual sampled at one of the sampling locations which are abbreviated along the x-axis. K=6 is plotted here. Each color in all plots represents a distinct genetic cluster. (c,d) DAPC scatterplots show the first two principal components for K=6.



**Figure 3:** Mahalanobis distances from MINOTAUR for each SNP pre-DFTD arrival (left plot) and post-DFTD arrival (right plot) for GEAs with mean annual temperature. SNPs are ordered by position along the chromosomes. The top 1% of loci with the largest Mahalanobis distance values pre-DFTD are indicated in red. (b) Post-hoc analyses found that a significantly greater number of candidate loci detected pre-DFTD arrival were not detected post-DFTD arrival. This trend was detected consistently across all eight of the abiotic environmental variables tested in genetic-environmental association analyses (output from the remaining seven variables found in supplementary figure 2).