Host-Pathogen Coevolution Between Tasmanian Devils (*Sarcophilus harrisii*) and Devil Facial Tumor Disease

Dylan G. Gallinson[1,2], Ryan C. McMinds[1,2], Mark J. Margres[2]

College of Public Health[1], Department of Integrative Biology[2], University of South Florida

Introduction

Over the last two decades, Tasmanian devils have suffered overall population declines of more than 80% due to the evolution of a species-specific transmissible cancer — devil facial tumor disease (DFTD). DFTD originated from a Schwann cell within a single female devil in northeastern Tasmania and is able to transmit between hosts by downregulating MHC (Cheng et al., 2019). Because DFTD is nearly 100% fatal, and devils are universally susceptible, epidemiological models initially predicted host extinction within a few decades. However, no local extinction events have occurred, and devils appear to be evolving rapidly in response to DFTD (Epstein et al., 2016). Furthermore, it appears as though DFTD is also evolving, as multiple lineages of the cancer exist (Kwon et al., 2020). Despite evidence of multiple tumor lineages, DFTD evolution remains poorly characterized and nothing is currently known about devil-DFTD coevolution. Here, we characterize the genetic underpinnings of devil and DFTD disease-related phenotypes by investigating the extent to which genotype-by-genotype interactions are driving variation in these traits.

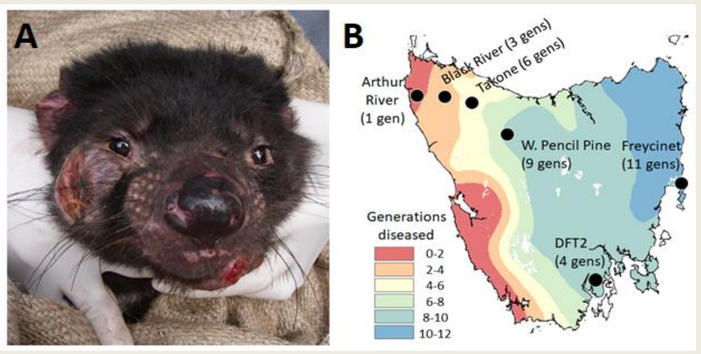

Figure 1. A. Tasmanian devil with DFTD. **B.** Movement of DFTD from east to west across Tasmania with number of devil generations since DFTD arrival denoted below each site in parenthesis. DFT2 is a second transmissible cancer derived independently from a male devil which is symptomatically equivalent to DFTD.

Hypotheses

Question 1: Is genetic variation spatially sorted within devils and DFTD, and does this sorting differ between devils and the cancer?

Hypothesis 1: Devils cluster into genetically distinct populations because they occupy a limited territorial range (i.e., minimal gene flow) and reproduce sexually. Conversely, although DFTD is equivalent in its spatial range, it is incapable of gene flow due to reproducing asexually, facilitating a heterogeneity of lineages within a given site.

Question 2: What is the host and pathogen genomic architecture underlying force of infection and virulence?

Hypothesis 2: Devil traits relevant to survival in the presence of DFTD are variable and heritable. Although novel beneficial mutations are unlikely in such a short timespan, large-effect variants which were previously neutral may already exist in the population.

Question 3: To what extent is coevolution driving evolutionary change within devils and DFTD?

Hypothesis 3: Coevolution is detectable and contributing to rapid evolution in both devils and DFTD and will be strongest in tumor virulence due to its dramatic effect on devil fitness.

Methods

Sample collection and data processing: A total of 1056 samples (507 devils and 549 tumors) were collected and sequenced using a probe-capture panel targeting all known cancer-associated loci. Sampling sites were selected along an east-west axis of Tasmania to represent varying durations of devil coexistence with DFTD. Sequencing reads were cleaned and aligned to the devil reference genome, mSarHar1.11 using BWA (Li 2013). The GATK pipeline (Van der Auwera & O'Connor 2020) was used to call SNPs.

Statistical analyses: Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010) was used to find genomic clusters of devils and DFTD. A genome-wide association (GWA) approach was used to investigate the genetics underlying force of infection and tumor virulence.

Figure 2. Overview of the study workflow. Raw reads were processed as above. A Bayesian Sparse Linear Mixed Model (BSLMM), implemented in GEMMA (Zhou & Stephens, 2012), was fit to the devil and tumor genome. To investigate devil-DFTD coevolution through genotype-by-genotype interactions, Analysis with a Two-Organism Mixed-Model (ATOMM; Wang et al., 2018) was used.

Genetic Variation Differs Spatially Between Host and Pathogen

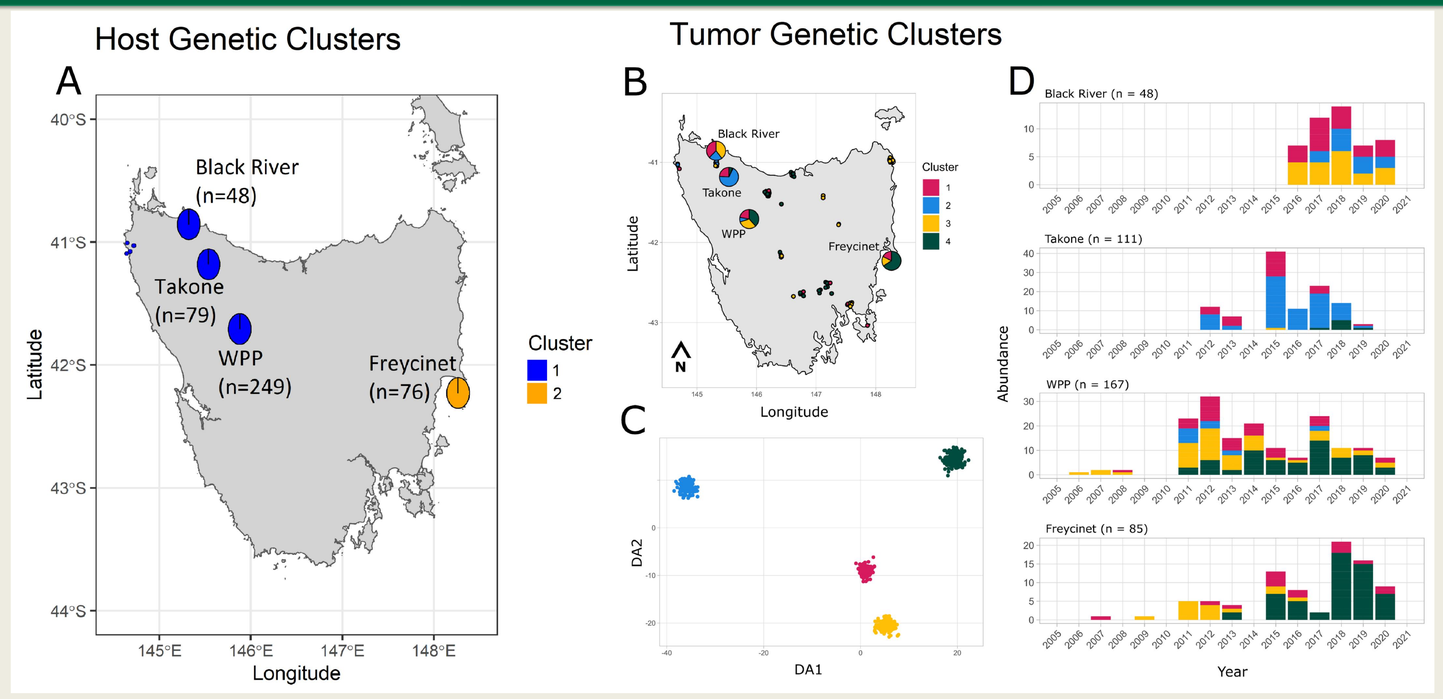

Figure 3. Devil and DFTD lineages identified via genomic clustering using DAPC. **A.** Geographic distribution of devil lineages across Tasmania show genetically distinct populations from east to west, indicating minimal gene flow. **B.** Geographic distribution of tumor clusters across Tasmania showing a high degree of multi-lineage coexistence at each site. Pie charts represent lineage proportions for highly sampled sites. **C.** Four tumor lineages identified with DAPC. **D.** Tumor lineage abundance from the four most densely sampled sites over time indicates the possibility for competition between lineages, with the potential for lineage replacement (e.g., Freycinet).

Disease-Relevant Traits Vary by Genome

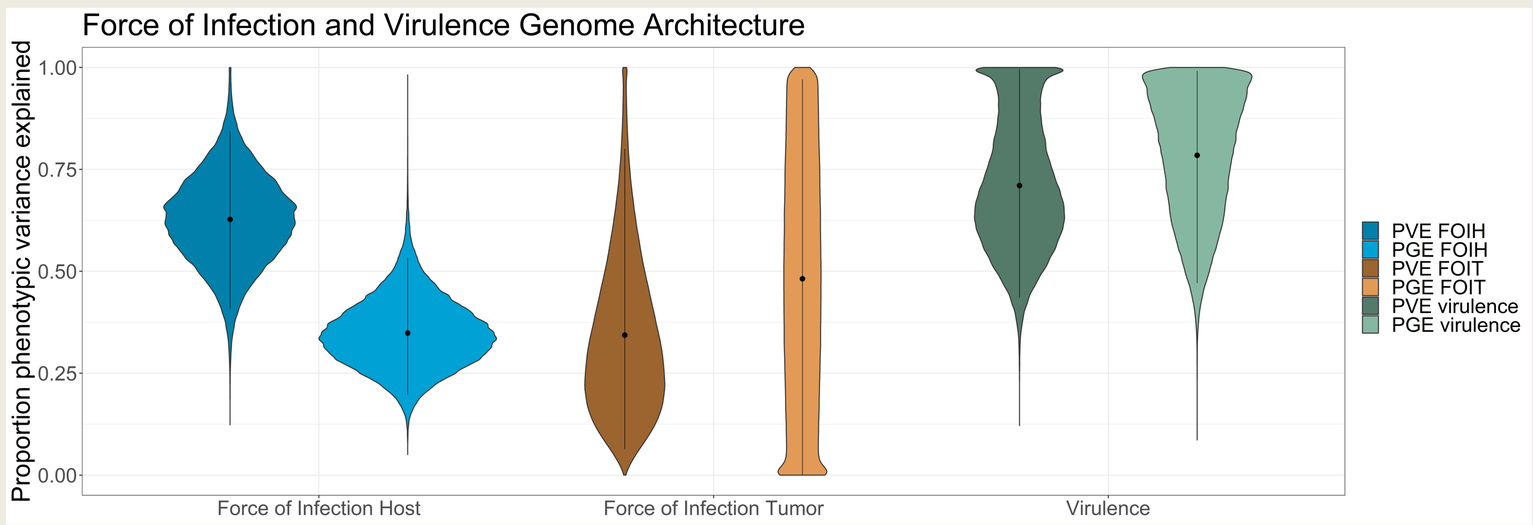

Figure 4. Genomic architecture for devil phenotypes (force of infection host; FOIH) and tumor phenotypes (force of infection tumor (FOIT) and virulence) obtained using a Bayesian Sparse Linear Mixed Model (BSLMM). Violin plots show proportion of phenotypic variance explained (PVE) and proportion of genetic variance explained (PGE) estimates generated through model fitting. PVE measures the cumulative effect of all SNPs input to the model; hence, high PVE values indicate much of the trait's variance is attributable to variation across genomes. PGE measures the proportion of genetic variance explained by the model's sparse effect terms (i.e., contribution of large-effect loci). Variation in DFTD virulence appears largely attributable to the tumor genome and is controlled by few large-effect loci. Variation in force of infection can be explained primarily by the devil genome and to a small degree by the tumor genome.

Genotype-by-Genotype Interactions Contribute to Force of Infection

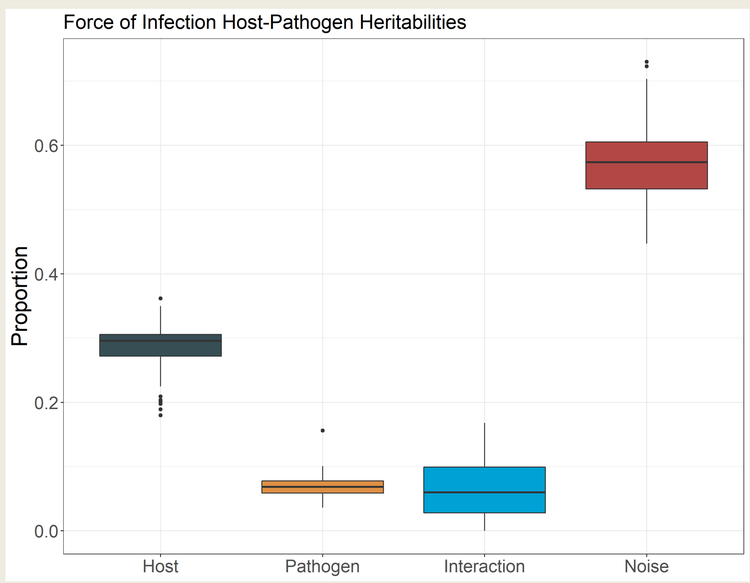

Figure 5. Force of infection heritability estimates obtained through model fitting using Analysis of a Two-Organism Mixed Model (ATOMM). Estimates refer to the proportion of host genome, pathogen genome, and genomic interactions explaining variance in force of infection.

Table 1. Top 5 significant genotype-by-genotype interactions for force of infection identified using genotype-by-genotype interaction tests in ATOMM.

Host			Pathogen		
Gene	CHR	Distance	Gene	CHR	Distance
RUSC1: Involved in MAPK-mediated Trk receptor signaling	4	+7kb	*AGPS*: Upregulated in primary tumors of multiple aggressive human cancers (Benjamin et al., 2013)	3	-60kb
HNF1A: Transcription factor required for expression of several genes in the liver and pancreas	1	Intron	*SKOR2*: Sequence specific double stranded DNA binding activity	1	-57kb
PDS5B: Negative regulator of cell proliferation, making it a possible tumor suppressor gene	3	3' UTR	LOC100917912: No annotated function	2	+2kb
SLC4A11: Borate cotransporter essential for borate homeostasis, cell growth, and cell proliferation	6	Arg → Gln	*AGPS*: Upregulated in primary tumors of multiple aggressive human cancers (Benjamin et al., 2013)	3	-60kb
DPP6: Single pass type II membrane protein	5	Intron	*GLRA3*: Glycine receptor subunit	6	Intron

References

Benjamin, D. I., et al. (2013). Ether lipid generating enzyme AGPS alters the balance of structural and signaling lipids to fuel cancer pathogenicity. Proceedings of the National Academy of Sciences of the United States of America, 110(37), 14912–14917

Cheng, Y., et al. (2019). Tasmanian devils with contagious cancer exhibit a constricted T-cell repertoire diversity. *Communications Biology*, *2*(1), 1–9

Epstein, B., et al. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications*, *7*(1), 12684

Kwon, Y. M., et al. (2020). Evolution and lineage dynamics of a transmissible cancer in Tasmanian devils. *PLOS Biology*, *18*(11), e3000926

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM

Van der Auwera GA & O'Connor BD. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media

Jombart, T., et al. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11, 94

Wang, M., et al. (2018). Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proceedings of the National Academy of Sciences*, *115*(24), E5440–E5449

Xiang Zhou and Matthew Stephens (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821–824

Conclusions

- Multiple tumor lineages co-occur within each site, facilitating the possibility of competition between tumor lineages and potentially increasing selective pressures for DFTD disease traits. Because devils form distinct genetic clusters, these selective pressures will likely differ from east to west and may result in differing evolutionary trajectories across space.
- High PVE and PGE for force of infection and virulence indicates evolution acting upon previously neutrally segregating alleles. Selection thus favors alleles already present in the population which confer the greatest fitness gains through large-effect phenotype changes.
- Heritability estimates attribute some variance in force of infection to host-pathogen genome interactions and extracting the variants with the most significant genotype-by-genotype interactions reveals potential signatures of coevolving sites.

Acknowledgements

We thank the Storfer Lab (Washington State University) for preparing and sequencing tissues, the Jones Lab (University of Tasmania) and the Hamede Lab (University of Tasmania) for field work collecting samples and metadata, and the McCallum Lab (Griffith University) for aiding in the statistical analysis. This work was funded by NSF DEB-2027446 to MJM and the College of Public Health Genomics Program at USF to RCM.

USF Health

MORSANI COLLEGE OF MEDICINE | COLLEGE OF NURSING | COLLEGE OF PUBLIC HEALTH | TANEJA COLLEGE OF PHARMACY | SCHOOL OF PHYSICAL THERAPY