

**Genomics of Rapid Evolution in the Tasmanian Devil
(*Sarcophilus harrisi*) and Tamarisk Beetles (*Diorhabda* spp.),
Wild Populations of Management Concern**

A Dissertation

Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

with a

Major in Bioinformatics and Computational Biology

in the

College of Graduate Studies

University of Idaho

by

Amanda R. Stahlke

Major Professor: Paul Hohenlohe, Ph.D.


Committee Members: Luke Harmon, Ph.D., Scott Nuismer, Ph.D., Christine Parent, Ph.D.


Department Administrator: David Tank, Ph.D.

May 2021

Authorization to Submit Dissertation

This dissertation of Amanda R. Stahlke, submitted for the degree of Doctor of Philosophy with a Major in Bioinformatics and Computational Biology and titled "Genomics of Rapid Evolution in the Tasmanian Devil (*Sarcophilus harrisii*) and Tamarisk Beetles (*Diorhabda* spp.), Wild Populations of Management Concern" has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:  Date: 4/29/2021
 Paul Hohenlohe, Ph.D.

Committee Members:  Date: 4/29/2021
 Luke Harmon, Ph.D.

 Date: 4/29/21
 Scott Nuismer, Ph.D.

 Date: 4/29/2021
 Christine Parent, Ph.D.

Department Administrator:  Date: 4-29-21
 David Tank, Ph.D.

Abstract

We now recognize the pervasiveness of rapid evolution in the wild; yet it remains a challenge to identify the key patterns and processes to predict outcomes and inform management decisions. In this dissertation, I employ genomic tools and multi-scale analyses over molecules, space, and time to improve our understanding of the genomic context of rapid evolution in two systems of management concern: the Tasmanian devil, *Sarcophilus harrisii*, and a classical biological control (biocontrol) agent, the tamarisk beetle (*Diorhabda* spp.). In these taxonomically-distinct systems, the common management goal is essentially to allow or provide beneficial genetic variation for adaptation to novel selective forces while limiting the potentially adverse effects of fixation and introgression. By spanning the tree of life and the unique circumstances of these systems, we can better identify strategies informative to systems of management concern.

Tasmanian devils (*Sarcophilus harrisii*) are evolving in response to a unique transmissible cancer, devil facial tumor disease (DFTD), first described in 1996. Persistence of wild populations and the recent emergence of a second independently evolved transmissible cancer suggest that transmissible cancers may be a recurrent feature in devils. In **Chapter 1**, I use range-wide RAD capture data and publicly available marsupial genomes to compare signatures of selection across temporal scales. I test the hypothesis that genes or gene pathways under contemporary selection have also been subject to historical selection. I documented genome-wide contemporary and historical selection, but limited evidence of recurrent selection, with no functional enrichment shared among gene sets. My results are consistent with a novel, polygenic evolutionary response of devils to DFTD and can inform management actions to conserve adaptive capacity of devils by identifying high priority targets for genetic monitoring and maintenance of functional diversity.

In **Chapter 2**, I examine the genome-wide outcomes of introduction, spread, and gene flow in the four cryptic species of the tamarisk beetle (*D. carinata*, *D. carinulata*, *D. elongata*, and *D. sublineata*), introduced to control invasive tamarisk (*Tamarix* spp.) in North America. I assembled a *de novo* draft reference genome of *D. carinulata* and used reference-based RADseq to characterize the genetic variation associated with establishment, spread, and gene flow. I found differential establishment and spread among the six released populations. Despite evidence of a substantial genetic bottleneck among collections of *D. carinulata* in North America and low levels of genetic diversity, populations continue to establish and spread, with one ecotype dominating the southward expansion front. I confirmed that *Diorhabda* hybridizes widely in the introduced range. Genetic diversity was greater in populations with hybrids, highlighting potential for increased adaptive capacity. These results provide a snapshot of

introduced *Diorhabda* populations and lay the foundation for further application of genomics to understand contemporary eco-evolution in classical biological control programs.

In **Chapter 3**, I present high-quality assemblies of all four *Diorhabda* species using iterative, third-generation sequencing technologies, in collaboration with the United States Dep. Agriculture, Agricultural Research Service Ag100 Pest Initiative. I annotate these nearly chromosome-level assemblies with population genetic data to validate sex chromosome identification, compare the arrangement of single-copy orthologs, and characterize repeat content among the introduced biocontrol species. These resources dramatically improve our ability to identify mechanisms of rapid evolution in reproductive barriers, adaptation to novel environments, and physiological responses to plant hosts.

Acknowledgements

I would like to acknowledge Paul Hohenlohe, my Major Professor, for his kindness and patient guidance, for providing fertile ground to explore and grow, for helping me realize the potential within myself and our projects. Members of my committee Christine Parent, Scott Nuismer, Erkan Buzbas, and Luke Harmon, greatly advanced my work and development as a scientist by asking tough questions and hosting delightful conversations.

This research and my training have benefited immensely by collaborators beyond the University of Idaho. I started this journey with the foundational training of my undergraduate research advisor Zeynep Ozsoy and Team *Diorhabda* founding member Dan Bean. I am incredibly grateful for smart, creative, and kind collaborators including Ellyn Bitume, Ruth Hufbauer, and Eliza Clark. Natalie West and John Gaskin hosted me for an internship that re-shaped my career trajectory in the best way, as well as ideas and approaches in biocontrol. Scott Geib, Sheina Sim, and Anna Childers have illuminated the world of high-quality reference genomes. I am honored to continue working alongside these folks.

The technical training and friendships forged in the Bioinformatics and Computational Biology (BCB) Program carried me through hurdles of all types. BCB Program Chair Dave Tank shined a welcoming light on paths I didn't know existed. I am grateful to have learned with and from Sarah Hendricks, Kenetta Nunn, Austin Patton, Bob Week, Megan Ruffley, Mason Linscott, Katie Peterson, Alex Fraik, Chava Castaneda Barba, Clint Elg, Breanna Siple, Yesol Sapozhnikov, and Kristen Martinet. The Landscape Genetics DGS Course led by the wonderful Lisette Waits was pivotal, not only in my educational training, but also by introducing me to a network of incredible colleagues and collaborators including Brenna Forester, Amanda Xuereb, and Erin Landguth. Jane Lucas was an accidental mentor/roommate and wonderful friend in my final years. I gained teaching experience that made me a better human and scientist through the DGS course, BIO115L with Trish Hartzell and Martina Ederer, in the ConGen Workshops with Gordon Luikart and Brian Hand, and as a member of the Carpentries group led by JT Van Leuven.

This work was supported by IBEST (the Institute for Bioinformatics and Evolutionary Studies); specifically, through administrative wonders by Lisha Abendroth, genomics support by Dan New, Matt Fagnan, and Sam Hunter, computational support by Benji Oswald, grant support by MollyAnn Jones, and leadership by Jack Sullivan and Barrie Robinson. The work I present here benefited greatly from wonderful colleagues in the halls, at IBEST/CMCI lunches, on couches or the floor of PEES, on the trails of Moscow Mountain, and later in zoom calls. Genomics and bioinformatics were

supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the NIH under grant number P30 GM103324. I was financially supported by the Bioinformatics and Computational Biology Program at the University of Idaho in partnership with IBEST (the Institute for Bioinformatics and Evolutionary Studies), and USDA AFRI NIFA Predoctoral grant 2020-67034-31888.

Work in Chapter One 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. Sample collection was funded under Australian Research Council grants DE170101116, DP110102656, LP0989613 and LP0561120, an ARC Future Fellowship FT100100031, and multiple awards from the University of Tasmania Foundation Eric Guiler Research Grant. This work was made possible by collaborators on the devil project including Kim Andrews, Soraia Barbosa, Cody Wiench, Brendan Epstein, Mark Margres, Chris Kozakiewicz, Lauren Ricci, Lanie Smith, Matt Lawrence, Sebastian Comte, David Hamilton, Manuel Ruiz, Sebastian Comte, David Hamilton, Manuel Ruiz, Barbara Schonfeld, Hamish McCallum, Andrew Storfer, Menna Jones, and Rodrigo Hamede.

Chapter Two presents a 10X library constructed at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Sequencing was carried out at the UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory. Anne Veillet and Meaghan Clark made the huge sample sizes possible by preparing approximately one million libraries. This research was supported by USDA grant Agriculture and Food Research Initiative grant COLO-2016-09135 and a USDA NIFA Hatch project 1012868. Collections in Xinjiang were supported by the Stillinger Herbarium Expedition Grant and would not have been possible without the help of Daoyuan Zhang and Yan-Feng Chen from the Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences. Ben Bloodworth and James Tracy provided coordinates for beetle collections in NM and TX. This work was supported in part by the U.S. Department of Agriculture.

Dedication

For my parents, my brother, my grandmothers, and my partner

Table of Contents

Authorization to Submit Dissertation.....	ii
Abstract	iii
Acknowledgements	v
Dedication	vii
Table of Contents	viii
List of Tables.....	x
List of Figures	xii
Statement of Contribution	xvii
Chapter 1: Contemporary and historical selection in Tasmanian devils (<i>Sarcophilus harrisii</i>) support novel, polygenic response to transmissible cancer.....	1
Abstract	1
Introduction	1
Methods and Materials.....	4
Results	6
Discussion	8
Conclusion.....	12
Chapter 2: Hybridization and range expansion in tamarisk beetles (<i>Diorhabda</i> spp.) introduced to North America for classical biological control	17
Abstract	17
Introduction	17
Methods and Materials.....	22
Results	26
Discussion	29
Chapter 3: <i>De novo</i> chromosome-level genome assemblies and synteny analysis of four cryptic tamarisk biocontrol agent species: <i>Diorhabda carinulata</i> , <i>D. carinata</i> , <i>D. elongata</i> , and <i>D. sublineata</i>	39
Abstract	39

Introduction	39
Methods	42
Results	46
Discussion	49
Conclusion.....	52
Literature Cited	65
Appendix A - Supplementary to Contemporary and historical selection in Tasmanian devils (<i>Sarcophilus harrisi</i>) support novel, polygenic response to transmissible cancer.....	81
Appendix B - Supplementary to Hybridization and range expansion in tamarisk beetles (<i>Diorhabda</i> spp.) introduced to North America for classical biological control.....	103

List of Tables

Table 1.1 Number of adults sampled before and after the year of first detection of DFTD at each site. See Table A.1 for sample size for each year at each locality.....	13
Table 3.1 Input sequence characteristics for PacBio HiFi reads, genome characteristics estimated by Genomescope (Vurture et al. 2017) from HiFi k-mers, and Hi-C alignments according to the PhaseGenomics Hi-C QC tool. For GenomeScope, estimates of genome characteristics based on 21-mer histograms for the HiFi reads following adapter removal.	53
Table 3.2 Primary contig assembly statistics and final Hi-C scaffolded assembly statistics after manual curation.....	54
Table 3.3 Classified repeat content for the assembled genomes of each <i>Diorhabda</i> species.....	55
Table A.1 Number and year of sampling across six localities.....	86
Table A.2 ANGSD genotype calling settings.	88
Table A.3 Estimates of effective population size.....	89
Table A.4 Number of shared genes within 10000 bp of the top 1% of SNPs for the DCMS list of contemporary candidates and each intermediate test. The total number of genes in each list is found in the diagonal. Populations are abbreviated as follows: FEN = Fentonbury, FOR = Forestier, FREY = Freycinet, wuk = wukalina/Mt. William, NAR= Narwantapu, and WPP = West Pencil Pine.	90
Table A.5 Annotated Tasmanian devil gene IDs of within 1000 bp of the top 1% of composite SNPs; i.e., candidates for contemporary selection, provided at: https://github.com/Astahlke/contemporary_historical_sel_devils/blob/master/contemporary/angsd_2019-01-18/next/composite_stat/2019-2-22/results/annotation_top1.0/composite.snps.everything.top.genes.100000bp.txt	91
Table B.1 Sampling locality identification codes (Locality ID), site names, latitude and longitude coordinates, dates and sample sizes for each collection effort. Locality IDs include the country or state sampled and symbols for relevant groups as follows: * = native range, Δ = source population, ∇=release sites, ● = hybrid zone, ■ = <i>D. carinulata</i> expansion front.	104
Table B.2 Available data on releases of the different <i>Diorhabda</i> species by year and state. Note: This table does not reflect every release since not all data was accessible at time of publication. Provided at: https://github.com/Astahlke/DiorhabdaPopulationStructure/blob/master/info/Supp%20tab%203%20Diorhabda%20release%20table.xlsx	106
Table B.3 Detailed Structure results for one repetition of $K = 4$ (presented in Figure 2.2), including sample ID, locality ID, inferred cluster assignments, and 90% credible intervals, available at	

https://github.com/Astahlke/DiorhabdaPopulationStructure/blob/master/structure_analysis/ancestry_confidenceintervals.csv	107
Table B.4 Estimates of π (nucleotide diversity), F_{IS} (inbreeding), and number of private alleles for hybrid populations, pure populations, and native populations of <i>Diorhabda</i> (spp.) by sample size (N). Superscript letters indicate significant differences at $p < 0.05$	108

List of Figures

- Figure 1.1 Map of the six contemporary sampling locations relative to disease prevalence over time (red lines) with the year of first detection labeled at each site. B) Reduced, unrooted time-calibrated phylogeny (Mitchell et al. 2014) of marsupials used to estimate genome-wide historical selection on the devil lineage with estimated divergence times (Ma) indicated along edges. Devil cartoon by David Hamilton. Wallaby, koala, and opossum digital images retrieved from <http://www.shutterstock.com/amplicon>. From top to bottom: The tammar wallaby (*Notamacropus eugenii*), koala (*Phascolarctos cinereus*), Tasmanian devil (*Sarcophilus harrisii*), and South American grey-tailed opossum (*Monodelphis domestica*)..... 14
- Figure 1.2 Results of each elementary test of contemporary selection across populations and the composite scores for final candidate (filled points) and noncandidate (opaque grey points) SNPs, ordered by chromosome and colored by population when applicable. From top to bottom: Change in allele frequency (Δaf), Mathieson and McVean (mm) (Mathieson and McVean 2013), $spatpg$ (Gompert 2016), de-correlated composite of multiple signals ($DCMS$) (Ma et al. 2015). 15
- Figure 1.3 A) Estimates of dN/dS and B) the proportion of sites under positive selection (Yang 2007) for historical candidates across the genome-wide background (red; N=1,982) and candidates for recurrent selection (genes with significant results for both historical and contemporary selection; black). Each point represents the respective result for a single gene. C) Distribution of $-\log(\text{FDR})$ for historical selection across all 6,193 genes tested (gray squares, non-significant at both scales; black squares, significant contemporary and non-significant historical; red circles, non-significant contemporary and significant historical; black circles, significant at both scales). 16
- Figure 2.1 Solid black points indicate sampling localities in western North America at original release sites and along expansion front of *Diorhabda carinulata* (squares), and across original release sites within the suspected hybrid zone of *D. carinata*, *D. elongata*, and *D. sublineata* (closed circles). Transparent filled points indicate known field-releases (Knutson et al 2019; Table B.3) for each respective species. Rivers are drawn in blue..... 34
- Figure 2.2 Genetic clustering using Structure (Pritchard et al 2000) for $K=4$ across all species. Each individual sample is represented by a bar. Individuals are grouped by collection site and ordered by localities of Table B.1. Groups from left to right are (A) *D. carinulata* Fukang source collection, then north to south from original release sites along expansion front; *D. carinata* and *D. sublineata* lab cultures and native range *D. elongata* collected in Greece follow. (B) Individuals collected within the hybrid zone of North America, from north to south. 35
- Figure 2.3 Genetic clustering at $K=3$ within populations of (A) *D. carinulata* and (B) *D. elongata*. 366

Figure 2.4 Population genetic statistics, π (top panel), the number of private alleles (middle), and F_{IS} (bottom) for the global SNP dataset, consisting of all individuals collected at each locality and grouped (from left to right) by collections within the <i>D. carinulata</i> range, laboratory cultures, the native <i>D. elongata</i> range, and those within the suspected hybrid zone (ordered as in Figure 2.2). Shading of individual bars indicates whether the collection was an original release site, site within the native range, natural colonization site, or lab culture (from darkest to lightest). Bars represent +/- the standard error for the respective statistic.....	37
Figure 2.5 Pairwise- F_{ST} compared to haversine distance between respective sites indicate patterns of isolation-by-distance associated with population structure for (A) <i>D. carinulata</i> collected in North America (Mantel's $r = 0.4481$, $p = 0.1011$ and (B) across the potential hybrid zone (Mantel's $r = 0.5108$, $p < 0.005$). Crossed-squares indicate pair-wise comparisons within ecotype, solid squares across ecotype, and filled circles for the hybrid zone. A linear model for each distribution projected behind points as a red line with standard error in grey to aid in visualization.....	38
Figure 3.1 Profiles of 21-mer histograms according to coverage and respective frequency and Genomescope (Vurture et al. 2017) model-fit for the CSS reads of all four species following adapter removal. The short peak with less coverage for all species indicates heterozygosity.	56
Figure 3.2 Hi-C contact heatmaps and curated scaffolds for assemblies of each <i>Diorhabda</i> species..	57
Figure 3.3 Distribution of RADseq markers by female and male coverage across population sampling of <i>D. carinulata</i> . Markers significantly associated with males or females are highlighted in red boxes.	58
Figure 3.4 Distribution of markers by female and male coverage for the <i>D. carinata</i> , <i>D. elongata</i> , and <i>D. sublineata</i> samples.	59
Figure 3.5 Manhattan plot of markers from male and female <i>D. carinulata</i> samples aligned against the scaffolded <i>D. carinulata</i> assembly. Scaffolds are ordered from largest to smallest. The second largest scaffold (labeled Chr2 here) is saturated with female biased markers, shown as significantly reduced coverage. The fourteenth and fifteenth largest scaffolds (Chr14 and Chr15) have markers significantly associated with males.....	60
Figure 3.6 Genome-wide repeat content annotated along ordered scaffolds of <i>D. carinulata</i> . Chromosome-level scaffolds are ordered by size, followed by sex chromosomes identified with population genomic data (Fig. 3.5), labeled here as ChrX and ChrY1 and ChrY2.....	61
Figure 3.7 BUSCO synteny between the improved assembly produced here with HiFi and Hi-C data (top) versus the first scaffolded assembly of <i>D. carinulata</i> using Shotgun and 10X Chromium data (Stahlke et al. in prep) (bottom). The HiFi + Hi-C chromosome-length scaffolds are ordered by length (Chr001 - Chr012), followed by the X chromosome, then one scaffold (Scaff129) at the end. The	

Shotgun + Chromium assembly is ordered according to scaffold length. Blocks are sized according to the number of BUSCO genes mapped to that chromosome or scaffold. BUSCO paths are colored according to chromosome assignment in the new HiFi + Hi-C assembly.	62
Figure 3.8 BUSCO synteny across assemblies of <i>D. sublineata</i> , <i>D. elongata</i> , <i>D. carinata</i> , and <i>D. carinulata</i> , ordered from left to right by native range west to east, and colored according to the assigned <i>D. carinata</i> chromosome. The small number of BUSCOs that were not found ('Missing') are shown in the bottom row. Blocks are sized according to the number of BUSCO genes mapped to that chromosome or scaffold. Chromosomes are ordered by size within each species, followed by the X chromosome. <i>D. carinulata</i> was the only species with a twelfth chromosome-level scaffold.	63
Figure 3.9 Coverage of reads from samples of <i>D. sublineata</i> , <i>D. elongata</i> , and <i>D. carinata</i> that did not align concordantly to the <i>D. carinulata</i> assembly. Chromosome-level scaffolds are ordered by size within each species, followed by the X chromosome.	64
Figure A.1 Mean coverage of individuals across populations at targeted loci.	92
Figure A.2 The folded allele frequency spectra for each population before DFTD became prevalent. wuaklina/Mt. William is not presented because it was first sampled in 2004, eight years after DFTD was first described at that locality.	93
Figure A.3 Un-adjusted p-values for all SNPs of each population analyzed with mm (Mathieson and McVean 2013).	94
Figure A.4 Adjusted p-values for all SNPs of each population analysed with mm (Mathieson and McVean 2013).	95
Figure A.5 Un-adjusted p-values for all SNPs across all populations analysed with spatpg (Gompert 2016).	96
Figure A.6 Adjusted p-values (Francois et al. 2016) for all SNPs across all populations analysed with spatpg (Mathieson and McVean 2013; Gompert 2016).	97
Figure A.7 Allele frequency change (Δaf) for each population separately. SNPs in the top 1% are indicated by more opaque points. The threshold line for the top 1% within each population is indicated by a dashed line.	98
Figure A.8 Signatures of selection as detected by mm for each population separately. SNPs in the top 1% are indicated by more opaque points. The threshold line for the top 1% within each population is indicated by a dashed line.	99
Figure A.9 Binned initial allele frequency distributions for DCMS candidates and non-candidates across all populations with samples before DFTD became prevalent.	10000
Figure A.10 Correlations among elementary statistics: afchange = Allele frequency change (Δaf); mm = Mathieson & McVean (2013). Correlations are clustered by similarity along the x-axis.	101

Figure A.11 Histogram of shared MSigDB (Subramanian et al. 2005) gene set overlaps between contemporary and historical candidates, i.e. the signature of ongoing selection, in a permutation test of 1000 draws with replacement. We observed no shared gene set overlaps between contemporary and historical candidates. Of the 1000 permutations, 100% had more overlapping sets than observed.	10202
Figure B.1 Sampling sites and original source population locations from the native range of <i>D. elongata</i> .	10909
Figure B.2 Sampling sites and original source population locations from the native range of <i>D. carinulata</i> .	11010
Figure B.3 Percent of paired reads that mapped concordantly to the <i>D. carinulata</i> genome per sample, according to group. The mean of 72.78% is plotted as a horizontal line.	11111
Figure B.4 Change in likelihood values from Structure results for the global SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.	11212
Figure B.5 Modal structure results for the global SNP dataset for $K=2$, $K=3$, and $K=4$, from top to bottom.	11313
Figure B.6 Structure results for the global SNP dataset across all ten repetitions of $K=4$.	11414
Figure B.7 Histograms of individual q-values (as a hybrid index) from modal $K=4$ Structure (Pritchard et al 2000) results for bi-parental hybrid individuals. Each histogram is symmetric about the 0.50 line and colors indicate the assigned ancestry for the respective cluster. Ordered leading with most abundant pair: A) <i>D. carinata</i> x <i>D. sublineata</i> hybrids ($n=24$) have both extreme (near 0 and 1) and intermediate values suggesting a lack of barriers. B) <i>D. elongata</i> x <i>D. carinata</i> hybrids ($n=3$) and C) <i>D. elongata</i> x <i>D. sublineata</i> hybrids ($n=3$) both have more extreme q-values with one likely back-crossed hybrid (0.25, 0.75).	11515
Figure B.8 Change in likelihood values from Structure results for the <i>D. carinulata</i> SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.	11616
Figure B.9 Change in likelihood values from Structure results for the <i>D. elongata</i> SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.	11717
Figure B.10 Estimated ancestry according to latitude for individuals from sites indicated by color and shape of A) <i>D. sublineata</i> from the global set of Structure results and B) <i>D. carinulata</i> Fukang ecotype from the <i>D. carinulata</i> substructure results. The linear model for each distribution is plotted as a red line with standard error in grey shading.	11818
Figure B.11 Pairwise- F_{ST} compared to haversine distance between respective sites indicate patterns of isolation-by-distance associated with population structure for <i>D. elongata</i> collected from the native range in Greece (Mantel's $r = 0.086$, $p = 0.492$). Crossed-squares indicate pair-wise comparisons	

within ecotype, solid squares across ecotype. A linear model for each distribution projected behind points as a red line with standard error in grey to aid in visualization.....	11919
Figure B.12 Directionality index (ψ) for both <i>D. carinulata</i> ecotypes (origins) in North America. Positive stepwise increasing values suggest a directional founding effect. Each ecotype is outlined by a box, with admixture zones (Fig. 2.3) overlapping.....	12020

Statement of Contribution

In Chapter 1: ARS conceived and designed the comparison of genome-wide contemporary and historical selection, conducted bioinformatics analyses, and led the writing of the study. Collaborators conceived and designed the contemporary selection analyses, conducted field work to collect samples, and performed contemporary devil genomic library data generation.

In Chapter 2: ARS contributed to conception and design of the study, performed field and laboratory work, conducted bioinformatic analyses, and wrote the study. Collaborators assisted with design the study, fieldwork and sampling, genomics laboratory work, and writing.

In Chapter 3: ARS conceived and designed the study, performed bioinformatic analyses, and wrote the study. Collaborators conducted laboratory work and some bioinformatic analyses.

Chapter 1: Contemporary and historical selection in Tasmanian devils (*Sarcophilus harrisii*) support novel, polygenic response to transmissible cancer

Forthcoming in Proceedings of the Royal Society B

Abstract

Tasmanian devils (*Sarcophilus harrisii*) are evolving in response to a unique transmissible cancer, devil facial tumour disease (DFTD), first described in 1996. Persistence of wild populations and the recent emergence of a second independently evolved transmissible cancer suggest that transmissible cancers may be a recurrent feature in devils. Here we compared signatures of selection across temporal scales to determine whether genes or gene pathways under contemporary selection (6-8 generations) have also been subject to historical selection (65-85 million years), and test for recurrent selection in devils. First, we used a targeted sequencing approach, RAD-capture, to identify genomic regions subject to rapid evolution in approximately 2,500 devils in six populations as DFTD spread across the species range. We documented genome-wide contemporary evolution, including 186 candidate genes related to cell cycling and immune response. Then we used a molecular evolution approach to identify historical positive selection in devils compared to other marsupials and found evidence of selection in 1,773 genes. However, we found limited overlap across time scales, with historical selection detected in only 16 contemporary candidate genes, and no overlap in enriched functional gene sets. Our results are consistent with a novel, multi-locus evolutionary response of devils to DFTD. Our results can inform management actions to conserve adaptive potential of devils by identifying high priority targets for genetic monitoring and maintenance of functional diversity in managed populations.

Introduction

Species are subject to selection by pathogens throughout their evolutionary history, shaping lineage diversification and leading to complex cellular and molecular defensive mechanisms (Barreiro and Quintana-Murci 2010). Still, emerging infectious diseases (EIDs) can cause mass mortality and, given sufficient reproduction and genetic variation, initiate rapid adaptive evolution in a naïve host population (Epstein et al. 2016). Although the prevalence and severity of EIDs in wildlife populations is now well-recognized (De Castro and Bolker 2005; Smith et al. 2006; Jones et al. 2008; Lafferty

2009), we are just beginning to understand the evolutionary impacts of disease in wildlife. We have a relatively short recorded history of infectious disease in wildlife, and therefore a limited ability to predict outcomes or intervene when warranted (McCallum and Dobson 1995; Storfer et al. 2020).

High-throughput DNA-sequencing techniques and high-quality annotated reference genomes have revolutionized our ability to monitor and identify mechanisms of evolutionary responses to pathogens (McKnight et al. 2017; Brandies et al. 2019; Storfer et al. 2020). Inter-specific comparisons of non-synonymous and synonymous variation (dN/dS) within protein-coding regions have long been used to identify positive selection at immune-related loci (Li 1993; Yang 1998; Shultz and Sackton 2019). At the population level, rapid evolution in response to disease can be detected by tracking changes in allele frequency before, during, and after the outbreak of disease (Mathieson and McVean 2013; Gompert 2016). Intra-specific comparisons across populations can reveal to what extent the evolutionary response to disease is constrained by limited genetic mechanisms or variation for adaptation (Yeaman et al. 2018). Reduced representation techniques such as restriction-site associated DNA-sequencing (RADseq) (Baird et al. 2008) have made the acquisition of genome-wide, time-series genetic data more accessible in non-model systems (Andrews et al. 2016). By integrating these resources and tests of selection at differing temporal scales, we can assess whether species that show rapid evolution in response to contemporary pathogens also show evidence of historical selection to similar pathogens.

A striking example of an EID acting as an extreme selective force in wildlife is devil facial tumour disease (DFTD), a transmissible cancer first described in 1996 afflicting wild Tasmanian devils (*Sarcophilus harrisi*) (Hawkins et al. 2006). Tasmanian devils are the largest extant carnivorous marsupial, with contemporary wild populations restricted to the Australian island of Tasmania. As a transmissible cancer, DFTD tumour cells are transmitted between hosts, behaving as a pathogen (Pearse and Swift 2006). Transmission typically occurs as devils bite each other during the mating season after devils have reached sexual maturity (Hamede et al. 2008; Hamilton et al. 2019). With few notable exceptions documenting regression (Pye et al. 2016a), DFTD tumors escape recognition, become malignant, and can kill their hosts within six months (Siddle et al. 2013). Starting from a single Schwann cell origin (Murchison et al. 2010), DFTD has now swept across nearly the entire species range (Fig. 1.1A). Devil populations have declined species-wide ~80% (Lazenby et al. 2018) with local declines in excess of 90% (McCallum et al. 2009). Nonetheless, population genomic studies have shown that devils are rapidly evolving in response to DFTD (Epstein et al. 2016; Hubert et al. 2018; Fraik et al. 2020), and DFTD has been spontaneously cleared (i.e., regressed) in some individuals (Pye et al. 2016a). Long-term field studies and simulation modelling have predicted that

cyclical co-existence or DFTD extirpation are more likely scenarios than devil extinction (Wells et al. 2019). This is particularly alarming because devils have notoriously low genome-wide diversity, attributed to climate- and anthropogenic-induced bottlenecks (Brüniche-Olsen et al. 2014; Brüniche-Olsen et al. 2018; Patton et al. 2019). Depleted genetic diversity at immune-related loci has likely further contributed to DFTD vulnerability (Siddle et al. 2007).

Despite transmissible cancers being exceedingly rare across animals, a second independent transmissible cancer in devils, DFT2, was described in 2014 (Pye et al. 2016b; James et al. 2019). Comparative and functional analyses of DFTD and DFT2 showed similar drivers of cancerous mutations and tissue type of origin (Stammnitz et al. 2018). Low genetic diversity, chromosomal fragility (Deakin et al. 2012), a reportedly high incidence of non-transmissible neoplasms (Peck et al. 2019), and injury-prone biting behaviour (Hamede et al. 2013) may contribute to a predisposition to transmissible cancers in devils (Patchett et al. 2020). These findings suggest that transmissible cancers may be a recurring selective force in the Tasmanian devil lineage. If so, this leads to the hypothesis that the genes and genetic pathways associated with the ongoing evolutionary response to DFTD may have experienced recurrent historical selection in the devil lineage from previous transmissible cancers.

Because of the threat of DFTD and DFT2 to devil populations, there are ongoing conservation efforts, including the establishment of a captive devil insurance meta-population. The insurance population is managed to maintain genome-wide genetic diversity and serve as a source for re-introductions in an effort to increase genetic diversity and size of wild populations (Hogg et al. 2017). To inform conservation efforts, it is important to understand what types of genetic variation in natural populations may allow for evolutionary rescue from disease and maintain adaptive potential for future threats (Hohenlohe et al. 2019). Given evidence for rapid evolution in response to DFTD, monitoring of genetic variation at candidate adaptive loci could help evaluate adaptive potential of wild populations (Hohenlohe et al. 2019; Wright et al. 2020). In heavily managed (e.g. captive) populations, loci associated with an adaptive response to disease could be included in genotyping panels for maintaining genetic diversity (Grueber et al. 2019).

Here we identify targets of selection and signatures of adaptation at both contemporary (6-8 generations) and historical (65-85 million years) scales in Tasmanian devils. First we test for evidence of contemporary genomic response to selection by genotyping thousands of individuals sampled at several time points across six populations, using RAD-capture (Ali et al. 2016) to target nearly 16,000 loci (Margres et al. 2018a). Next we identify signatures of historical selection in the devil lineage by comparing across marsupial species with annotated genomic sequence data. Then, we test for

evidence of recurrent selection by examining shared contemporary and historical signatures of selection, in terms of either specific loci, genes or functional genetic pathways.

If transmissible cancer is a novel selective force acting on Tasmanian devils, we expect that genes under contemporary selection by DFTD will be different from those with a signature of historical positive selection. Alternatively, if transmissible cancer is a recurrent selective force in the devil lineage that targets the same set of genes repeatedly, we may expect a conserved response among populations and an overrepresentation of the same genes or pathways under both contemporary and historical selection. However, if there are multiple genetic pathways that could be involved in a response to recurrent transmissible cancers, we may expect a polygenic response across contemporary populations and little overlap between contemporary and historical timescales. These alternatives can inform conservation efforts to manage genetic diversity for resilience in natural devil populations, and any genes or functional pathways that show both contemporary and historical selection may be relevant to cancer resistance more broadly.

Methods and Materials

Contemporary Selection

We used the RAD-capture method (combining RADseq and sequence capture) (Ali et al. 2016) to conduct targeted genotyping of single-nucleotide polymorphisms (SNPs) across 2,562 unique individuals from multiple Tasmanian devil populations, sampled both before and after DFTD appeared in each population (Fig. 1.1A, Table 1.1, Table A.1) (Hamede et al. 2015; Fraik et al. 2020). We constructed RAD-capture libraries following Ali et al (2016), using the restriction enzyme *PstI* and a capture array targeting 15,898 RAD loci selected for membership in one of three functional categories: 1) those showing signatures of DFTD-related selection from previous work (Epstein et al. 2016), 2) loci close to genes with known cancer or immune function, and 3) loci widely distributed across the genome (See Margres et al. 2018a for more details on the development of this array.; Fraik et al. 2020). See Appendix A for multiplexing, read processing, and SNP genotyping details.

To account for the expected high rates of genetic drift within populations, we used a composite statistic to compare signatures of selection across populations. We identified candidate SNPs as the top 1% of a de-correlated composite of multiple signals score (DCMS) (Ma et al. 2015), which combined the results of three analyses: change in allele frequency in each population after DFTD (Δaf), and two methods that estimate strength of selection from allele frequencies at multiple time points in multiple populations, the method of Mathieson & McVean (Mathieson and McVean 2013) (mm), which allows the estimated selection coefficient to vary over space; and $spatpg$ (Gompert 2016), which allows the selection coefficient to vary over time and space. Individuals were assigned

to generational-cohorts based on their estimated years of birth (Table A.1). We estimated Δaf for five locations at which we had sampling both before and after DFTD was prevalent, according to DNA collection date and estimated date of birth, combining multiple cohorts when applicable (Table 1.1; Table A.1). Both time-series methods (*mm* and *spatpg*) incorporate estimates of effective population size, which ranged from 26-37 according to a two-sample temporal method (Jorde and Ryman 2007; Do et al. 2014) (Table A.3). DCMS reduces the signal-to-noise ratio by combining p-values from different tests at each SNP while accounting for genome-wide correlation among statistics. We included SNPs with results from at least eleven of the twelve individual tests (Δaf for five populations, *mm* for all six populations, and *spatpg*) and weighted based on the statistics with results at that SNP. To characterize the role of standing genetic variation in rapid evolution (Kelly and Hughes 2019), we visualized the initial allele frequencies of each population for each analysis of contemporary evolution (Figs. A.2, A.9). We evaluated repeatability among populations by comparing population-specific p-values of Δaf and *mm* with the R package *dgconstraint* for a similarity index called the C-score, where 0 indicates no similarity between populations (Yeaman et al. 2018). See Appendix A for details of each analysis.

Historical Selection

We combined existing genomic resources for the South American grey-tailed opossum (*Monodelphis domestica*) (Mikkelsen et al. 2007) and tammar wallaby (*Notamacropus eugenii*) (Renfree et al. 2011) from the Ensembl database (Aken et al. 2016) and the recently published transcriptome assembly of the koala (*Phascolarctos cinereus*) (Johnson et al. 2018) to identify genome-wide signatures of positive selection in devils, relative to these other species using the branch-site test of PAML (Phylogenetic Analysis by Maximum Likelihood) (Yang 1997, 2007). We compiled alignments of orthologous genes and reduced the marsupial time-calibrated phylogeny of Mitchell et al. (Mitchell et al. 2014) to those species for which annotated full genomes are available (Fig. 1B). The branch-site test compares likelihood scores for two models which estimate dN/dS among site classes of a multi-sequence alignment, allowing dN/dS to exceed 1 (positive selection) in a proportion of sites along a single branch in the alternative model. We reduced the potential for false positives by filtering any putative orthologs with extreme sequence divergence ($S > 2$), measured as the sum of synonymous mutations per gene (*S*), and ensuring alignments of nucleotides were longer than 100 bp (Anisimova et al. 2001; Bielawski and Yang 2005). We identified historical candidates with the likelihood-ratio test, comparing the likelihoods of the alternative and neutral models with one degree of freedom and an $\alpha = 0.05$. Historical candidates were those with estimates of dN/dS > 1 along the devil branch and

FDR > 0.05 after correcting for multiple testing (Shaffer 1995). See Appendix A for details regarding orthology identification and PAML implementation.

Recurrent Selection

We refer to genes under both contemporary and historical selection as candidates for recurrent selection. To test whether genes under contemporary selection differed from genes under historical selection, we first tested for significant overlap with Fisher's one-tailed test. To test for differences in the strength of selection, we compared the distributions of dN/dS and the proportion of sites per gene found under positive selection among candidates for recurrent selection to all other historical candidates from the genome-wide background using nonparametric tests of equality, the Kolmogorov-Smirnov test (Massey Jr 1951), which is more sensitive to the centre of the distributions, and the Anderson-Darling test (Scholz and Stephens 1987), which is more sensitive to extreme values of the distribution and often has more power. To identify and compare key mechanisms of adaptation among candidate genes from each set, we used gene ontology (GO) term enrichment analysis using the SNP2GO package (Szkiba et al. 2014), the PANTHER web-interface (Mi et al. 2018), and in gene sets of the molecular signatures database (MsigDB), using the subset of genes tested for each test as the respective background set (Subramanian et al. 2005). We capitalized on the wealth of ongoing research in devils and DFTD by comparing our contemporary and historical candidates to those previously identified using different datasets and analytical approaches (Epstein et al. 2016; Wright et al. 2017; Hubert et al. 2018; Margres et al. 2018a; Fraik et al. 2020). See Appendix A for details of these comparisons.

Results

Genomic Data

To test for contemporary selection, we sampled a total of 2,562 individuals across six localities of Tasmania before and after DFTD prevalence (Table 1.1; Table A.1; Fig. 1.1A), with a RAD-capture array (Margres et al. 2018a). After filtering, we mapped a total of 517.7 million reads against targeted loci. The mean final coverage of targeted loci was 14.8x, with 76.6% of all samples having coverage of at least 5x (Fig. A.1). After filtering, we retained 14,585 – 22,878 SNPs for downstream analysis, depending on the sampled time point and population.

Evidence for contemporary selection

Among each elementary test for selection signatures, 161 – 232 SNPs (depending on population) were in the top 1% of allele shifts following disease (Δaf), 209 – 217 were in the top 1% of mm scores, and 213 were in the top 1% of spatpg scores (Table A.4, Figs. A.7-8). Across populations and elementary tests for contemporary selection (Δaf , *mm*, *spatpg*), p-values were not correlated

(Pearson's $r < 0.155$ for all tests; Fig. A.10). The computed repeatability indexes for population-specific responses Δaf and mm were $C_{\Delta af} = 4.86$ ($p = 1e-04$) and $C_{mm} = 3.72$ ($p = 1e-04$), which implies a low, but significant level of repeatability (Yeaman et al. 2018). In the top 1% of DCMS scores (≥ 1.167), we identified 144 candidate SNPs for contemporary selection by DFTD; of these, 79 had annotated genes (186 total) within 100 kb (Fig. 2.2; Table A.5). The initial frequencies for candidate SNPs were not skewed toward intermediate frequencies (Fig. A.9).

Comparing our contemporary candidates and those previously identified in devils with selection and genome-wide association analyses (Wright et al. 2017; Hubert et al. 2018; Margres et al. 2018a; Fraik et al. 2020), we found many overlapping genes (discussed below). Notably, we found significant enrichment of candidates previously associated with DFTD-related phenotypes in females (14 genes, $p = 4.2e-08$, Odds ratio=7.3) (Margres et al. 2018a). Gene ontology enrichment analysis found middle ear morphogenesis (GO:0042474) significantly enriched among contemporary candidate SNPs (FDR < 0.05). Five candidate SNPs were within the 100 kb window of two genes associated with this term: EYA1 and PRKRA. Both EYA1 and PRKRA are involved in cell proliferation and migration and implicated in tumour suppression and angiogenesis (Meurs et al. 1993; Wu et al. 2013; Cai et al. 2018).

Evidence for historical selection

Of the 18,788 genes annotated in the devil reference genome, 6,193 had 1-to-1 orthologs in at least three of the four marsupial genomes and an appropriate sequence divergence ($S < 2$). Using the branch-site test for positive selection in PAML, we found a total of 1,773 genes to be candidates for historical positive selection (Table A.6). Estimates of dN/dS spanned the full range of possible values, from 1.05 – 999 and proportion of sites with substitutions per gene ranged from 0.01 – 0.78 (Fig. 1.3). The majority of genes were classified as having a molecular function of binding (GO:0005488) or catalytic activity (GO:0003824); a plurality involved in cellular processes (GO:0009987) or biological regulation (GO:0065007); and a plurality as participating in the Wnt signalling pathway (P00057). None of these pathway classifications were significantly enriched.

Recurrent selection

Of the 186 contemporary candidate genes, 68 had 1-to-1 orthologs among the four marsupials and were tested for historical selection. Sixteen genes showed evidence of historical selection and are thus candidates for recurrent selection (dN/dS > 1 , FDR < 0.05 ; Table A.6). Contemporary candidates were not enriched for historical selection according to Fisher's test (Odds ratio = 0.0, $p = 1$). Among the 16 recurrent candidates, dN/dS estimates spanned 15.7 – 999 and proportion of sites per gene 0.01 – 0.25 (Fig. 1.3). According to the Anderson-Darling and Kolmogorov-Smirnov tests of equality,

neither distributions of dN/dS estimates (Fig. 1.3A; A.D. $p = 0.86$; K.S. $p = 0.58$), nor proportion of sites (Fig. 1.3B; A.D. $p = 0.49$; K.S. $p = 0.48$) differed between candidates for recurrent selection (in black) and historical candidates (in red).

After correcting gene set enrichment for multiple testing (FDR < 0.05) and requiring at least 10 genes in the background set, we did not find functional enrichment of any MSigDB gene sets among recurrent candidates or shared between both contemporary and historical sets. Importantly, the permutation test of shared gene sets found fewer shared between historical and contemporary selection than expected by chance ($p < 0.001$, Fig. A.11).

Discussion

Contemporary Responses to DFTD

Using a targeted set of nearly 16,000 loci, we detected widespread evidence of a response to selection by DFTD across the Tasmanian devil genome. Our results extend previous work that has shown genomic evidence of a response to DFTD in wild populations (Brüniche-Olsen et al. 2016; Epstein et al. 2016; Hubert et al. 2018; Fraik et al. 2020). Here we greatly increased the sample size of individuals and genetically independent populations for greater power, resulting in strong evidence of a response to selection widely distributed across the genome. We found greater similarity across populations within analytical approaches than among methods within populations and relatively low, but significant repeatability across populations. This result is consistent with rapid, polygenic evolution facilitated by selection for standing variation within populations that was present prior to disease arrival. This timescale (3 - 8 generations) would likely be too short for new mutations or migration to play a substantial role in DFTD response, and genetic variation is shared across the species range, despite geographic population structure (Hendricks et al. 2017; Fraik et al. 2020).

In line with previous population genomic studies (Epstein et al. 2016; Wright et al. 2017; Hubert et al. 2018; Margres et al. 2018a; Fraik et al. 2020), our analysis of contemporary evolution detected a putatively adaptive response related to the immune system, cell adhesion, and cell-cycle regulation (Table A.5). Our GO enrichment result for middle ear development (GO:0042474) among contemporary candidates may highlight selection for interactions with the peripheral nervous system and cell proliferation. Genes annotated with nervous system associations may indicate selection for behavioural changes (Hubert et al. 2018), or highlight importance and vulnerability of peripheral nerve repair by Schwann cells in devils, given the prevalence of biting and Schwann cell origin of DFT (Patchett et al. 2020). Significant overlap for genes associated with devil infection status (case-control), age, and survival (Margres et al. 2018a) among our contemporary candidates is a strong indicator that these contemporary candidates likely confer relevant phenotypic change. We also

confirmed five (CRBN, ENSSHAG00000007088, THY1, USP2, C1QTNF5) of seven candidates identified previously (Epstein et al. 2016) in a genome scan for loci under selection from DFTD in three of the same populations (Freycinet, Narawntapu, and West Pencil Pine). In contrast, we identified those five and only two more (TRNT1 and FSHB) of 148 candidates from a re-analysis of that same dataset which studied population-specific responses (Hubert et al. 2018).

Among genes that have been associated with host variation responsible for tumour regression on devils (Wright et al. 2017; Margres et al. 2018b), we found only JAKMIP3, a Janus kinase and microtubule binding protein (Margres et al. 2018b), in our list of contemporary candidates. However, we found devil regression candidates TL11, NGFR, and PAX3, which encodes a transcription factor associated with angiogenesis (Fang et al. 2014); as well as GAD2, MYO3A, and unannotated ENSSHAG00000009195 (Margres et al. 2018b), among population-specific candidates for allele frequency change (Δaf), possibly reflecting differences in test sensitivities. Overall, the paucity of candidates shared between our contemporary analysis and regression studies suggests that regression may not be the dominant form of phenotypic response to DFTD; to date tumour regression has only been detected in a few populations (Margres et al. 2018b) not represented in our study.

Historical selection in the devil lineage

With our genome-wide molecular evolution approach (Yang 2007), we found widespread historical positive selection across the devil genome in about 28% of all 6,249 orthologs tested (Table A.6). The branch-site test is known to be less conservative than related models, particularly when divergence among species is large (Zhang 2004), but the rates of historical selection we found in devils are similar to those described in other taxa; e.g. 23% of genome-wide orthologs among 39 avian species using a similar approach (Shultz and Sackton 2019).

We did not find preferential positive selection for immunity-related genes, as has been shown in primates (Barreiro and Quintana-Murci 2010), eutherian mammals more generally (Kosiol et al. 2008), and birds (Shultz and Sackton 2019). Instead, we found the highest proportion of pathways under historical selection to be functionally classified within the Wnt pathway, a signalling cascade regulating cell adhesion and implicated in carcinogenesis (Nelson and Nusse 2004). As genomic resources grow and improve in marsupials (Brandies et al. 2019), interspecific analyses for positive selection at finer scales may reveal more recent and specific selection targets in Tasmanian devils. Our ability to detect historical selection due to transmissible cancer in devils could be improved by genome assembly efforts among more closely related Dasyuridae, as well as complementary annotation.

Comparing Contemporary and Historical Timescales

Remarkably few transmissible cancers have been discovered in nature (Metzger et al. 2015; Strakova and Murchison 2015), and yet two of those independent clonally-transmitted cancers have been discovered in Tasmanian devils in less than 20 years. This and the observed rapid evolutionary response to disease suggest that transmissible cancers may be a recurrent event in devils. We found no significant overlap of historical and contemporary selection at either individual genes or functional gene sets. This does not rule out the possibility of prior transmissible cancers in devils; but it suggests that if transmissible cancers have been a recurrent feature of devil evolution prior to DFTD, they did not generally impose selection on the same set of genes or genetic pathways that show a contemporary response to DFTD. Nonetheless, the 16 candidate genes showing both historical and contemporary evidence for selection (Table A.7) raise interesting targets for understanding adaptively important variation in devils.

The 16 candidate genes for recurrent selection (Table A.7) are generally related to three main themes: transcription regulation, the nervous system, and the centrosome. Four of these candidates for recurrent selection were previously associated with disease-related phenotypes (Margres et al. 2018a). We additionally found 82 historical candidates previously identified in the top 1% of SNPs associated with disease-related phenotypes with three represented in the top 0.1% associated with large-effect sizes for female case-control and survival (Margres et al. 2018a). This overlap lends support to the hypothesis of recurrent selection by transmissible cancers, but was not significant ($p = 1$, odds ratio = 0). Both our contemporary selection analysis and the genome-wide association study (GWAS) approach used by Margres and colleagues (Margres et al. 2018a) are statistically limited by small populations, sample size, and the time scale over which DFTD-related selection has occurred. By considering the complement of these results together, the overlapping historical, GWAS, and contemporary candidates may still be targets of recurrent selection along similar functional axes, potentially including transmissible cancer.

The low prevalence of candidates for recurrent selection and lack of shared functional gene set enrichment between both contemporary and historical signatures of selection suggest a novel response to DFTD compared to historic selection in the devil lineage. However, there are alternative hypotheses. For example, there could be redundancy in genetic mechanisms underlying resistance to transmissible cancers, potentially as a result of repeated selection for resistance, allowing selection to act across many loci (Casanova and Abel). That is, the low genetic diversity observed in devils could be the result of widespread historical purifying selection resulting from transmissible cancers or other diseases (Guiler 1992), or historical bottlenecks due to climate change and habitat loss (Brüniche-

Olsen et al. 2014; Brüniche–Olsen et al. 2018; Patton et al. 2019), that prevent a response to selection under DFTD at loci that are still associated with disease phenotypes.

The widespread contemporary evolution we found in devils reflects the recent prediction (Miller and Metcalf 2019) that response to an emergent disease is most likely controlled by many genes conferring quantitative resistance (Gandon and Michalakis 2000), for example by reducing the within-host growth rate of tumors. DFTD is predicted to become less virulent in the short-term (Wells et al. 2019; Patton et al. 2020). If DFTD persists long-term in the devil population with ongoing coevolution, it may lead to diversifying selection for specific, qualitative host resistance mechanisms (Miller and Metcalf 2019). Indeed, phylodynamic analysis of DFTD as it spread across Tasmania supports the hypothesis that devils may be mounting a response; transmission rates have decayed such that DFTD appears to be shifting from emergence to endemism (Patton et al. 2020). Although host-genomic variation was not jointly considered in that study, the combined evidence of multiple studies demonstrating rapid evolutionary response of devils to DFTD, including this one, support these interpretations.

Conservation Implications

Calls have been made to consider the historical context of adaptation when proposing conservation management solutions based on genomic results (Kardos and Shafer 2018). Our analysis of historical selection largely supports the hypothesis that DFTD is a newly emerging and novel selective force, distinctly shaping today's remaining wild devils. The targets of novel selection that we identified (Fig. 1.2, Supp Table A.4) and their functional roles should be considered for prioritization of monitoring and conservation in light of DFTD. At the same time, the wide distribution of contemporary candidates across the genome also highlights the importance of standing genetic variation to continue to respond to unique selective forces, including local environmental factors (Fraik et al. 2020). Genomic monitoring could be useful for maintaining both functional diversity at candidate loci and genome-wide variation in captive populations (Wright et al. 2015; Grueber et al. 2019; McLennan et al. 2020) and in the wild. Multiple genomic tools are available for targeted monitoring of large sets of loci (e.g. Campbell et al. 2015; von Thaden et al. 2020) and could be used to track adaptive evolution and potential in the form of genetic diversity (Hohenlohe et al. 2019). However, before management decisions are made for specific genes, further work would need to identify favoured alleles and fitness effects for the genes we identified (Table A.5) .

DFTD has yet to reach devils in the far west (Fig. 1.1A) and continues to circulate throughout the island. To maintain long-term adaptive capacity in the face of similar recurrent selective forces

including DFT2 and potential future transmissible cancers, our results warrant (1) the monitoring of genetic variation in broad functional groups and (2) management strategies to maintain genetic diversity across those broad groups. Although these populations were not subject to DFT2 at the time of writing, an important and interesting future direction should examine the evolutionary response to DFT2 and could compare loci under selection by the two independent transmissible cancers. This study could provide a list of candidate loci for development of a genotyping panel for either purpose, with flexibility to target many or fewer loci. At the same time, given urgent and unpredictable present-day threats including not just emerging diseases but environmental change and population fragmentation, it is important that monitoring and population management also focus on maintaining genetic variation across the genome.

Conclusion

Our results suggest that the contemporary evolutionary response to DFTD is mostly novel compared to the genome-wide signature of historical selection. Comparing the degree of overlap and distributions among contemporary and historical candidates did not support recurrent selection on a common set of genes in response to transmissible cancer. Our work contributes to mounting evidence of possible mechanisms by which devil populations are persisting and rapidly evolving in the face of DFTD despite overall low genetic diversity and population bottlenecks (Brüniche-Olsen et al. 2016; Epstein et al. 2016; Pye et al. 2016a; Margres et al. 2018a; Margres et al. 2020). Broadly, this type of approach can be applied to analyses of novel threats in wildlife populations in the current era of anthropogenic global change to guide monitoring and management actions focused on genetic adaptive potential.

Table 1.1 Number of adults sampled before and after the year of first detection of DFTD at each site. See Table A.1 for sample size for each year at each locality.

Location	Year of First Detection	Samples		Total
		Before	After	
wukalina/Mt. William	1996	0	155	155
Freycinet	2001	300	382	682
Forestier	2004	131	552	683
Fentonbury	2005	99	169	268
West Pencil Pine	2006	52	348	400
Narawntapu	2007	224	150	374
Total		806	1756	2562

Figure 1.1 Map of the six contemporary sampling locations relative to disease prevalence over time (red lines) with the year of first detection labelled at each site. B) Reduced, unrooted time-calibrated phylogeny (Mitchell et al. 2014) of marsupials used to estimate genome-wide historical selection on the devil lineage with estimated divergence times (Ma) indicated along edges. Devil cartoon by David Hamilton. Wallaby, koala, and opossum digital images retrieved from <http://www.shutterstock.com/amplicon>. From top to bottom: The tamar wallaby (*Notamacropus eugenii*), koala (*Phascolarctos cinereus*), Tasmanian devil (*Sarcophilus harrisii*), and South American grey-tailed opossum (*Monodelphis domestica*).

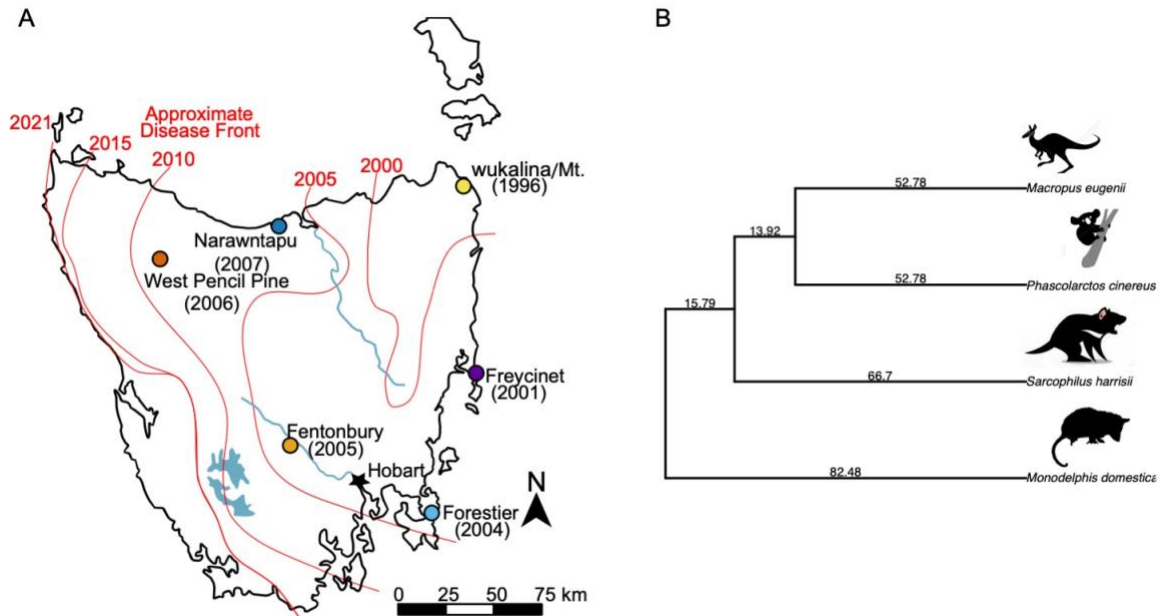


Figure 1.2 Results of each elementary test of contemporary selection across populations and the composite scores for final candidate (filled points) and noncandidate (opaque grey points) SNPs, ordered by chromosome and colored by population when applicable. From top to bottom: Change in allele frequency (Δaf), Mathieson and McVean (mm) (Mathieson and McVean 2013), $spatpg$ (Gompert 2016), de-correlated composite of multiple signals ($DCMS$) (Ma et al. 2015).

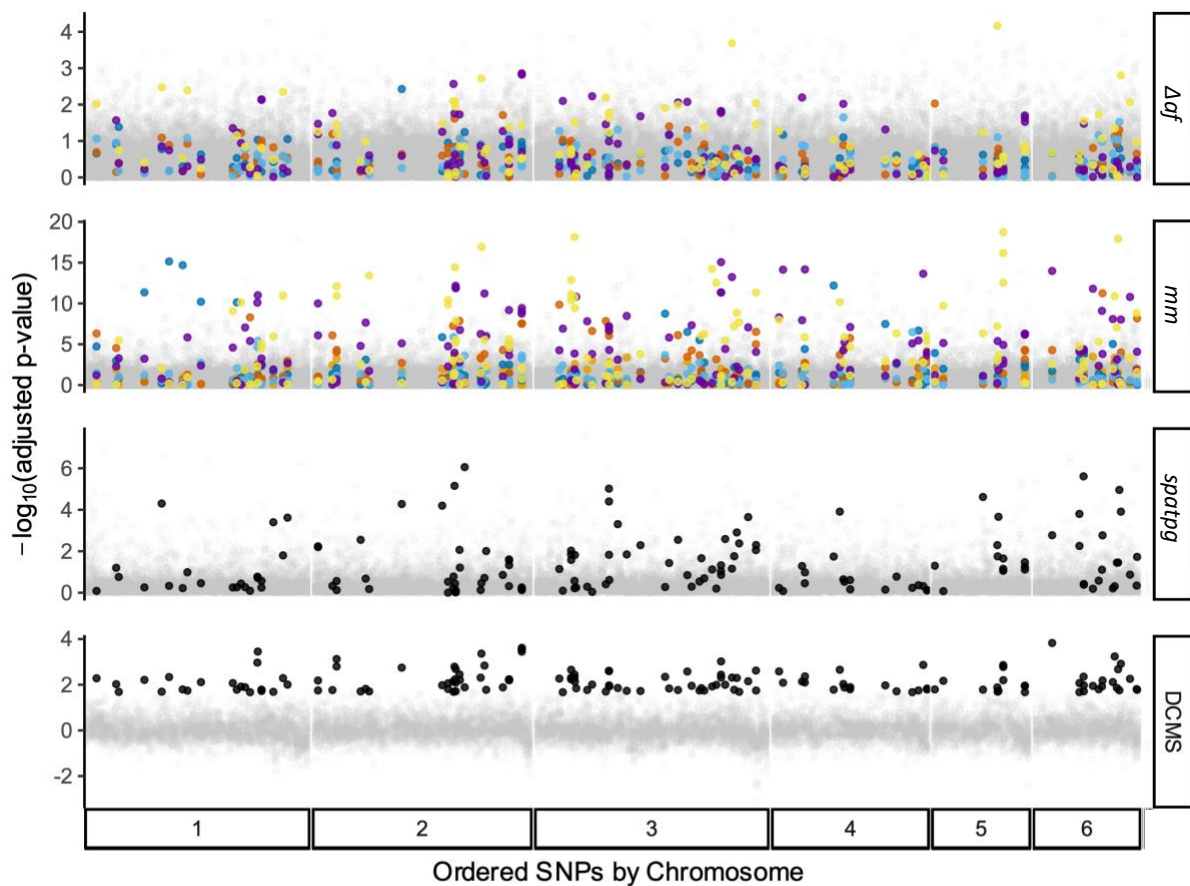
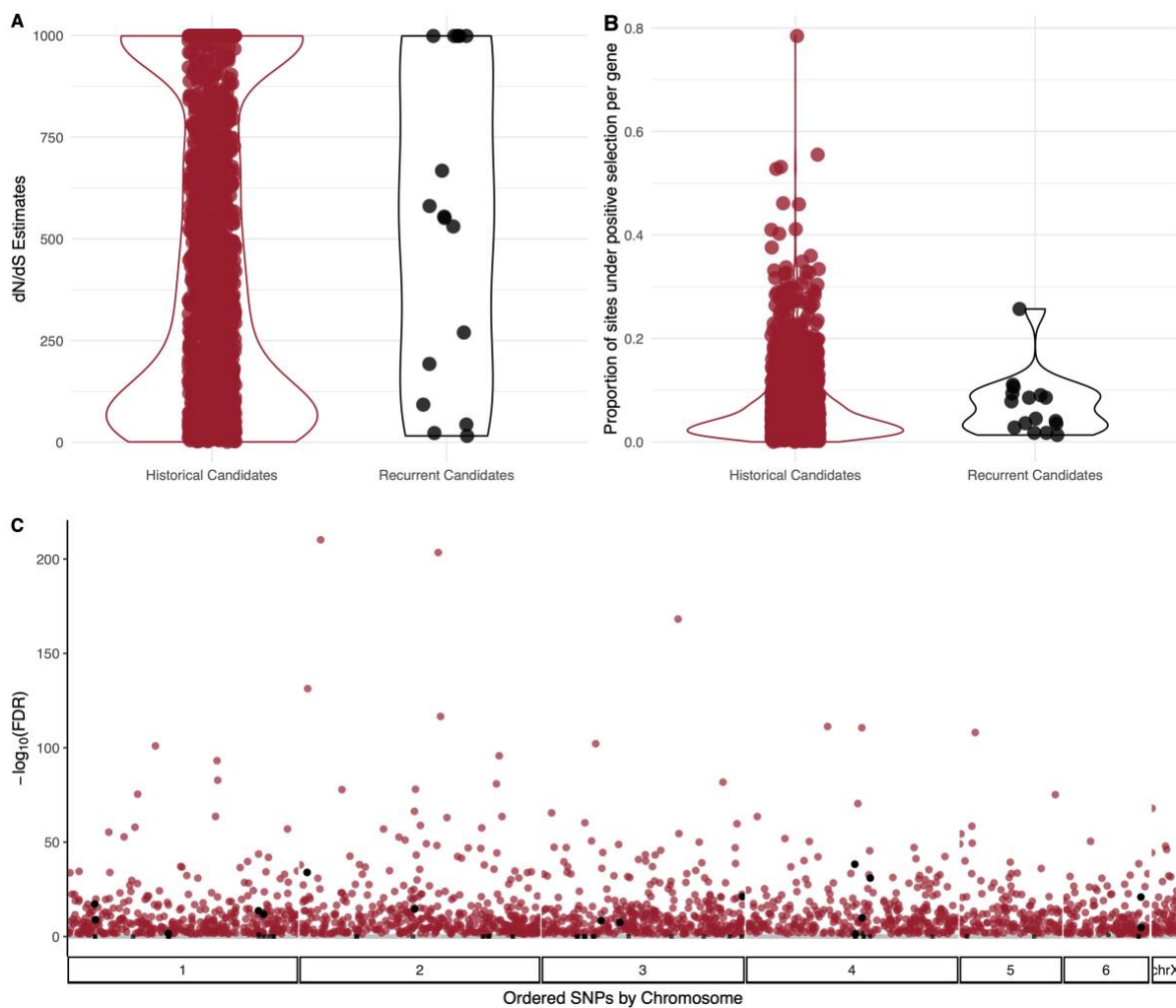


Figure 1.3 A) Estimates of dN/dS and B) the proportion of sites under positive selection (Yang 2007) for historical candidates across the genome-wide background (red; N=1,982) and candidates for recurrent selection (genes with significant results for both historical and contemporary selection; black). Each point represents the respective result for a single gene. C) Distribution of $-\log_{10}(\text{FDR})$ for historical selection across all 6,193 genes tested (gray squares, non-significant at both scales; black squares, significant contemporary and non-significant historical; red circles, non-significant contemporary and significant historical; black circles, significant at both scales).



Chapter 2: Hybridization and range expansion in tamarisk beetles (*Diorhabda* spp.) introduced to North America for classical biological control

Abstract

As planned invasions, classical biological control (biocontrol) agents present important opportunities to understand the mechanisms of establishment and spread in a novel environment. In some biocontrol systems, multiple source populations (ecotypes) or congeneric species are released to increase the chance of ecological matching across diverse habitats. The ability of these biocontrol agents to spread and adapt, and their effects on local ecosystems, depends on genomic variation and the consequences of admixture in the novel environment. Here we examine the genome-wide outcomes of introduction, spread, and hybridization in four cryptic species of a biocontrol agent, the tamarisk beetle (*Diorhabda carinata*, *D. carinulata*, *D. elongata*, and *D. sublineata*), introduced from six localities across Eurasia to control the invasive shrub tamarisk (*Tamarix* spp.) in western North America. We assembled a *de novo* draft genome and applied RADseq to over 500 individuals from laboratory cultures, the native ranges, and across the introduced range. Despite evidence of a substantial genetic bottleneck among samples of *D. carinulata* in North America and low levels of genetic diversity, populations continue to establish and spread, with one *D. carinulata* ecotype dominating the southward expansion front. We confirmed that *D. carinata*, *D. elongata*, and *D. sublineata* hybridize in the field to varying extents, with *D. carinata* x *D. sublineata* hybrids being the most abundant. Genetic diversity was greater at sites with hybrids, highlighting potential for increased ability of hybridized populations to adapt and expand. Our results demonstrate the complex patterns of genomic variation that can result from introduction of multiple ecotypes or species for biocontrol, and the importance of understanding them to predict and manage the effects of biocontrol agents in novel ecosystems.

Introduction

Human-mediated translocations and climate change have reshaped range limits and previous barriers to gene flow on a global scale (Capinha et al. 2015). Understanding the genomic consequences of translocations and admixture between introduced populations is key to both preventing the spread of invasive species and to improving the conservation of threatened species (Roderick and Navajas 2003; Fauvergue et al. 2012; McFarlane and Pemberton 2019). However, the spontaneous nature of

accidental introductions makes it difficult to study the outcomes and consequences of eco-evolutionary processes occurring in these systems, since, among other factors, the introduction history and founding population sizes are unknown. Classical biological control programs (hereafter, biocontrol) are essentially planned, intentional invasions. In a typical biocontrol program, highly host-specific natural enemies (agents) are collected in their native range and introduced into a novel environment to control invasive pests (targets) (McFadyen 1998). Biocontrol systems thus provide an unmatched opportunity to study invasions from a genomic perspective because, compared to their respective target invasive species, biocontrol agents were introduced relatively recently and almost always intentionally, with known source location, introduction localities, and sometimes known introduction population sizes (Marsico et al. 2010). Despite the opportunity biocontrol agents represent, genomic tools have rarely been used to identify and characterize the consequences of founder effects or evolutionary mechanisms contributing to establishment, persistence, range expansion, or rapid evolution in classical biocontrol agents of invasive species (Hopper et al. 2019; Szűcs et al. 2019; Leung et al. 2020; Muller-Scharer et al. 2020; Sethuraman et al. 2020).

Biocontrol scientists have, on several occasions, released several individuals from different locations in the native range, with either known, or inferred, differences in phenotypes. These different populations are referred to as “ecotypes”. Different ecotypes are often released in the hope that some will better match the novel environment (Frick 1970; Room et al. 1981; DeBach and Rosen 1991; Smith et al. 2018). While this practice may increase the chance of ecological matching across a diverse range of target habitat, it also opens the door for novel phenotypes to arise upon hybridization of different ecotypes. Admixture among different populations, and, at an extreme, hybridization between different species, may present the genetic novelty and diversity necessary to overcome the bottleneck imposed by introduction and adapt, but at the risk of yielding undesirable traits or decreases in fitness (Kolbe et al. 2004; Fauvergue et al. 2012; Rius and Darling 2014; Lommen et al. 2017). The outcomes of multiple agent releases have been understudied, while the consequences of hybridization and admixture among divergent biocontrol agents are even less well understood (but see Szűcs et al. 2011; Szucs et al. 2012; Szűcs et al. 2012; Szűcs et al. 2021). Biocontrol efforts could be enhanced if hybridization resulted in increased genetic diversity, providing the raw material for the regional evolution of more efficacious ecotypes, or increased fitness in populations with higher genetic diversity (Tracy and Robbins 2009; Szucs et al. 2012; Szűcs et al. 2012; Bean et al. 2013; Bitume et al. 2017; Szűcs et al. 2021). If hybridization resulted in forms with less host specificity than seen in parental forms, for example, biocontrol safety could be compromised; though there have been no documented cases of evolution in fundamental host range (the range of host species that the agent on which the agent can complete development) (Van Klinken and Edwards 2002). Given the potential

consequences, post-release monitoring of biocontrol programs involving the release of different populations or species is necessary (Hufbauer 2008; McFarlane and Pemberton 2019). Recently developed molecular tools can now also be routinely employed to track biocontrol agents at the population genetics level.

The case of *Diorhabda* spp. (Coleoptera: Chrysomelidae), a leaf beetle released to control invasive woody shrubs of the genus *Tamarix* (hereafter, tamarisk), provides a system in which questions relating to the genomic consequences of invasion can be addressed. Tamarisk is native to North Africa and Eurasia and has become invasive in riparian areas across the western United States and northern Mexico. Stands of tamarisk can form dense monotypic thickets that cause substantial economic and environmental damage including increased fire intensity and frequency (Drus et al. 2013), increased evapotranspiration (Nagler et al. 2014), diminished soil mycorrhizae critical for native plant species (Meinhardt and Gehring 2012) as well as a number of other negative impacts on native flora, wildlife habitat and recreation (Di Tomaso 1998; Zavaleta 2000; Gaskin and Schaal 2002; Shafroth et al. 2005). The large extent of the tamarisk invasion, which is estimated to cover at least 360,000 hectares (Nagler et al. 2011), coupled with the high value of ecologically-sensitive riparian areas and the cost of conventional control which runs in the millions of dollars (US) per project (Knutson et al. 2019), provided impetus for development and implementation of a biocontrol program.

The first agent released for tamarisk biocontrol was the northern tamarisk beetle, *Diorhabda carinulata*, originally introduced in 2001 at eight locations in North America (Fig. 2.1) (DeLoach et al. 2003). Initial field releases of *D. carinulata* established at five northern locations but failed to establish below the 38th parallel in California and Texas (Lewis et al. 2003) leaving many heavily invaded river systems without a biocontrol option. To address this problem and improve ecological matching across the diverse invaded range (Sands and Harley 1980), three additional tamarisk-feeding *Diorhabda* ecotypes were collected from ecologically distinct locations in North Africa and Eurasia and introduced primarily in tamarisk-infested areas of western Texas (Tracy and Robbins 2009; Michels et al. 2013; Knutson et al. 2019). These ecotypes were elevated to species status in a taxonomic revision based in part on morphology of the genital sclerites (Tracy and Robbins 2009). As a result, a total of four closely-related, cryptic species in the genus *Diorhabda* have established in N. America: *D. carinulata* from Fukang, China and Chilik, Kazakhstan; *D. carinata* from Karshi, Uzbekistan; *D. sublineata*, from Sfax, Tunisia; and *D. elongata* from Crete and Posidi Beach, Greece (Tracy and Robbins 2009).

The *Diorhabda* system is one of the few examples in which contemporary evolution has been demonstrated in a biocontrol agent of invasive plants (see also Szucs et al. 2012; Szűcs et al. 2012; Szűcs et al. 2019 for work on *Longitarsus jacobaeae*). Evolution of response to photoperiod signals enabled rapidly southward expanding *D. carinulata* populations to enter diapause in closer synchrony with the seasonal timing of senescence of tamarisk stands growing in more southern and warmer climates, where the growing season is longer than in the north (Dalin et al. 2010; Bean et al. 2012; Hultine et al. 2015). Evolution in diapause induction accompanied faster dispersal than initially expected (Nagler et al. 2014), challenging the conventionally expected constraints to evolution, high genetic drift and loss of genetic diversity due to repeated founder effects at the frontier (Excoffier et al. 2009; Slatkin and Excoffier 2012). However, range expansion of *D. carinulata* may be quite different from the conventional prediction. Genetic diversity could instead be generated and preserved by negative density-dependent dispersal (Birzu et al. 2019) and migration en masse (‘swarming’) common among mobile insects (Sullivan 1981), especially at range expansion fronts as host resources are depleted and aggregation pheromones draw individuals to mating sites (Cosse et al. 2005). In addition to photoperiod, released and range-expanding *Diorhabda* populations encounter myriad novel selective forces, including changes in annual temperature variation, host plant genetic background, and other biotic interactions, many of which are correlated with latitude (Hultine et al. 2015; Lee et al. 2018). The role of genetic diversity in this rapid range expansion has not yet been tested.

Additionally, secondary contact between *Diorhabda* species in North America has likely initiated admixture among this cryptic species complex, uniquely providing a window to the stability of cryptic species upon secondary contact and represent test cases of (non)ecological speciation (Smith et al. 2018). In this case, only *D. carinulata* and *D. carinata* are sympatric in the native range, *D. elongata* is in parapatry with *D. sublineata* to the west and *D. carinata* to the east, and *D. sublineata* and *D. carinata* are the only species completely allopatric (Tracy and Robbins 2009). While no intermediate forms indicative of hybridization among the four *Diorhabda* species were found in beetles collected from the native range, (Tracy and Robbins 2009), laboratory experiments showed that *D. carinata*, *D. elongata*, and *D. sublineata* can readily occur and back-cross with viable eggs. In contrast, hybrids and back crosses between *D. carinulata* and the other three species showed significantly reduced egg viability and male sterility (Bean et al. 2013). Later regional studies described intermediate morphotypes between *D. carinata*, *D. elongata*, and *D. sublineata* in Texas and surrounding states (Michels et al. 2013; Knutson et al. 2019). Hybridization among all *Diorhabda* species may be especially likely and may increase efficacy because the native host plants and the

biocontrol targets (*Tamarix chinensis* and *T. ramosissima*) exist primarily in N. America as hybrids (Williams et al. 2014).

Evaluating the status of *Diorhabda* populations in terms of hybridization, genetic diversity, and range expansion is a high priority in management of the tamarisk invasion. A long-standing goal is to evaluate and enhance *Diorhabda* as a tamarisk control option, as the tamarisk invasion and the expense of controlling the shrub at a regional scale has heightened interest in biocontrol among regional resource managers (Bean and Dudley 2018). Recently, because some native species now utilize tamarisk, including an endangered bird subspecies, the southwest willow flycatcher (SWFL) (*Empidonax traillii extimus*) known to nest in the shrub, a new challenge has been presented to coordinate riparian restoration efforts with declining density of tamarisk brought about by biocontrol with rapidly evolving tamarisk beetles (Sogge et al. 2008; Hultine et al. 2010).

Existing resources to monitor hybridization and range expansion include mitochondrial COI (mtCOI) haplotypes and morphological markers. These conventional and accessible methods have generally been in agreement for species identification of expanding populations (Bean et al. 2013; Ozsoy et al. 2018; Knutson et al. 2019; Ozsoy et al. 2019; Ozsoy et al. 2021), but both morphology and COI can lead to incorrect species assignments when hybridization is common, and neither can accurately quantify proportions of ancestry (Wayne and Jenks 1991; Rieseberg et al. 1993) or be used to identify the genetic-basis of ecologically-relevant traits and inform predictions. Thus, molecular genetic analysis at the whole genome level is critical for a more detailed analysis of field populations to inform management strategies.

Here we characterize the genetic and geographic signatures of establishment, hybridization, and range expansion in introduced populations of the four tamarisk beetle species. We produced a *de novo* draft genome assembly of the northern tamarisk beetle, *D. carinulata*, and used it as a reference for reduced representation genomic sequencing of over 500 individuals from both the native and introduced range. Our primary goals were to (1) identify genetic variation characteristic of the six source populations of the four parental species, (2) quantify prevalence and levels of hybridization in the introduced range, and (3) examine the consequences of population bottlenecks and admixture on genome-wide diversity across broad, landscape-wide expansion fronts and the hybrid zone. We predicted that admixture among populations and hybridizing species would lead to an increase in genetic variation, while isolation of disconnected patches or range expansion would lead to a decrease. We aim here to build the molecular genetic foundation to monitor and predict evolution in introduced *Diorhabda* species and improve our understanding of mechanisms and consequences of range expansion and hybridization in novel environments.

Methods and Materials

Whole genome assembly of D. carinulata

We developed a *de novo* draft genome assembly using adults from an inbred line established from field-collected beetles in Lovelock, NV (40.02°N, 118.52°W), where *D. carinulata*, originally sourced from Fukang, China (44.17°N, 87.98°E), were released in 2001. We sampled reproductively-active males twice from this line, one at the fifth and one at the twelfth generation.

We combined two sequencing approaches for reference genome assembly. First, we extracted gDNA from the head, thorax, and dissected testes of the G5 male and constructed a library for whole-genome shotgun sequencing (WGS) using the NEBNext Ultra II DNA library. This WGS library was sequenced in one lane of a MiSeq platform (Illumina) using v3 reagents to produce paired 300-bp reads, resulting in approximately 10.4 million read pairs. Second, we conducted 10X Chromium sequencing, which produces long-distance synthetically linked reads (Weisenfeld et al. 2017). We isolated high molecular weight gDNA from the dissected testes of the G12 male, using a MagAttract kit (Qiagen). The UC Davis Genome Center prepared a Chromium 10X library with v1 chemistry. This 10X library was sequenced on a lane of HiSeq 4000 at UC Berkeley and resulted in 708.96 million reads with an average length of 139.50 bp following quality and adapter trimming.

We first assembled the WGS MiSeq 250-bp reads from the G5 male *D. carinulata*. We used a windowed adaptive trimmer, Sickle v1.33, to remove adapters and low quality reads from this library (Joshi and Fass 2011) and retained approximately 10.3 million pairs trimmed to an average length of 249.8 bp. We built contigs and scaffolds from these reads with SPAdes v3.7.1 with one iteration of BayesHammer error correction; k-mer values 21, 33, 55, 77, 99, 127; mismatch careful mode turned on; repeat resolution and mismatchCorrector enabled; and coverage cutoff turned off (Bankevich et al. 2012). Then, we incorporated the 10X Chromium synthetic long reads from the G12 male to scaffold these contigs using the ARCS+LINKS pipeline (Yeo et al. 2018). Briefly, we extracted barcodes with the 10X software Long Ranger v2.1.6, aligned these barcoded reads to the SPAdes assembly with BWA-MEM v0.7.17 (Li 2013), then supplied the SPAdes assembly and alignments to the ARCS (v1.0.1) + LINKS (v1.8.5) pipeline in default mode. ARCS uses the evidence of the synthetic linked reads to construct a graph of linkages for LINKS to then resolve phased scaffolds. We assessed quality of this assembly in terms of overall contiguity using QUAST v5.0.2 (Mikheenko et al. 2018), as well as the completeness of single-copy conserved orthologous genes using BUSCO

v5.0.0 with the Insecta database (insecta_obd10) composed of 75 species and 1,367 orthologs (Simão et al. 2015).

Individual genotyping across the native range, introduced range, and lab cultures

Between 2014 and 2017, we field-collected adult beetles near original source collection sites for *D. elongata* and *D. carinulata* in Eurasia, sites of the first releases of all four species in North America, along the *D. carinulata* expansion front in Utah, Nevada, and Arizona, and across the suspected hybrid zone in New Mexico and Texas (Fig. 2.1, Table B.1, Figs. B.1-B.2). To represent genetic variation of source populations of *D. sublineata* and *D. carinata*, we sampled from laboratory colonies, derived from the same original collections used for the introductions, and maintained at the Palisade Insectary, Colorado Department of Agriculture (see (Bean et al. 2007b) for details of laboratory culturing). In total we sampled 566 beetles, from 37 locations and two laboratory cultures, for population genomic analysis. At each site, we sampled beetles from trees within a 1 km radius and limited the collection of beetles to no more than 5 individuals per tree where possible. We could not find adults at 19TX and instead collected third-instar larvae which were reared to the adult stage under laboratory conditions. Individuals for 31NV, 32UT, 33NV, and 34UT were sampled from the second generation of laboratory cultures established from these sites. All samples were adult beetles transferred as live individuals to coolers with dry-ice or immediately to a -80°C freezer. All samples were stored at -80°C until DNA extraction.

To prepare restriction-site-associated DNA sequencing (RADseq) libraries, DNA was extracted from individual beetles using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's protocol. The abdomens of all individuals were removed to avoid DNA of developing embryos, gut microbes, or consumed plant material and allow for later morphological characterization. Samples were treated with 4 µL Rnase A (Qiagen) to eliminate RNA contamination. DNA sample concentration was quantified for each individual by fluorometric quantification (Qubit 2.0 HS DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA).

In total we prepared 634 individually-barcoded RADseq samples across eight single-digest RADseq libraries using the 8-bp restriction enzyme *SbfI* (Ali et al., 2016). Of those 634, 37 samples were replicated individuals to validate bioinformatic parameter choices and mitigate poor-performing barcodes. Adapter-ligated libraries were multiplexed to achieve approximately 69.7 million reads and paired-end sequenced to 150-bp on an Illumina HiSeq 4000 across two lanes (Vincent J. Coates Genomics Sequencing Laboratory, UC Berkeley). Samples from 47TX-52TX, *D. sublineata* and *D. carinata* cultures, and replicates of ten samples from the first round of sequencing were sequenced in

a NovaSeq lane for 80.3 million additional reads. In total, we obtained 149.99 million paired-end reads across all eight RADseq libraries.

We used Stacks 2.5 (Rochette et al. 2019b) to process raw fastq reads, call genotypes, and produce population genetic statistics. First, raw sequencing reads of each library were filtered for PCR duplicates using `clone_filter`. Then reads were de-multiplexed by individual barcode and re-oriented using the `--bestrad` flag in `process_radtags`, allowing for 3 mismatches and discarding reads with low-quality scores (Catchen et al. 2013; Rochette et al. 2019b; Stahlke et al. 2020). Each processed sample was then aligned to our *D. carinulata* draft genome using the `--very_sensitive` flag of `bowtie2` v2.2.9 (Langmead and Salzberg 2012a), and sorted with `SAMtools` v1.9 (Li et al. 2009). We called genotypes for all sequenced individuals together in the Stacks 2 module `gstacks` with the default `maruki_low` model (Maruki and Lynch 2017). Then, we required that retained sites were present in the majority of all samples (`-R 50`) and extracted a random SNP from each ordered locus. Individuals were further filtered to retain those with $>4x$ effective coverage and $<75\%$ missing genotypes with `VCftools` v0.1.16 (Danecek et al. 2011). We removed 82 individuals from the dataset with this filter. Finally, genotypes were called again, variant calling format (vcf) files generated, and population genetic summary statistics evaluated using `Stacks populations`. We refer to the final catalog of SNPs comprising all 552 individuals as the global dataset.

Source Population Ancestry and Hybridization

We used `Structure` v2.3.4 (Pritchard et al. 2000) to characterize population structure and admixture. Given that founding populations were from distinct sources across Eurasia and the potential for rapid range expansion to lead to dramatic differences in allele frequency among sites, we used the uncorrelated allele frequency model and allowed the alpha parameter to be inferred for each population (Falush et al. 2003). For each K from 1 to 10, we executed 10 independent runs, allowing a burn-in period of 10,000 steps and 10,000 Markov chain Monte Carlo replicates, and printed the estimation of 90% credible intervals. We used `PopHelper` 2.3.0 (Francis 2017) to visualize results and characterize the posterior probability across values of K . After assessing global ancestry assignment with `Structure`, we checked for relationships between missing genotype rates, alignment rates and individual species ancestry assignment (Fig. B.3).

We compiled published data (Hudgeons et al. 2007; Knutson et al. 2012; Michels et al. 2013; Knutson et al. 2019; Pratt et al. 2019) and grey literature describing original releases (directly from the native range) and redistribution efforts (translocations from original release localities) to guide our inference of ancestry assignment (Table B.2). Then we used the `Structure` ancestry assignments of parental species from lab cultures, single source population release sites (e.g., Fukang ecotype *D.*

carinulata at 1WY and Chilik ecotype at 34UT) (Table B.2), and the native range (46CH and 37CR-43GR) to guide our ancestry inference at the remaining localities and quantify hybridization in the remaining samples. Using those diagnostic samples, we examined the confidence intervals across independent runs to conservatively identify the threshold at which ancestry could be confidently inferred, $q = 0.067$ (i.e. the lower-bound of the 90% credible interval), below which admixture identification could be unreliable and due to technical biases (Caniglia et al. 2020). Finally, we visualized the distributions of q-values ('hybrid index') for pairs of inferred ancestral taxa to estimate the degree of back-crossing among pairs (McFarlane and Pemberton 2019).

To characterize the distribution of established and range expanding source populations within *D. carinulata* and *D. elongata* (i.e., population substructure), we constructed population maps for individuals that had *D. carinulata* or *D. elongata* ancestry (respectively) above the 0.066 threshold, then re-filtered SNPs as above in the Stacks populations module. We then re-ran Structure for each subset of individuals. This secondary analysis also provided a check for sensitivity of population genetic statistics and ancestry assignment to unbalanced sampling in the global Structure analysis (Meirmans 2019).

We were interested in preliminarily testing whether the distribution of *Diorhabda* species ancestry or ecotypes showed evidence of local adaptation or environmental filtering related to latitudinal variation, which serves as a proxy for photoperiod, temperature, and host-genotype variation. To test for a relationship between ancestry and latitude, we constructed linear models of *D. carinulata* source population ancestry and *D. sublineata* ancestry within the hybrid zone predicting the relevant q-value from latitude.

Genomic Diversity

We quantified differentiation among all individuals within and across sites using population-based measures of genomic differentiation. We used the populations module of Stacks to calculate π (nucleotide diversity), F_{IS} (the inbreeding co-efficient), and private alleles. We present π and F_{IS} calculated from only variant sites, and private alleles were counted among all sites (variant and invariant).

To test whether genetic diversity results (π , F_{IS} , and private alleles) were statistically different between hybrids versus pure species and native versus introduced, we performed a one-way ANOVA in R. We grouped samples within localities according to Structure results using the threshold described above. Native range populations included 46CH, 43GR, 44GR, 41 CR, 37CR, 39CR, and 38CR. We excluded laboratory colony populations from this analysis.

To quantify isolation-by-distance (IBD), we calculated the distances between the latitude and longitude coordinates of each sampling locality using the haversine formula and constructed a pairwise distance matrix (Sinnott 1984). Then, using the *vegan* package (Oksanen et al. 2019), we conducted a Mantel test with 999 permutations on pair-wise F_{ST} and distance matrices (Hutchison and Templeton 1999) for the following groups: (a) introduced *D. carinulata*, (b) native range *D. elongata*, and (c) the suspected hybrid zone. After characterizing IBD, we tested for a linear relationship between latitude of collection site and Chilik ancestry for the *D. carinulata* SNP dataset and, separately, *D. sublineata* ancestry using the global SNP dataset.

To examine the consequences of rapid range expansion in *D. carinulata*, we tested for the signal of asymmetrical range expansion with two origins (1WY and 34UT) across all of introduction sites of *D. carinulata*. Although 1WY is not the source population for *D. carinulata* released in Nevada, Colorado, and Utah, it is the best representative population for Fukang releases (Table B.2). We filtered genotypes in the *Stacks* populations module for only those populations. We used the *rangeExpansion* R package v0.0.0.9000 to estimate the strength of founder effects, the fit of the predicted range expansion dynamic, and the directionality index (ψ), which is a measure of directional clines in allele frequency created by successive colonization events (Peter and Slatkin 2013).

Results

Genome assembly

Using SPAdes, we first obtained 90,101 scaffolds with an N50 of 52.957 KB, an L50 of 1,917, a total length of 382.446 MB, a BUSCO score of 83.9% Complete single-copy orthologs of the 1,367 genes in the BUSCO Insecta set, (86.2% of those represented a single time and 0.7% duplicated), 1.8% fragmented, and 14.3% missing. We improved this assembly by incorporating the 10X Chromium synthetic long reads with the LINKS+ARCS hybrid *de novo* genome assembly approach. This *de novo* draft *D. carinulata* reference genome was composed of 84,491 scaffolds with an N50 of 720.04 Kbp from 103 scaffolds (L50), a total length of 382.502 MB, and more complete, single-copy orthologs according to BUSCO with 98.2% complete, (97.4% single, 0.8% duplicated), 1.1% fragmented, and .7% missing.

SNP Genotyping

We constructed a total of nine bestRAD libraries containing over 552 individuals across lab cultures, the native range, and the introduced range. We removed an average of 39.52% reads identified as PCR duplicates across libraries and retained or recovered a total of 86.28 million reads after initial filtering and demultiplexing. Individual reference alignment rates against the *de novo* draft assembly

of *D. carinulata* were 72.77% on average, but we noted two distinct peaks near 75% and 100% that reflected species assignments (Fig. B.3). Effective coverage averaged 24.1x after removing individuals with coverage < 4x. In the global SNP dataset, 153,453 loci were merged across paired-end reads for an average locus length of 854.16 bp (std. error = 2.70) and a total of 2,247,632 nucleotide sites. We retained 1,457 SNPs with an average of 8.1% missing data. In the subset of samples with *D. carinulata* ancestry, we retained 2,629 SNPs across 182 samples. In the subset of samples with *D. elongata* ancestry, we retained 1,766 SNPs across 99 samples.

Source Population Ancestry and Hybridization

Structure analysis suggested differential establishment, range expansion, and admixture among all six source populations. We present first the results from the global dataset consisting of all samples from all collections. The greatest change in likelihood, (i.e., the Evanno method; (Evanno et al. 2005), occurred at $K=2$, splitting *D. carinulata* from the other three introduced *Diorhabda* spp. and hybrids (Figs. B.4-B.5). At $K=4$, change in likelihood values plateaued (i.e. the ΔK method; Fig. B.4) and matched species identities for lab cultures, native range, and isolated release sites. The modal ancestry assignments for $K=4-8$ were identical across reps after aligning clusters, i.e., no additional clusters were recovered (Fig. B.6). Therefore, we used cluster assignments from $K=4$ to infer parental species and hybrid ancestry.

Using a threshold of $q=0.067$ for inferring ancestry from a parental species, we found pure ancestry (no interspecific admixture) in 179 *D. carinulata*, 176 *D. sublineata*, 69 *D. carinata*, 93 *D. elongata*. We did not find evidence of hybridization among laboratory cultures or in the source populations for *D. carinulata* (46CH) and *D. elongata* (44GR-41CR). Most of the hybridization was found at sites near the Pecos River (28NM, 27NM, 26NM, 21NM, 20NM, and 18TX), near where all four species were released (Figs. 1-2, Table B.2) (Knutson et al. 2019). We identified 32 likely hybrids in the suspected hybrid zone: 24 *D. carinata* x *D. sublineata*, three *D. carinata* x *elongata*, three *D. elongata* x *sublineata*, one *D. carinulata* x *D. sublineata*, and one triad hybrid between *D. carinulata* x *D. carinata* x *D. sublineata*. Ranges of q -values for *D. carinata* x *D. sublineata* hybrids included extreme and intermediate values (Fig. B.7A); whereas the other hybrid pairs had ranges less than 0.25 or greater than 0.75 (Fig. B.7B and C). The two putative *D. carinulata* hybrids were from 6KS (G5_rep) and 28NM (G48_rep). G48_rep was assigned largely to *D. sublineata* ancestry with $q_{sublineata} = 0.912$, while G5_rep was assigned to tri-specific ancestry with $q_{carinulata} = 0.664$, $q_{sublineata} = 0.071$, $q_{carinata} = 0.265$ (Fig. 2.2, Table B.3).

We further investigated hierarchical substructure among previously identified ecotypes within *D. carinulata* and *D. elongata* (Fig. 2.3), indicated within the global structure results in a minority of

runs (Fig. B.5). We found that the Chilik ecotype, first introduced in 34UT, was the dominant ecotype represented along the Virgin River expansion front (Figs. 1 and 3A). Individuals from sites 5CO and 11CO in eastern Colorado showed admixture with the Chilik ecotype (Fig. 2.3A), likely due to anthropogenic movement of individuals from western Colorado, which was colonized by individuals from southeast Utah (Table B.2). We detected *D. carinulata* farther east than its suspected range in 6KS, 28NM, and 18TX and assigned to Fukang ancestry (Fig. 2.3A). Some hybrid individuals showed ancestry from a third cluster (other *Diorhabda* species; Fig. 2.3).

We found highly significant linear associations with large residuals between latitude and both *D. sublineata* in the global SNP dataset (Adjusted $R^2 = 0.433$; $p < 0.01$) (Fig. B.9A) and the Chilik ecotype in the *D. carinulata* (Adjusted $R^2 = 0.239$; $p < 0.01$) (Fig. B.9B), demonstrating that latitudinal variation may have influenced differential establishment and spread among genotypes.

Genome-wide diversity

We found significant differences in genetic diversity metrics among hybrid, pure, and native range populations. Nucleotide diversity (π) was significantly greater at sites with hybrids (mean = 0.0863) than in either the sites with pure individuals (mean = 0.0335) or from the native range (mean = 0.0381), supporting the prediction that hybridization could increase genetic diversity (Fig. 2.4). The number of private alleles was greater among sites within the native range (mean = 17.86) than in introduced sites with pure individuals (mean = 7) and hybrids (mean = 7), supporting a bottleneck upon introduction for *D. carinulata* and *D. elongata* (Fig. 2.4). Sites with hybrids also had a significantly higher average F_{IS} value (mean = 0.157) than either the sites with pure individuals (mean = 0.0454) or within the native range (mean = 0.027; Fig. 2.4), consistent with recent hybridization and/or assortative mating among species. All results were significant at $\alpha = 0.05$ (Table B.4).

Taking genetic diversity results into a spatial context, we characterized the relationship between divergence and geographic distance among collection sites. Overall evidence for IBD was weak (Fig. 2.5): Mantel's $r = 0.4481$ ($p = 0.1011$) for *D. carinulata*, Mantel's $r = 0.08$ ($p = 0.49167$) for native range *D. elongata*, Mantel's $r = 0.5108$ ($p < 0.005$) across the suspected hybrid zone. Only the suspected hybrid zone had a positive IBD signal significant at $\alpha = 0.05$. Pairwise comparisons across introduced *D. carinulata* ecotypes and native range *D. elongata* had greater F_{ST} values and were further apart (Fig. 2.5, Fig. B.10).

We detected asymmetric range expansion and estimated two distinct origins for *D. carinulata* ecotypes (Fig. B.11; $p < 0.01$). Origin 1 was estimated to be near 1WY (44.86°N, 108.18°W), Origin 2 near the Hoover Dam (35.97°N, 114.64°W), and the origin of their union in central Nebraska

(42.63°N, 101.08°W). The directionality index (ψ) was greatest between 32UT (an introduction site of the Chilik ecotype) and 1WY (a Fukang introduction site) at 0.144, supporting two distinct introduced populations, and the least non-negative between 2CO and 18TX (0.0142). Orienting ψ from greatest to least generally reflected suspected expansion fronts within and among the two source populations (Fig. B.11).

Discussion

To better understand the genetic consequences of admixture and range expansion in the *Diorhabda* biocontrol system, we built a crucial foundation by describing the current distribution of species and ecotypes in the introduced range, contact zones, admixture among populations, and genome-wide diversity. We discuss these results in the context of eco-evolutionary processes, highlight implications for the tamarisk biocontrol program in North America, and discuss opportunities for further study of this system to improve our understanding of contemporary evolution and biocontrol of invasive plants.

Distribution of parental taxa and hybridization

We found clear evidence for differential establishment of introduced *Diorhabda* populations and spread of these beetles from release sites across the introduced range. On one hand, among many original release sites and along suspected colonization routes, individual ancestry assignments were largely composed of respective source populations, suggesting that the extant populations at these release sites remained stable for several generations and naturally expanded along riparian corridors in a predictable pattern. The Fukang release in Lovell, WY (1WY) appeared stable, with uniform ancestry represented (Fig. 2.3A) and a lack of inbreeding (Fig. 2.4) despite being relatively isolated. In terms of spread, ancestry assignment (mostly Chilik) and range expansion signatures supported natural spread from 34UT to 12AZ along the Virgin R. corridor. On the other hand, the eco-evolutionary mechanisms driving differential establishment and spread among *Diorhabda* populations will require further study. For example, it is unclear why very little *D. elongata* ancestry was detected in N. America despite many releases in Texas and New Mexico (Fig. 2.1). The appearance of the *D. carinulata* Fukang ecotype in Kansas (6KS), New Mexico (28NM), and Texas (18TX) (Fig. 2.3A) was surprising given the lack of establishment reported previously (Bean et al. 2012), although releases of that population did occur near there (Fig. 2.1). We have preliminary evidence of environmental filtering or adaptation among *Diorhabda* source populations, with genomic clines forming for the *D. carinulata* Chilik ecotype and *D. sublineata* along latitudinal gradients. However, the distributions are confounded by release history (Fig. 2.1, Table B.2), and the large residuals of this model due to the presence of the *D. carinulata* Fukang ecotype at southern latitudes of NM and TX

(Fig. B.9B), suggest that specific genetic variation, rather than broad ecotype identity, could be important.

Differential rates of hybridization among *Diorhabda* spp. suggest that there may be several outcomes of multiple closely related taxa released together, depending on reproductive barriers, at least. Hybridization frequency in the introduced range supported predictions of (non)ecological speciation theory (Czekanski-Moir and Rundell 2019) based on the native range distributions of species in Eurasia (Tracy and Robbins 2009), that reproductive isolation maintained by geography (allopatric *D. sublineata* and *D. carinata*) is more likely to break down upon secondary contact than in species that are sympatric in their native range which have evolved isolating mechanisms (*D. carinata* and *D. carinulata*). Although we were not surprised to find hybrids between *D. carinata*, *D. elongata*, and *D. sublineata* (Bean et al. 2013; Bitume et al. 2017; Knutson et al. 2019), the high abundance ($N=24$) and intermediate distribution of individual q -values among *D. carinata* x *D. sublineata* hybrids in particular suggest a lack of reproductive barriers or even increased fitness relative to parental species, while the distributions of q -values among the other pair-wise hybrids were extreme and could indicate pre- or post-zygotic barriers to hybridization (Fig. B.7) (McFarlane and Pemberton 2019). In contrast, it is surprising to find *D. carinulata* hybrids ($N=2$) with any of the other species, because laboratory crosses found little to no mating success among those pairs; although *D. carinulata* males could reproduce with females of the other species (Bean et al., 2013). Sex-biased asymmetry of hybridization could be examined by employing existing mtCOI efforts with genome-wide ancestry analyses like these (Petit and Excoffier 2009; Ozsoy et al. 2018; Ozsoy et al. 2019; Ozsoy et al. 2021). We did not recover mtDNA loci in this dataset. The elevated F_{IS} observed in localities with multiple ancestry (Fig. 2.4; e.g., 6KS and 28NM with both ‘pure’ samples and hybrids) reflects that parental species were still detected as partially isolated, sympatric populations, not a panmictic population, in the introduced range (i.e. the Wahlund effect) (Waples 2015). Examining the mechanisms that contributed to speciation of *Diorhabda* in the native range (e.g., divergent environments, ecological interactions, sexual selection) and the role of those barriers in the introduced range present would improve our understanding of the stability of cryptic species broadly and in biocontrol (Rundle and Nosil 2005; Fišer et al. 2018; Smith et al. 2018).

Genomic Diversity during Range Expansion in D. carinulata

The genomic basis of rapid range expansion in *D. carinulata* provides yet another example of the long-held ‘genetic paradox of invasions’, that a reduction of genetic diversity during introduction does not preclude establishment and spread (Estoup et al. 2016). While we found a clear signature of population bottleneck by comparing the number of private alleles between samples from the native

range (46CH) to those in N. America (Luikart et al. 1998), neither the number of private alleles, nor nucleotide diversity (π) declined along range expansion fronts (Fig. 2.4). The weak signal of asymmetrical range expansion (Fig. B.11) and lack of significant IBD (Fig. 2.5) suggests that, at least at the time of sampling, population expansion was not unidirectional or smoothly distributed across the landscape, possibly reflecting anthropogenic-mediated dispersal. Alternatively, these results could support a role for negative density-dependent dispersal in preserving genetic diversity. If large census and effective population sizes are maintained by this mechanism, it could facilitate rapid local adaptation in novel environments and evolution along expansion fronts by assortative mating for co-dispersers (Burton et al. 2010).

To better understand range expansion in this biocontrol system and others like it, a different approach that incorporates anthropogenic movement, spatial distribution of habitat, and informed dispersal parameters would likely yield a more informative result regarding the genomic mechanisms, routes, and impacts of range expansion. These rangeExpansion results indicated suggested a contact point for the two introduction sites in Nebraska, far outside the range geographic possibilities, likely due to the Fukang *D. carinulata* observed in Kansas (6KS) and Texas (18TX) (Fig. 2.3B). Similarly, several of our sampled sites in the range expansion analysis were not connected by natural dispersal even though they were most genetically similar. We could build upon this work with biologically-realistic simulations (Haller and Messer 2019; Landguth et al. 2020) and approximate Bayesian computing (Estoup and Guillemaud 2010) to more accurately assess genomic mechanisms and consequences of rapid range expansion. For example, remote sensing of *D. sublineata* defoliation and expansion has shown that tamarisk continuity and area width predict dispersal distance along a riparian corridor (Ji et al. 2017), whereas rangeExpansion assumed a continuous habitat and natural dispersal (Peter and Slatkin 2013). Further work investigating the ongoing range expansion of *D. carinulata* should also examine possible roles for few loci of large effect (Dlugosch et al. 2015) and plasticity (Bay et al. 2017) operating in the evolution of CDL in *D. carinulata*.

Eco-evolutionary processes influence the impacts of biocontrol

Evolutionary processes have long been of interest in the field of biocontrol (e.g. Simmonds 1963), and has included efforts to mitigate potential negative effects of losing genetic diversity by augmenting population sizes, and of avoiding ecological mismatch through introducing individuals from deliberately targeted locations in the native range. Despite this long interest, only in recent years are the potential consequences of eco-evolutionary processes on the success of biological control programs being acknowledged and explored (Szűcs et al. 2019). Here we provide further evidence that more detailed and nuanced information, including genomic data, can help us better understand

the eco-evolutionary processes occurring among introduced biocontrol agents. Our work specifically documents population expansion despite a dramatic genetic bottleneck, differential establishment and spread among source populations, and differential admixture among those populations.

Widespread hybridization may have implications for both the safety and efficacy of *Tamarix* biocontrol in North America. Considering that previous laboratory experiments with hybrids created in the lab showed changes in phenotypes related to both fecundity and host-preference (Bitume et al. 2017), the abundance of admixture we observed warrants further study. These laboratory results have not been verified in the field, and host preference was measured to the third generation (Bitume et al. 2017). A recent regional decline in *Diorhabda* population density, and extirpation from some previously occupied areas, has been noted in Texas, Oklahoma, and Kansas (Knutson et al. 2019). This underscores the possibility that hybrid breakdown could compromise fitness of *Diorhabda* in the field, an interpretation consistent with the observation that up to 57% of field-collected hybrids displayed abnormal genitalic sclerites (Knutson et al. 2019). Testing of these traits in field-collected, genotyped populations is critical to better understand changes in risk or efficacy of the biocontrol program due to hybridization. Our dataset could be used to develop a panel of markers for more rapid and cost-effective identification of hybrids with targeted sequencing (e.g., RAD-capture, GTseq) (Meek and Larson 2019; Reid et al. 2020).

Our population genomics approach presents a much-needed tool to monitor biocontrol releases of multiple populations and cryptic species, highlighted by a notable discrepancy between morphological analyses (Knutson et al. 2019). The laboratory cross found by Bitume et al., (2017) to be the most fecund relative to parental types, *D. carinata* x *D. sublineata*, were the hybrid pairs we found to be most abundant and widely distributed here, but they were not previously described in the morphological analysis of hybridization in this region. One possible explanation is that the sampling design of Knutson *et al.*, (2019) did not include sites farther north into New Mexico, where the bulk of our *D. carinata* x *D. sublineata* hybrids were found. However, we also found some of these hybrids in Texas, in close proximity to the locations sampled by Knutson *et al.*, (2019). Our dataset could be used to improve the accuracy of morphological hybrid identification by validating morphological markers (Padiál et al. 2010; Griffin et al. 2020).

Our inferences regarding the mechanisms and impacts of range expansion and hybridization are currently limited because *Diorhabda* collections were sometimes transported without detailed documentation regarding population sizes or population sources. Therefore, we cannot determine based on these data alone whether any hybrid genotype or ancestry combination is more successful without more complete records. For example, we know that many *D. carinulata* releases were made

in New Mexico (Table B.2), but the source, precise release sites, precise release numbers, and establishment rates are largely unknown. In general, the practice of anthropogenic movement, often undocumented within management units, presents an interesting tradeoff. On one hand, this increases the availability and likely efficacy of biocontrol across users; but on the other, it makes it more difficult for biocontrol research to understand the patterns of range expansion and adaptation. Evolutionary biologists, biocontrol scientists, and the stakeholders of target invasive species would be well-served with improved records and catalogs of genetic material from biocontrol agent source populations, releases, and follow-up monitoring. Currently there is no entity charged with genomic monitoring of biocontrol releases and these efforts rely on short-term funding and idiosyncratic academic-governmental relationships for each biological system.

The application of genomic approaches in biocontrol systems has the potential to improve both our understanding of contemporary evolutionary processes and management of invasive species (Roderick and Navajas 2003; Szűcs et al. 2019; Leung et al. 2020; Muller-Scharer et al. 2020; Sethuraman et al. 2020). The draft assembly of *D. carinulata* is one of only two currently available reference genomes (Bouchemousse et al. 2020) of biocontrol agents of invasive plants, both in the family Chrysomelidae, subfamily Galerucinae, an inviting opportunity for broader comparative work. The resource we developed here is unique in that it can be used for four intentionally released biocontrol agents (the four *Diorhabda* spp.), compared to *Ophraella communa*, which was not intentionally released and is spreading adventively (Müller-Schärer et al. 2014). Nonetheless, only two genomes for any biocontrol agents of invasive plants represents a substantial missed opportunity considering that there are over 332 established invasive plant biocontrol agent species worldwide (Schwarzlander et al. 2018). Further development of the *D. carinulata* reference genome, including annotation, would greatly improve our ability to identify SNPs, structural variants, and genes associated these traits of interest. Still another opportunity is presented by coupling genomic studies of biocontrol agents with genomic resources from the invasive pests (Lee et al., 2018) to examine co-evolutionary interactions (Sun et al. 2020). Our results provide a baseline timepoint upon which we can build to further disentangle the mechanisms of rapid range expansion and consequences of hybridization in *Diorhabda*.

Figure 2.1 Solid black points indicate sampling localities in western North America at original release sites and along expansion front of *Diorhabda carinulata* (squares), and across original release sites within the suspected hybrid zone of *D. carinata*, *D. elongata*, and *D. sublineata* (closed circles). Transparent filled points indicate known field-releases (Knutson et al 2019; Table B.3) for each respective species. Rivers are drawn in blue.

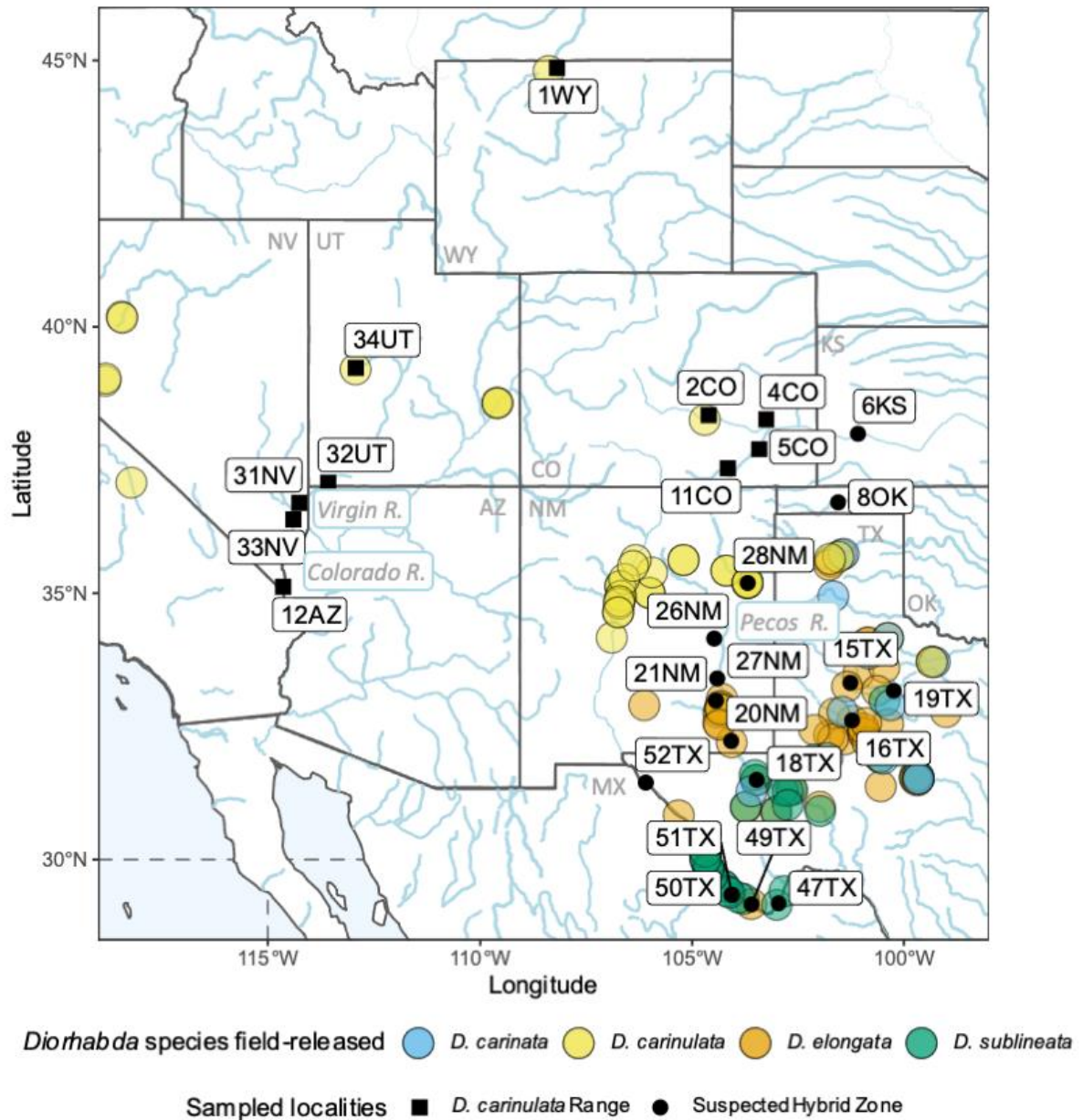


Figure 2.2 Genetic clustering using Structure (Pritchard et al 2000) for $K=4$ across all species. Each individual sample is represented by a bar. Individuals are grouped by collection site and ordered by localities of Table B.1. Groups from left to right are (A) *D. carinulata* Fukang source collection, then north to south from original release sites along expansion front; *D. carinata* and *D. sublineata* lab cultures and native range *D. elongata* collected in Greece follow. (B) Individuals collected within the hybrid zone of North America, from north to south.

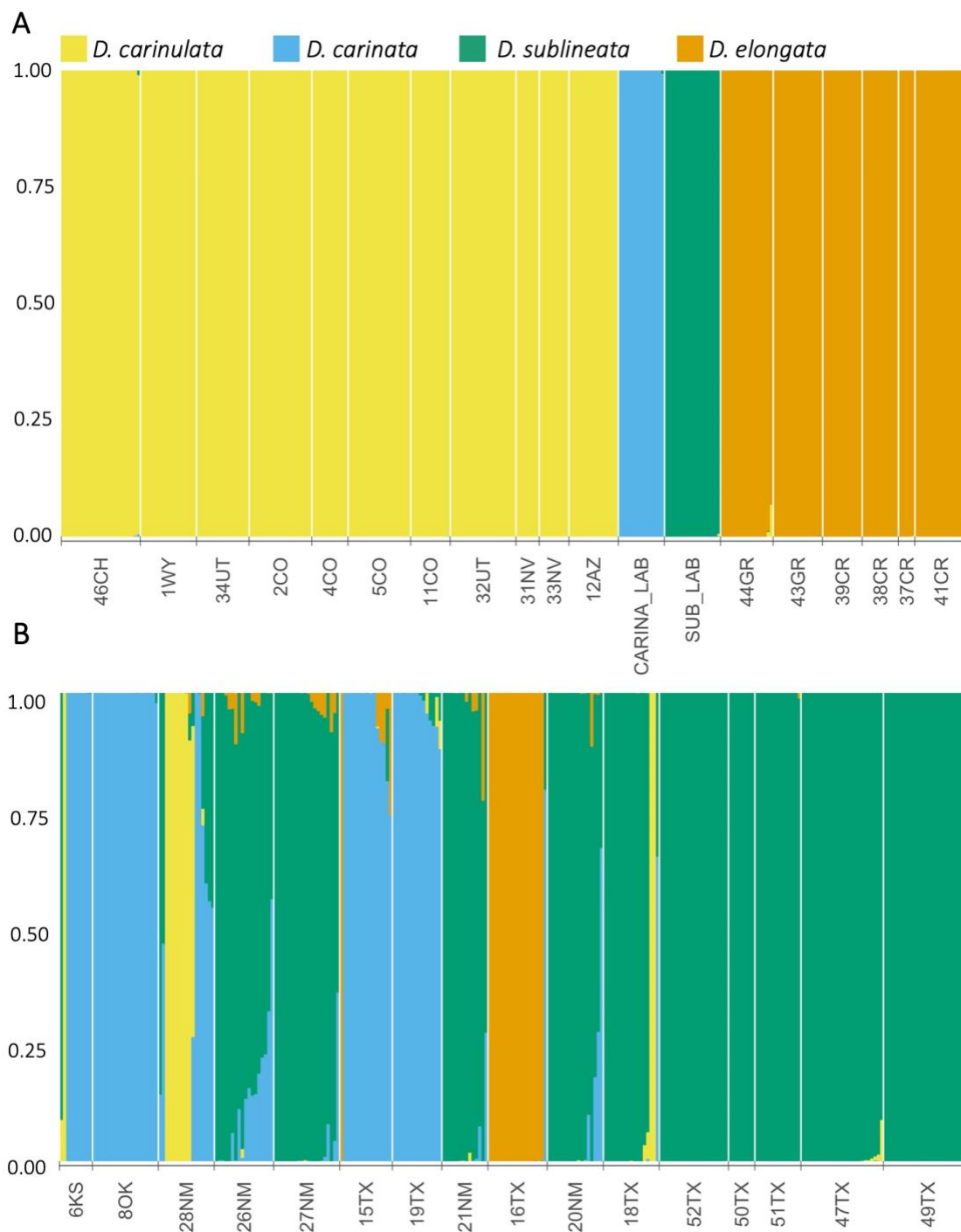


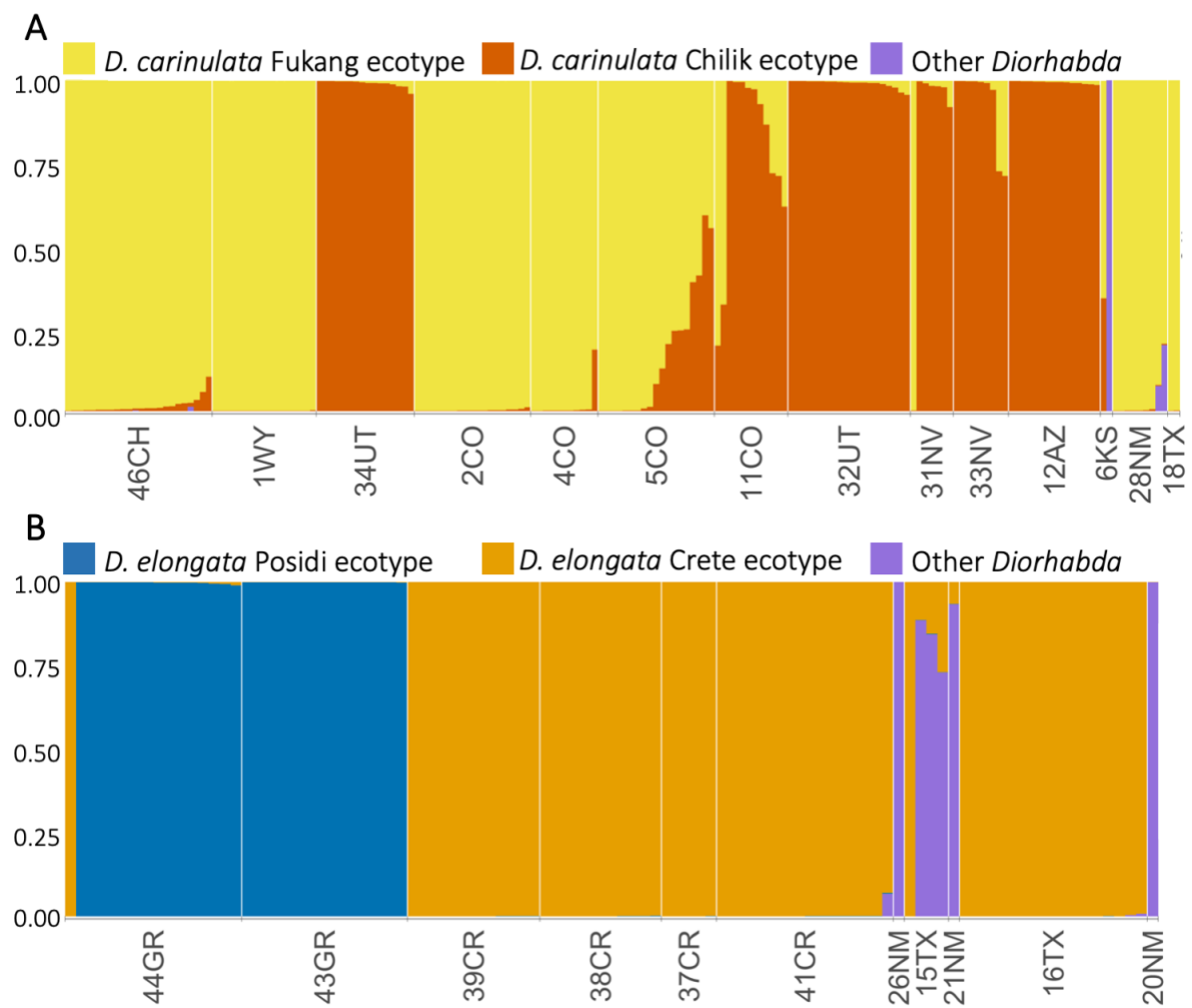
Figure 2.3 Genetic clustering at K=3 within populations of (A) *D. carinulata* and (B) *D. elongata*.

Figure 2.4 Population genetic statistics, π (top panel), the number of private alleles (middle), and F_{IS} (bottom) for the global SNP dataset, consisting of all individuals collected at each locality and grouped (from left to right) by collections within the *D. carinulata* range, laboratory cultures, the native *D. elongata* range, and those within the suspected hybrid zone (ordered as in Figure 2.2). Shading of individual bars indicates whether the collection was an original release site, site within the native range, natural colonization site, or lab culture (from darkest to lightest). Bars represent +/- the standard error for the respective statistic.

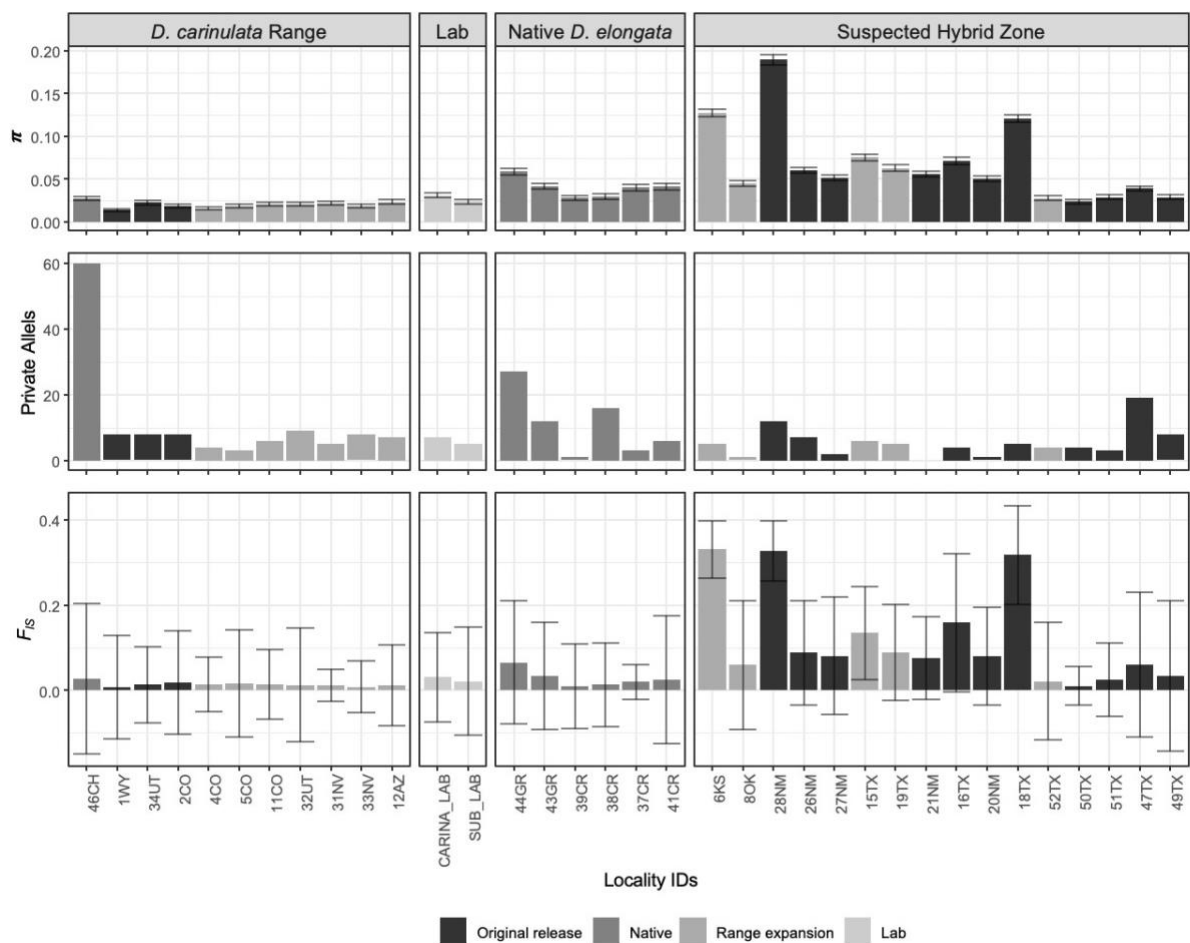
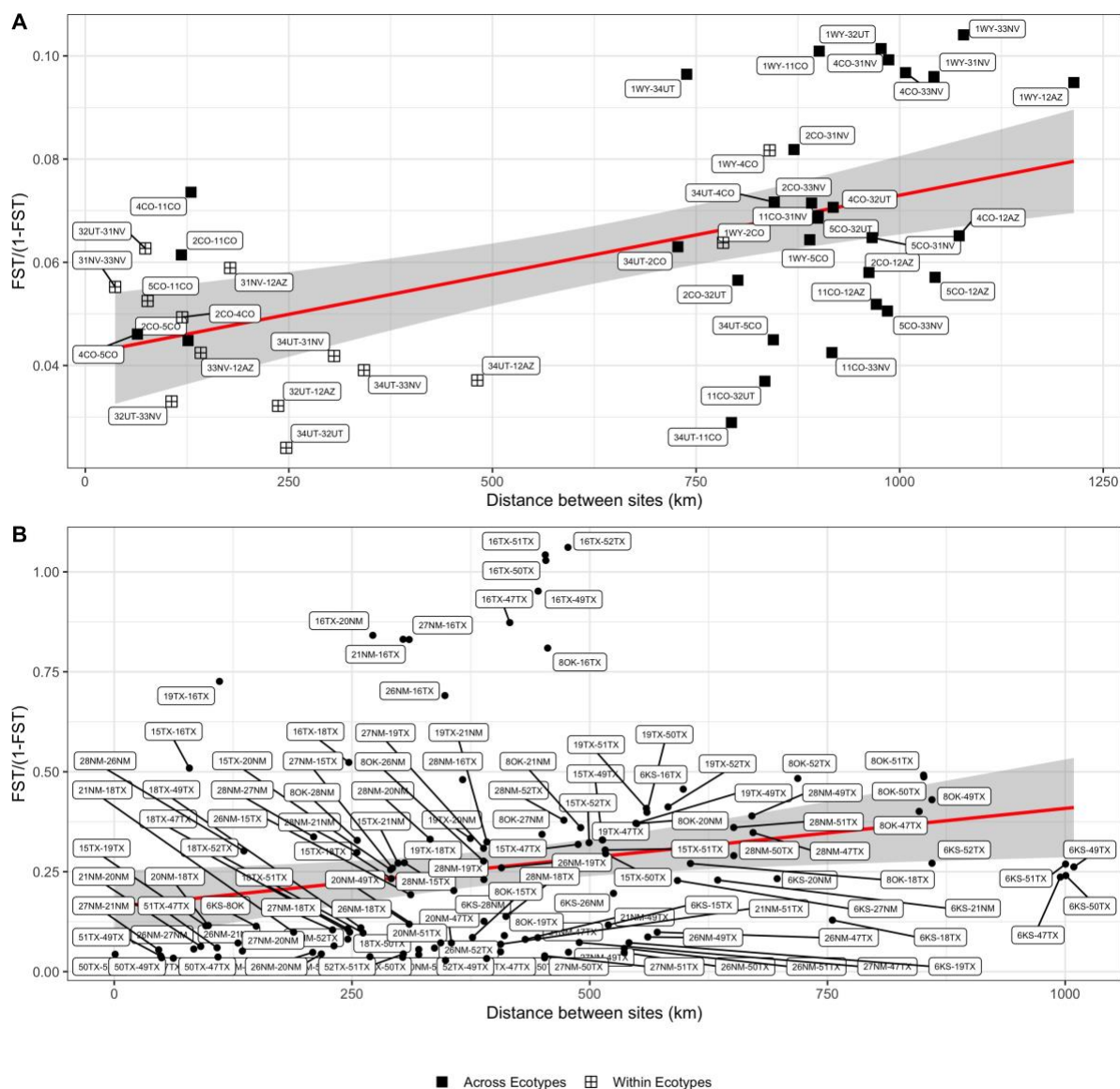


Figure 2.5 Pairwise- F_{ST} compared to haversine distance between respective sites indicate patterns of isolation-by-distance associated with population structure for (A) *D. carinulata* collected in North America (Mantel's $r = 0.4481$, $p = 0.1011$ and (B) across the potential hybrid zone (Mantel's $r = 0.5108$, $p < 0.005$). Crossed-squares indicate pair-wise comparisons within ecotype, solid squares across ecotype, and filled circles for the hybrid zone. A linear model for each distribution projected behind points as a red line with standard error in grey to aid in visualization.



Chapter 3: *De novo* chromosome-level genome assemblies and synteny analysis of four cryptic tamarisk biocontrol agent species: *Diorhabda carinulata*, *D. carinata*, *D. elongata*, and *D. sublineata*.

Abstract

Cryptic species of classical biological control agents (BCA), natural enemies from the native range of a target pest introduced for control in the invaded range, offer unique and important opportunities to examine the outcomes of anthropogenic translocations and secondary contact among closely related species. Unfortunately, genomic resources for any BCA are lacking, and none exist among closely related species. Here we generated PacBio HiFi and Illumina Hi-C data from a single male of each of four cryptic BCA species, the tamarisk beetles *Diorhabda carinata*, *D. carinulata*, *D. elongata*, and *D. sublineata*, notable for documented rapid evolution and hybridization in the introduced range that could impact the biocontrol program. Our approach yielded high-quality, nearly chromosome-level assemblies for all four species. The 90th percentile of scaffolds according to length was between 4.257 - 18.376 MB and included 12-14 chromosome-level scaffolds. Total assembly lengths were in agreement with flow-cytometry estimated sizes between 417-481 MB. We annotated repeat content and conserved single-copy orthologs, provided relevant gene evidence with RNAseq data from *D. carinulata*, and incorporated existing population genomic data to identify sex chromosomes. All four *Diorhabda* assemblies were annotated with about 50% repeat content which could indicate a role for transposable elements in modulating rapid evolution and species boundaries. Likely telomeres, centromeres, and, in *D. carinulata*, male sex chromosomes were annotated with especially high repeat content. Chromosomal synteny analysis according to conserved single-copy orthologs revealed chromosomal homology among *D. carinata*, *D. elongata*, and *D. sublineata*, contrasted by likely rearrangements between *D. carinulata* and the other species, reflecting the current frequencies of hybridization in the introduced range. Together these assemblies and annotations provide insight to the outcomes of secondary contact among cryptic species and possible mechanisms of rapid evolution.

Introduction

Classical biological control agents (BCA) are organisms intentionally introduced from the native range of a target invasive species for control in the invaded range (McFadyen 1998). With known source populations, well-characterized host-ranges, intentional releases, and follow-up monitoring, BCA present excellent and important systems for a range of eco-evolutionary questions (Szűcs et al. 2019; Muller-Scharer et al. 2020; Sethuraman et al. 2020). Biocontrol agents that are cryptic species, i.e., recently diverged taxa with little morphological differentiation (Bickford et al. 2007), have been recently recognized for their important role in offering control within distinct host and environmental niche space (Smith et al. 2018); however, the stability, i.e., species coexistence, of cryptic BCA in the introduced range has been rarely studied.

Genome architecture plays a fundamental role in determining the outcomes of secondary contact by shaping the landscape of recombination at the chromosomal scale (Noor et al. 2001). For example, loci with suppressed recombination, such as the Y chromosome, or regions of rearrangement, can broaden the effects of genes involved with reproductive barriers due to extended linkage (Rieseberg 2001). Among phytophagous insects, exemplar natural systems of ecological speciation (Funk et al. 2002), chromosomal rearrangements have been attributed to the evolution of host-choice and phenology (Feder et al. 2003; Dowle et al. 2017). Similar studies in closely related BCA of invasive plants ('weed biocontrol') could improve the predictability of secondary contact. Unfortunately, genomic resources for biocontrol insects of invasive plants are lacking: At the time of writing, there is only one publicly available reference genome for biocontrol agents of invasive plants (Bouchemousse et al. 2020). This dearth presents a missed opportunity, considering that 462 invasive plant biocontrol species have been released worldwide (Schwarzlander et al. 2018), several of them closely related, cryptic species (Smith et al. 2018).

Until recently, high-quality genome assemblies were largely inaccessible for budget- and sample-limited organisms, such as biocontrol agents. Small chitinous organisms present challenges for high-quality assemblies simply due to limitations in available tissue and accessibility of high molecular weight extraction from a single individual. New low-input libraries and long reads with single individuals are now accessible and revolutionizing insect genomics (Kingan et al. 2019). The PacBio Sequel II system now offers a large increase in yield relative to the required loading mass per single molecule real-time (SMRT) cell and increased accuracy from circular consensus sequencing (CCS) by the PacBio. Hi-C data, which capitalizes on chromatin contacts within chromosomes to reveal long-distance spatial proximity (Rao et al. 2014), can be used to orient and scaffold long contigs generated by PacBio data for chromosome-scale assemblies (Dudchenko et al. 2017). Consortia are capitalizing on these technologies to create fast, accurate, and efficient pipelines (Lewin et al. 2018).

Among these initiatives, the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) Ag100 Pest Initiative has developed pipelines to provide high-quality reference genomes to accelerate genomics research in insect pests (Childers et al. 2019). There is an opportunity to apply this approach to address the dearth of resources in biocontrol agents of plant pests.

An inviting test-case to apply the Ag100 genome assembly approach for the study of invasive plant biocontrol agents is presented by four cryptic species of the genus *Diorhabda* (Chrysomelidae), *D. carinata*, *D. carinulata*, *D. elongata*, *D. sublineata*, released to control the invasive riparian shrubs of the genus *Tamarix* (DeLoach et al. 2004). The first population introduced to N. America ca. 2001, *D. carinulata* from northern China (44.17°N; 87.98°E), did not successfully establish at southern latitudes due to premature diapause induction (Lewis et al. 2003; Herrera et al. 2005; Bean et al. 2007a). Additional populations were introduced from distinct regions across Eurasia occupying different climatic niches and preferences for various *Tamarix* spp., and later delimited as cryptic species (Lewis et al. 2003; Tracy and Robbins 2009). Following introduction, laboratory experiments indicated that hybridization could readily occur between *D. carinata*, *D. elongata*, and *D. sublineata*, and to a far lesser degree between *D. carinulata* and the others (Bean et al. 2013), despite a lack of any intermediate morphology in the native range (Tracy and Robbins 2009). Morphological analysis of field-collected individuals between 2014 and 2018 suggested widespread hybridization where all four species were released (Knutson et al. 2019). By 2011, the timing of diapause induction had evolved in *D. carinulata*, allowing rapid southward range expansion (Bean et al. 2012).

The importance of the role of contemporary evolutionary processes in shaping the biocontrol program and restoration outcomes has motivated ongoing genetic work in *Diorhabda* (Bean et al. 2013; Ozsoy et al. 2018; Ozsoy et al. 2019; Ozsoy et al. 2021). Recent genomic analysis of 1.4K SNPs aligned to a draft *de novo* assembly of *D. carinulata* (Stahlke et al. in prep) showed widespread hybridization between these three, but especially between *D. sublineata* and *D. carinata*, a cross which is known to result in increased fecundity (Bitume et al. 2017). Regarding the rapid range expansion of *D. carinulata*, this approach found a substantial bottleneck in allelic diversity comparing native range samples to those in the introduced range and low levels of nucleotide diversity, challenging the conventional expectations of the role of genetic diversity in limiting rapid evolution and supporting the paradox of invasions (Excoffier et al. 2009; Slatkin and Excoffier 2012). To further investigate the outcomes of hybridization and mechanisms of rapid range expansion, improved resources are necessary.

Here we present high-quality *de novo* genome assemblies of each *Diorhabda* species introduced for biocontrol in N. America, using *D. carinulata* as an anchor upon which to compare the other three species. With this in mind, we generate and assemble RNAseq data to annotate *D. carinulata*. We efficiently produce chromosome-scale phased assemblies from a single individual of each species, using low-input libraries to generate contigs from PacBio HiFi reads and, from the same individual, Illumina Hi-C reads for scaffolding. Then we annotate repetitive elements and incorporate existing population genomic data to identify sex chromosomes and regions of divergence between *D. carinulata* and the other species. To examine the role of genome structure in maintaining species boundaries upon secondary contact, we quantify the degree of synteny with conserved single-copy orthologs. In addition to an improved *D. carinulata* reference genome, we discuss new insight and opportunities to investigate the role of genome structure presented by these high-quality resources.

Methods

Sample Collection

For genome assembly, we used a single snap-frozen adult male beetle of each species from the respective culture maintained by the Palisade Insectary, Colorado Department of Agriculture. The *D. carinulata* culture was held for two generations from collections made in Delta, UT (39.23°N, 112.93°W) from *Tamarix ramosissima/chenensis*, where *D. carinulata* were from Chilik, Kazakhstan (43.6°N 78.25°E) released in 2001 (Carruthers et al. 2008). The *D. carinata* culture was established from individuals collected in Karshi (Qarshi), Uzbekistan (38.86°N, 65.72°E) from *Tamarix* (spp.) and has been kept in lab culture since 2002. The *D. elongata* culture was held for two generations from collections made near Cache Creek, CA (38.72°N, 121.80°W), where *D. elongata* were released in 2004 from collections made in Sfakaki, Crete, Greece (35.42°N, 24.69°E) on *Tamarix parviflora*. The *D. sublineata* culture was established from individuals collected in Sfax, Tunisia (39.23°N, 112.93°E) from *Tamarix* (spp.) and has been kept in lab culture since 2008. Unless specified, methods for each step of the genome assembly and annotation from these samples are the same for all species. Expected genome sizes of males and females from each *Diorhabda* species were estimated by flow-cytometry of propidium iodide-stained nuclei (Johnston et al. 2019).

Genome Assembly

HiFi & Hi-C library preparation and sequencing

High molecular weight DNA was extracted from the heads of individuals. To optimize the size and consistency of fragment length for accurate CCS HiFi reads, we used the Diagenode Megaruptor (Lakha et al. 2016), targeting 12-18 KB fragments in length. Each library was sequenced on a single 8M SMRT Cell on the Sequel II System with the Sequel II Sequencing Kit. *D. elongata* and *D.*

sublineata samples were both 50:50 on two cells for a combined single cell each. Then, we used the thorax of the same individual to generate Hi-C data for scaffolding.

Following individual barcode de-multiplexing, raw PacBio subreads were converted to fastq using bamtools v2.5.1 (Barnett et al. 2011). Then we removed PacBio adapters using a custom BLAST (McGinnis and Madden 2004) script against a database of sequences including the PacBio internal control sequence, blunt adapter, and C2 primer. We generated a k-mer distribution (k-mer length=21) of the filtered CSS reads with Jellyfish2 v2.2.9 (Marçais and Kingsford 2011) and supplied the k-mer histogram to GenomeScope (Vurture et al. 2017) for a complementary estimate of genome size, heterozygosity, and unique sequence.

To assemble the PacBio HiFi reads, we used the efficient, accurate, phased assembler Hifiasm v0.14 (Cheng et al. 2021) on these filtered CSS reads to produce the primary and alternate contig assemblies. Using the any2fasta tool (<https://github.com/tseemann/any2fasta>), we converted the primary and alternate haplotype resolved unitig graphs to fasta format.

We extracted, assembled, and annotated mitochondrial contigs using MitoFinder (Allio et al. 2020) from each resultant primary contig assembly using previously published *Diorhabda* mitochondrial genomes as the seeding reference (Stahlke et al. 2019). Based on phylogenetic relationships among *Diorhabda* (Bean et al. 2013), we used the published mitochondrial genomes of *D. carinata* (MK359256) as the reference for *D. carinata*, *D. elongata*, and *D. sublineata*; and *D. carinulata* (MK359257) for *D. carinulata*.

Scaffolding contigs with Hi-C

Raw paired-end Hi-C reads were aligned to the primary contig assembly using BWA-MEM (Li 2013) with the -5SP options employed to retain the chromosomal capture information of the paired reads. We removed PCR duplicates with samblaster (Faust and Hall 2014) and filtered the alignment to retain only primary alignments (-F 2316) in the resultant bam, using samtools (Li et al. 2009). We used the matlock tool provided by PhaseGenomics to convert the filtered bam to linked juicer formats and sort alignments (matlock.py), then evaluated the sorted, aligned Hi-C data to the respective contig bam (hic_qc.py). Finally, we visualized the Hi-C contacts heatmap and manually edited each candidate assembly with Juicebox (Durand et al. 2016) for translocations and suspicious fragments before exporting the final scaffolded assembly of each *Diorhabda* species.

Assessment

We assessed the quality of each assembly in terms of contiguity and length and completeness of conserved orthologous genes. We quantified contig and scaffold length distributions with bbtools

v38.79 (Bushnell 2014). We assessed completeness of each assembly by searching for conserved single-copy orthologs with BUSCO v5.0.0 (Benchmarking Universal Single-Copy Orthologs), (Simao et al. 2015; Waterhouse et al. 2017) using the MetaEuk gene prediction tool kit (Levy Karin, Mirdita, and Söding 2020) and the insecta_odb10 database, composed of 1,367 genes from 75 insect species.

Genome Annotation

Repeat Content

We constructed *de novo* lineage-specific repeat libraries for each species using the RepeatModeler pipeline v2.0.1 (Flynn et al. 2020). We classified the resultant repeat libraries and created a genomic features file (GFF) for assembly annotation with RepeatMasker v4.0.6 (Smit et al. 2015). We used the custom python script FindTelomeres.py (<https://github.com/JanaSperschneider/FindTelomeres>) to search for canonical telomeric repeats (TTAGG/CCTAA) within the first and last 500 bp of each scaffold.

Sex Chromosomes

To identify sex chromosomes for each species we first used existing population genomic data (Stahlke et al. in prep) and RADsex v1.1.3 (Feron et al. 2020), to efficiently compare the association between RADseq locus presence and phenotypically sexed individuals across RADseq loci. These individual samples were previously assigned to species ancestry using the *D. carinulata* draft reference genome and a balanced reference panel of each species (Stahlke et al. in prep). For sex marker identification here, we divided the population genomic samples into pure *D. carinulata* samples and those with some combination of ancestry from *D. carinata*, *D. elongata*, and *D. sublineata* from three different localities. Unfortunately, we did not have any pure, sexed samples available due to widespread hybridization in this region (Stahlke et al. in prep). Then we used RadSex v1.1.3 to identify markers with sex bias using the `distrib` and `signif` commands with default settings. To infer sex determination type and distribution of markers along assembled scaffolds, we visualized the results with the R package SexGenomicsToolkit (Feron 2021).

We extracted the sex-biased scaffolds we identified in *D. carinulata* associated with males (see Results) and used the lastal aligner of last v869 (Frith and Kawaguchi 2015) (Frith and Kawaguchi 2015) to find highly similar sequences in the other *Diorhabda* genomes. For this, we constructed a lastal database from each query reference fasta, then retained alignments meeting E-values less than 0.05. We used the last post-mark tool to discard alignments composed of only simple repeats (Frith 2011).

Reference-guided transcriptome assembly of *D. carinulata*

Species-specific gene evidence for submission to the automated NCBI RefSeq genome annotation platform was derived from an RNAseq experiment designed to later characterize diapause induction in *D. carinulata*. Adult beetles were collected in Fillmore, UT (39.12°N, 112.74°W) on September 17, 2017 and reared for a generation in the lab in reproductive conditions (16:8 Light:Dark) to reduce maternal effects. Virgin adults from the next generation (G2) were used to construct two families with one male and three females. After five days of oviposition, the next generation (G3) was reared at 16:8 and prepared for a switch. After adult eclosion, male-female pairs were formed in individual cups. On Day 5, we harvested three males for RNA isolation, then subjected three other pairs to a diapause-inducing photoperiod treatment (11:13). On Day 10 (five days after the switch), we harvested 3 males each from 16:8 and 11:13. In total, we harvested nine males from these treatments: Three on Day 5 at 16:8, three from Day 10 at 16:8 and three from Day 10 at 11:13. Whole beetles were snap-frozen in liquid nitrogen and stored at -80° C until RNA isolation.

RNA was isolated from individually dissected heads using the Qiagen RNeasy Micro Kit for each individual sample following the manufacturer's protocol. Beetle heads were crushed in liquid nitrogen with a disposable pestle for tissue disruption. The liquid nitrogen was allowed to evaporate while keeping the tube on dry ice. A total of 1050 uL RLT buffer was added. A syringe with 20-gauge needle was used for homogenization by passing the lysate at least 5-10 times. The lysate was centrifuged for 3 minutes at full speed. The supernatant was transferred to a fresh tube and used immediately.

From these samples, we constructed a library for transcriptome assembly. A cDNA library was constructed by the IBEST Genomics Resources Core from pooled equal mass RNA from these nine starting samples using a stranded RNA-Seq kit from Kapa. qPCR was performed to determine the quantity of sequenceable library. The resultant stranded library of nine pooled male heads was sequenced in one lane of a 2x300 Miseq run at the IBEST GRC, resulting in a total of 19.85 million raw paired reads.

We began reference-guided transcriptome assembly by visualizing raw reads in FASTQC (Andrews 2010) then removed adapters and trimmed reads with Trimmomatic v0.39 (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (Bolger et al. 2014). The remaining 19.811 million trimmed paired reads were aligned to this *de novo* *D. carinulata* reference genome with the fast, splice-aware HiSat2 v2.2.1 (Kim et al. 2015; Kim et al. 2019), then assembled transcripts with StringTie2 (Kovaka et al. 2019). Transcript sequences were extracted using gffread (Pertea and Pertea 2020).

We assessed the quality of the transcriptome assembly by aligning QuantSeq reads (Moll et al. 2014) generated from the same individuals (NCBI SRX10122873 - SRX10122884) to the resultant transcriptome assembly with bowtie2 v2.2.9 (Langmead and Salzberg 2012b). Additionally, we ran BUSCO v5.0.0 in transcriptome mode and quantified the number of fully assembled transcripts with blastx (Camacho et al. 2009).

Synteny across Diorhabda species

To improve our understanding of barriers to hybridization and the genomic landscape of speciation, we examined the chromosomal distribution of two independent results: 1) mapped BUSCOs, and 2) RADseq loci that did not align to the *D. carinulata* genome. For BUSCO comparisons, we merged the full table of BUSCO results for each species, then visualized orthologous gene abundance between scaffolds. Then, to examine unique loci between species, we used samples that were previously assigned to have dominant *D. carinata*, *D. elongata*, and *D. sublineata* ancestry (Stahlke et al. in prep). First, we aligned the RADseq loci of those individuals to the *D. carinulata* assembly presented here and captured the paired-end reads that did not align concordantly with bowtie2. Then, we aligned those paired-end reads to assemblies of *D. carinata*, *D. elongata*, and *D. sublineata*. For alignments, we used the --very-sensitive wrapper to allow for more divergent alignments based on sequence identity. We quantified genome coverage across 10 kb windows using the multicov tool of bedtools v2.26.0 (Quinlan 2014) and visualized the distribution of aligned reads for each species.

We computed summary statistics and prepared visualizations in R v4.0.3 (R Core Team 2020), employing dplyr v1.04 (Wickham et al. 2016) (Wickham Hadley, François Romain, Henry Lionel, Müller Kirill 2021), ggplot2 v3.3.3 (Wickham 2016), and ggalluvial v0.12.3 (Brunson and Read 2020).

Results

Genome assembly and assessment

By flow cytometry, genome sizes of *D. carinata* and *D. elongata* were estimated to be 443.6 MB and 472.5 MB. Genome sizes of both males and females were estimated to be 395.8 +/- 0.5 Mb and 404.1 +/- 0.6 MB *D. carinulata*, and 455.5 +/- 0.9 MB and 465.0 +/- 0.5 MB for *D. sublineata*, respectively, indicating male heterogamy and possible karyotypic differences between species.

For each of the species, we generated 1,205,352 - 2,881,478 PacBio HiFi reads, ranging from 43 - 37,739 bp and an average of 9,081 bp in length (Table 3.1). K-mer based assessment of PacBio HiFi reads for each species estimated genome sizes for each species (350-375 MB), with heterozygosity between 50-60% (Table 3.1; Fig. 3.1). The PacBio library for *D. carinulata* had the highest effective

k-mer coverage (33.5x), while *D. carinata*, *D. elongata*, *D. sublineata* each had about half the coverage (approx. 17x). Unique sequence for all libraries was estimated to be about 85% for all species.

Each primary contig assembly of these *Diorhabda* species resulted in long, nearly chromosome length contigs and very high rates of complete single-copy orthologs according to BUSCO (Table 3.2). For all four species, chromosome-level scaffolds were recovered in the largest 11-13 scaffolds of the L90 (Table 3.2, Fig. 3.2).

We successfully extracted mitochondrial genomes for all four species and circularized those of *D. carinata* and *D. carinulata*. Using the MitoFinder pipeline, we identified 16-20 of the 22 expected tRNAs and 9-13 of the 13 expected protein coding genes (depending on species), however several of these were fragmented.

Genome Annotation

Repeat Content

Classifying sequences from the *de novo* repeat libraries revealed a moderately high rate of repetitive content in all four *Diorhabda* species, between 47.05 - 53.89%, and scaling with genome size. The assembly of *D. carinulata* had the least repeat content and *D. elongata* had the most (Table 3.3).

Across species, the majority of transposable elements were unclassified DNA interspersed repeats (45.98 - 52.81%) (Table 3.3). We identified canonical telomere repeats in 13 scaffolds of *D. carinulata* (three end to end), 10 in *D. carinata* (four end to end), 16 in *D. elongata* (three end to end), and ten in *D. sublineata* (one end to end). Regions of high repeat content indicative of telomeres (at the end of scaffolds) and centromeres (in the middle of scaffolds) were evident by visualization of genome-wide repeats (Fig. 3.6).

Sex Chromosomes

The distribution of RADseq markers from population sampling supported male heterogamy for all four species, with several markers significantly detected in females compared to males (Figs. 3.3-4). Aligning these significant markers to each scaffolded assembly here, we identified the chromosome-scale scaffold from each species with significant sex-biased coverage for female samples, an X chromosome, which also had less coverage than other scaffolds in the Hi-C plots (Fig. 3.2). We also identified two scaffolds indicative of at least one Y chromosome in *D. carinulata* (Fig. 3.5), but we did not find significantly sex-biased RADseq loci indicative of a Y chromosome in the other species. Population structure was evidenced by patchy in the distribution of RADseq markers for the hybrid samples (Fig. 2.4).

Aligning the putative Y chromosome scaffolds (scaffolds 14 and 15) we identified in *D. carinulata* to the other three species, we recovered scaffolds with significant sequence homology that did not share homology with other chromosomes. In *D. carinata*, scaffold 13 had the greatest shared sequence identity with both Y scaffold candidates of *D. carinulata*. In *D. sublineata*, high-scoring homology was only found between scaffold 14 of *D. carinulata* and scaffolds 422 and 13.

Reference-guided transcriptome assembly of D. carinulata

Using the trimmed, paired-end RNAseq data from the nine male *D. carinulata* heads, we assembled 28,162 transcripts, comprising 562,547,083 nucleotides. The median, mean, and N50 contig lengths were 6,572.5 bp, 19,975.40, and 62,976 respectively. Single-end QuantSeq reads from the same individuals as those used to build the transcriptome aligned back to the assembly at an acceptable rate (85 - 95%). The assembled transcriptome and QuantSeq reads passed sufficient quality for submission to the NCBI RefSeq annotation platform.

Synteny across Diorhabda species

Upon visual inspection of BUSCO mapping across assemblies of *D. sublineata*, *D. elongata*, *D. carinata*, and *D. carinulata*, we identified conserved chromosomal homology among some chromosomes, as well as putative chromosomal rearrangements unique to *D. carinulata* and conserved between *D. carinata*, *D. elongata*, and *D. sublineata* (Fig. 3.8). For example, Chromosome 1 of *D. carinulata* shares orthology with Chromosome 1 and 2 of *D. carinata*, but Chromosome 1 and 2 of *D. carinata* are respectively completely orthologous with distinct chromosomes of *D. elongata* (Chr001 and Chr004, respectively) and (Chr001 and Chr002, respectively). Contrastingly, we did not find any such splits in seven of the 12 apparent chromosomes of *D. carinulata*. Examining pairwise orthology from west to east, e.g., *D. sublineata* vs *D. elongata*, we found the greatest number of rearrangements between *D. carinata* and *D. carinulata*.

To examine interactions between genome evolution and sequence divergence, we compared the number of reads that aligned to either *D. carinata*, *D. elongata*, or *D. sublineata*, and did not align to the *D. carinulata* genome (Fig. 9). We found that *D. carinata* had the highest average coverage of reads/10KB window with 569.30, followed closely by 564.63 for *D. sublineata*, then 502.43 for *D. elongata*. The number of alignments decreased moving east to west, with the fewest aligned reads against the *D. sublineata* genome and in the fewest regions.

Discussion

High-quality genomic resources

We produced high-quality, nearly chromosome-scale *de novo* genome assemblies from single individuals of *D. carinulata*, *D. carinata*, and *D. elongata* with a single pipeline that scaffolded PacBio HiFi reads with Hi-C data. These assemblies mark a dramatic improvement from the first draft assembly of *D. carinulata* (Stahlke et al. in prep), from an N/L50 of 3,561/.70 MB and the genome totaling approx. 84,000 scaffolds to an N/L50 of 5/33.547 MB and a total of 132 scaffolds (Fig. 7). This improvement is marked by an increase in total assembly length, from 382.502 MB to 414.77 MB, with 30% greater repeat content annotated, including telomeres, centromeres, and highly repetitive scaffolds associated with males (Y chromosomes) (Fig. 3.7). Similarly, the mitochondrial genomes we describe here are about 1 KB longer than the previously published mitochondrial genomes of *D. carinata* and *D. carinulata* (Stahlke et al. 2019), likely due to more accurate assembly of the repetitive control region and intergenic content. Using Hifiasm, we also recovered highly continuous alternate assemblies, i.e., phased haplotypes, for all four *Diorhabda* assemblies. These highly continuous assemblies will accelerate studies of hybridization and contemporary evolution by providing a basis for fine-scale maps of recombination and long-range context of linkage disequilibrium (McKinney et al. 2017; Dias-Alves et al. 2018) More broadly, our work sheds light on the role of genome evolution in determining the outcomes of secondary contact among cryptic BCA species and demonstrates an efficient approach to test hypotheses regarding stability.

Sex chromosomes

Our findings of male heterogamy and specific identification of the X chromosome in all four species present a significant opportunity to examine sex-chromosome specific research in *Diorhabda*. Sex chromosomes can harbor key genes related to reproductive barriers (Presgraves 2008; Dowle et al. 2017; Baiz et al. 2020; Schield et al. 2021) and play large roles in rapid evolution due to their differential inheritance and recombination (Charlesworth et al. 1987; Meisel and Connallon 2013; Lasne et al. 2017). Relevant to *Diorhabda*, loci involved with diapause regulation are commonly found on sex chromosomes (Hagen and Scriber 1989; Ikten et al. 2011; Chen et al. 2014; Fu et al. 2015; Pruischer et al. 2018). Differential introgression and reproductive barriers are expected along the X-chromosome in *Diorhabda* due to female-biased mortality observed in previous lab-crosses (Bean et al. 2013), male-biased dispersal due to semiochemicals and aggregation pheromones (Coss et al. 2005; Wertheim et al. 2005; Gaffke et al. 2018), and mito-nuclear discordance found in the native range (Bean et al. 2013) and in recent secondary contact between *D. carinulata* and *D. sublineata* (Ozsoy et al. 2021).

The lack of any signal of male-specific RADseq loci may indicate that these species lack a Y chromosome (Fig. 3.4). However, even at relatively high coverage and with the relatively high-quality assemblies as we present here, Y chromosomes can be hard to assemble. Because of lower effective population sizes, lack of recombination, and the dominance of X-chromosomes, they can become highly repetitive and may lack genes (Bachtrog 2013). Y chromosomes have been lost among many taxa, including multiple independent events in Coleoptera (Blackmon et al. 2017) (Blackmon, Ross, and Bachtrog 2017), although loss is hypothesized to be less common in Polyphaga due to achiasmatic pairing during male meiosis (Blackmon and Demuth 2014). Nonetheless, without additional species-specific population genomic data for *D. elongata*, *D. carinata*, and *D. elongata*, we could not confidently identify a Y chromosome, but did identify homologous sequences based on alignment. Species-specific genomic data and basic cytological karyotyping could be performed to validate the presence or absence of a Y chromosome in *D. carinata*, *D. elongata*, and *D. sublineata*.

Transposable Elements

Transposable elements are increasingly appreciated for their role in rapid, contemporary evolution (Stapley et al. 2015; Schrader and Schmitz 2019) and reproductive isolation (Dowle et al. 2017). The relatively high rate of transposable elements we recovered in all four *Diorhabda* assemblies (47-54%; Table 3.3), suggest that they should be considered for a number of traits in *Diorhabda*; for example, in diapause regulation (Yamashita et al. 2001; Robich et al. 2007; Yocum et al. 2011; Sasibhushan et al. 2012) and host specificity (Schoville et al. 2018). We found that TE and genome size scale accordingly across *Diorhabda*, similar to other arthropods (Talla et al. 2017; Blommaert et al. 2019). Similar rates of repetitive content were found in another Chrysomelid BCA known for rapid evolution to novel environments, *Ophrealla communa* (Bouchemousse et al. 2020), supporting family-wide patterns typical among insects (Petersen et al. 2019). Further work to understand the role of these repetitive elements should employ additional resources and tools (Hoede et al. 2014) to classify these largely unclassified repeats found in 16-28% of these assemblies (Table 3.3).

Synteny and the stability of cryptic species

The role of chromosomal rearrangements in driving ecological speciation and preventing gene flow has been debated (Rieseberg 2001; Faria and Navarro 2010). With chromosome-scale assemblies for all four introduced *Diorhabda* species, previous laboratory crossing experiments (Bean et al. 2013; Bitume et al. 2017), and current distributions of introduced-range hybridization (Ozsoy et al. 2018; Knutson et al. 2019; Ozsoy et al. 2021) we can address these hypotheses. We found that coexistence of cryptic BCA species upon secondary contact may be predicted by genome structure and native range species distributions. Sympatric *D. carinulata* and *D. carinata* have reduced rates of shared

BUSCO synteny; while peripatric and allopatric *D. sublineata*, *D. elongata*, and *D. carinata* (ordered west to east), have conserved chromosomal assignment of BUSCOs (Fig. 3.7) (Tracy and Robbins 2009). The lower rates of synteny between *D. carinulata* and all three species also reflect the outcomes of laboratory crosses (Bean et al. 2013), with genital mismatch (Sota and Kubota 1998) leading to high rates of mortality for *D. carinulata* females and low rates of egg-hatch. Conserved synteny between *D. carinata*, *D. elongata*, and *D. sublineata* suggests one fewer barrier to reproductive isolation and reflects both the success of laboratory crosses (Bean et al. 2013; Bitume et al. 2017) and high frequencies of hybridization where all three species were introduced (Knutson et al. 2019).

D. sublineata x *D. carinata* hybrids, compared to hybrids with *D. elongata*, have been described at greater frequency, with evidence of back-crossing in N. America (Stahlke et al. in prep), and showed increased fecundity in laboratory crosses (Bitume et al. 2017), suggesting a lack of reproductive barriers evolved in allopatry compared to peripatric *D. elongata*. The relatively conserved synteny we observed among the three hybridizing species according to BUSCO mapping was supported by alignment rates of RADseq loci, wherein alignment success was lowest for *D. elongata* suggesting increased sequence divergence despite chromosomal homology. A comparison of islands of divergence (Turner et al. 2005) among these species would allow us to examine these hypotheses more directly. For example, among sympatric and allopatric bumblebees, allopatric species had fewer islands of divergence than those in sympatry which had evolved barrier loci in regions of reduced recombination (Christmas et al. 2021). Notably, *D. elongata* has the largest genome among all four *Diorhabda* and the highest percentage of bases identified as repetitive content, which may suggest a role for TEs in generating reproductive barriers.

Future directions with gene evidence

The highly continuous genome assemblies we produced here are likely to facilitate studies of gene family evolution across the genus (Li et al. 2019); for example, to investigate gene family evolution related to herbivory and host-specialization (Goldman-Huertas et al. 2015; Calla et al. 2017; Comeault et al. 2017). To fully realize the potential of these *Diorhabda* assemblies, additional species-specific gene evidence should be generated for multiple life-stages and tissue-types. However, the mito-nuclear discordance observed among *D. carinata*, *D. elongata*, and *D. sublineata* in the native range (Bean et al. 2013) suggests incomplete lineage sorting, which should be explicitly modeled to account for hemiplasy (Szöllősi et al. 2015; Guerrero and Hahn 2018). Nonetheless, the gene evidence and transcriptome we generated here from *D. carinulata* will be useful to study the evolution of diapause.

Conclusion

These assemblies are part of a broader project, the Ag100 Pest Initiative (Childers et al. 2019), to address the needs of agriculture in the United States and contribute high quality reference genomes of agricultural importance to the Earth BioGenome Project (Lewin et al. 2018). This pipeline was shown to be a highly effective approach across these four *Diorhabda*, with all four meeting the Earth BioGenome standards for continuity of contigs, with the NG50 of contigs >10 MB, and completeness, with BUSCO single-copy orthologs over 95%. Additionally, our work represents the first chromosome-level assemblies in BCA of invasive plants and presents a path forward to address the current dearth of genomic resources therein.

Table 3.1 Input sequence characteristics for PacBio HiFi reads, genome characteristics estimated by Genomescope (Vurtture et al. 2017) from HiFi k-mers, and Hi-C alignments according to the PhaseGenomics Hi-C QC tool. For GenomeScope, estimates of genome characteristics based on 21-mer histograms for the HiFi reads following adapter removal.

Input Data Statistics	<i>D. carinulata</i>	<i>D. carinata</i>	<i>D. elongata</i>	<i>D. sublineata</i>
Raw PacBio HiFi Reads				
Number of Sequences	2,881,478	1,649,607	4,531,085	6,180,692
Total length (bp)	26,169,167,863	15,252,588,416	41,421,756,279	56,674,344,695
Minimum length (bp)	43	44	48	49
Average length (bp)	9,081.90	9,246.20		
Maximum length	37,739	39,481	27,247	31,295
GenomeScope Estimates				
Heterozygosity	0.59%	0.48%	0.51%	0.51%
Genome Haploid Length (MB)	340.621	360.197	362.213	362,326,287.00
Genome Repeat Length (MB)	47,956,677.50	58,337,352.50	60,999,126.50	61,018,284.00
Genome Unique Length (MB)	292.58	301.86	301.21	301.310
Genome Unique Content (%)	0.00%	83.80%	83.16%	83.16%
Model Fit	96.93%	96.72%	97.21%	98.22%
Read Error Rate	0.13%	0.21%	0.14%	0.14%
Hi-C QC				
Same Strand (%)	37.08%	35.72%	37.85%	37.08%
Informative Read Pairs (%)	63.47%	63.30%	68.89%	67.48%
Noninformative Read Pairs	0.62%	0.27%	0.27%	0.23%
Long Contacts	57.07%	57.04%	59.01%	56.96%
Intercontig Contacts	14.91%	14.57%	24.10%	24.44%
Usable high-quality read pairs	2,129.30	2,673.77	1,473.79	1,046.62

Table 3.2 Primary contig assembly statistics and final Hi-C scaffolded assembly statistics after manual curation.

Primary Assembly Statistics	<i>D. carinulata</i>	<i>D. carinata</i>	<i>D. elongata</i>	<i>D. sublineata</i>
Number of Scaffolds	132	83	299	422
Number of Contigs	143	109	355	467
Total Length (MB) Scaffolds	417.239	448.661	481.404	456.344
Total Length (MB) Contigs, %	417.237,	448.658,	481.400,	456.339,
Scaffold N/L50 (MB)	5/33.547	6/40.833	6/39.447	6/36.832
Contig N/L50 (MB)	8/21.665	7/24.528	12/12.818	13/13.817
Scaffold N/L90	12/14.207	11/18.376	14/4.257	13/13.817
N/L90 (MB)	22/8.164	19/6.897	44/2.534	55/1.025
Largest Scaffold	68.588	48.62	47.753	44.198
Largest Contig (MB)	37	46	40	29
Number of scaffolds > 50 KB	48	29	87	173
% main genome in scaffolds > 50	99.49%	99.66%	99.16%	98.84%
Contig BUSCO Single	98.70%	98.40%	98.50%	98.50%

Table 3.3 Classified repeat content for the assembled genomes of each *Diorhabda* species.

Repeat Classification	<i>D. carinulata</i>	<i>D. carinata</i>	<i>D. elongata</i>	<i>D. sublineata</i>
SINEs	0.07	0.06	0.05	0
LINES	8.64	8.27	7.23	8.12
LINE1	0.04	0	0.01	0
LINE2	0.96	0.78	0.73	0.74
L3/CR1	0.36	0.34	0.29	0.31
LTR elements	1.98	2.1	1.58	1.7
ERV1	0	0	0	0
ERV1-MaLRs	0	0	0	0
ERV Class I	0.01	0	0.01	0.01
ERV Class II	0	0	0	0
DNA elements	18.78	17.32	15.55	17.67
hAT-Charlie	1.07	0.72	1.13	0.91
TcMar-Tigger	0.17	0.14	0.12	0.15
Unclassified	16.55	19.33	28.4	22.04
Total interspersed repeats	46.01	47.06	52.81	49.53
Small RNA	0.37	0.35	0.41	0.12
Satellites	0.05	0.5	0.03	0
Simple repeats	0.75	0.7	0.63	0.7
Low complexity	0.11	0.11	0.1	0.11
Total bases masked	47.05%	48.63 %	53.89%	50.39 %

Figure 3.1 Profiles of 21-mer histograms according to coverage and respective frequency and Genomescope (Vurture et al. 2017) model-fit for the CSS reads of all four species following adapter removal. The short peak with less coverage for all species indicates heterozygosity.

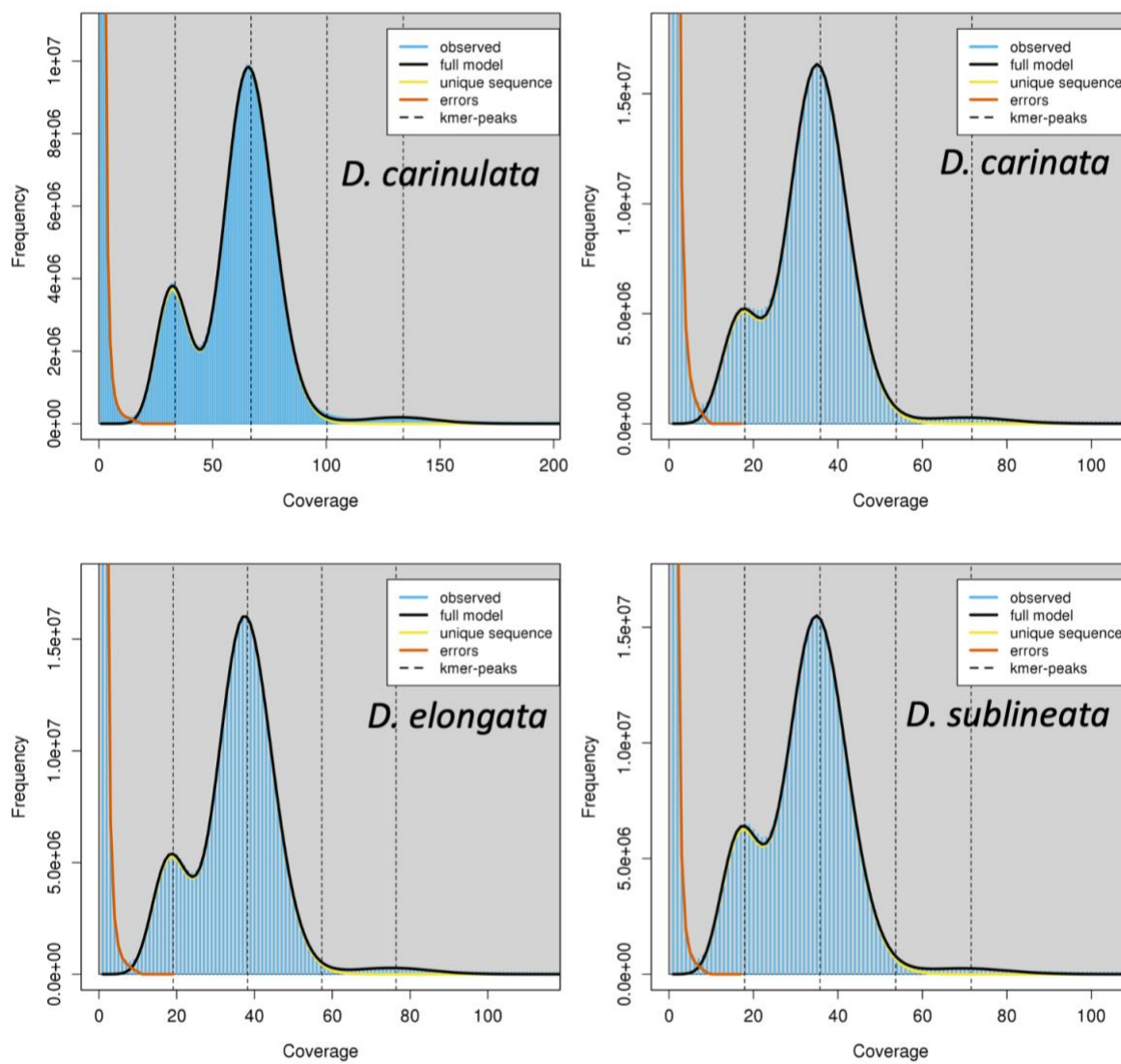


Figure 3.2 Hi-C contact heatmaps and curated scaffolds for assemblies of each *Diorhabda* species.

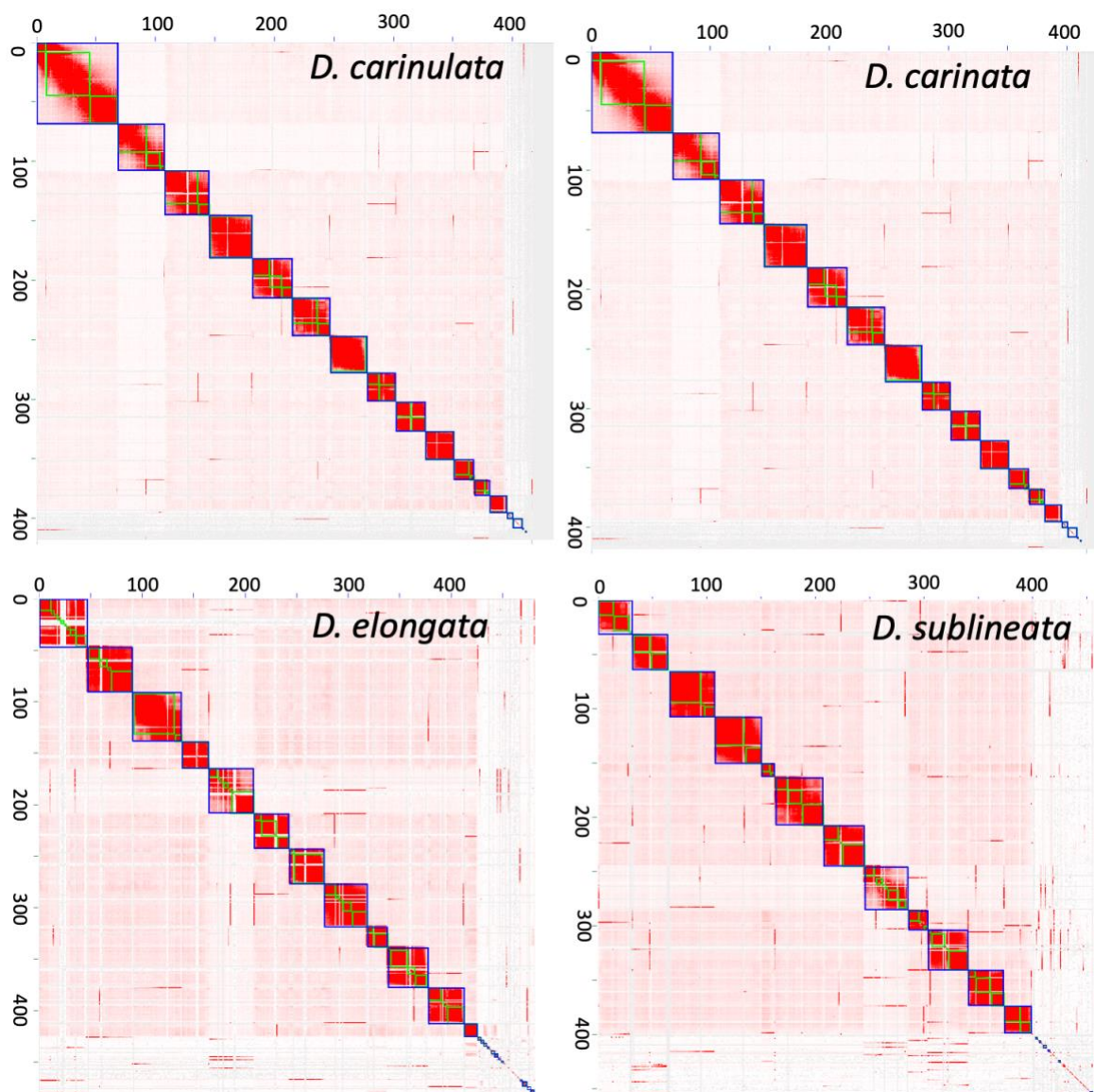


Figure 3.3 Distribution of RADseq markers by female and male coverage across population sampling of *D. carinulata*. Markers significantly associated with males or females are highlighted in red boxes.

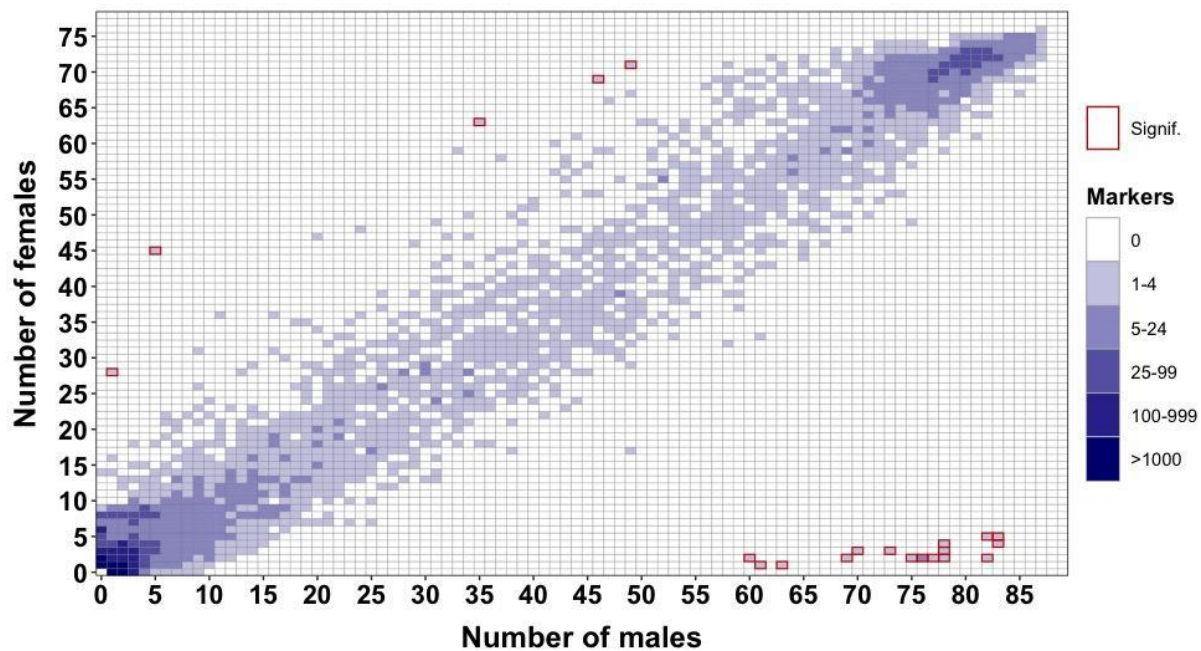


Figure 3.4 Distribution of markers by female and male coverage for the *D. carinata*, *D. elongata*, and *D. sublineata* samples.

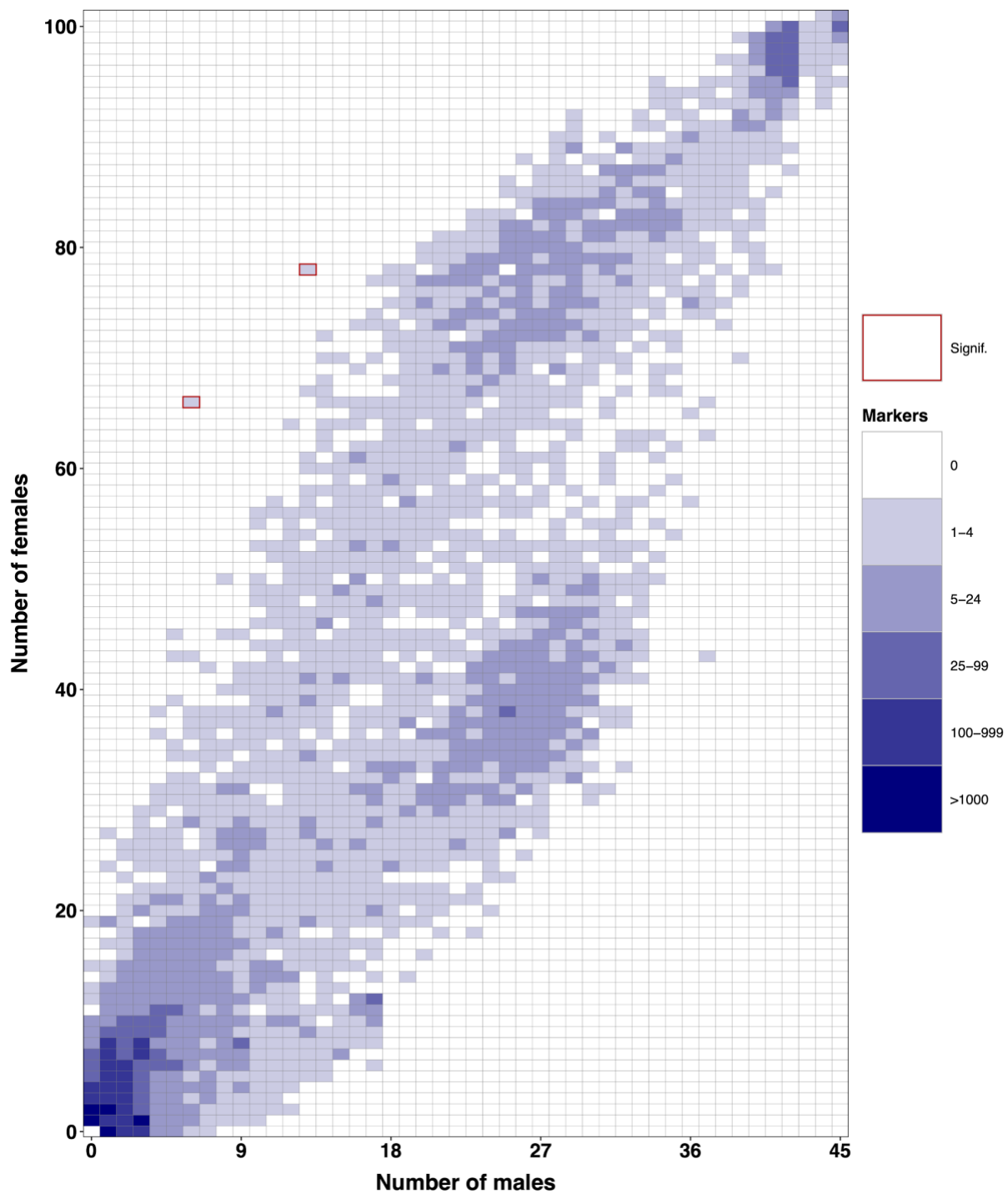


Figure 3.5 Manhattan plot of markers from male and female *D. carinulata* samples aligned against the scaffolded *D. carinulata* assembly. Scaffolds are ordered from largest to smallest. The second largest scaffold (labeled Chr2 here) is saturated with female biased markers, shown as significantly reduced coverage. The fourteenth and fifteenth largest scaffolds (Chr14 and Chr15) have markers significantly associated with males.

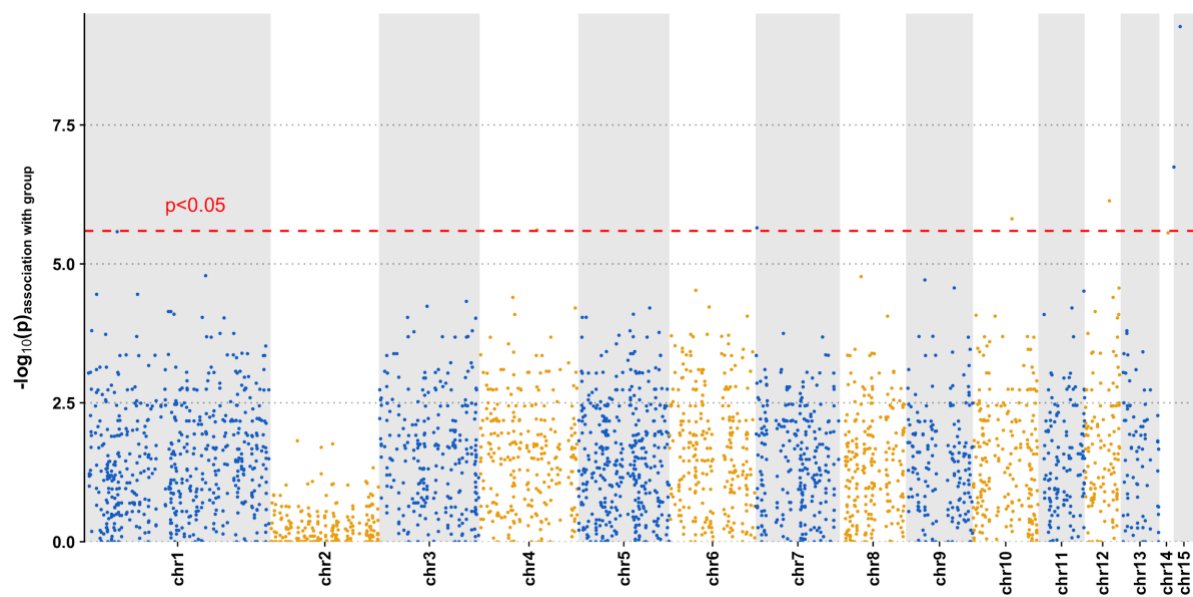


Figure 3.6 Genome-wide repeat content annotated along ordered scaffolds of *D. carinulata*. Chromosome-level scaffolds are ordered by size, followed by sex chromosomes identified with population genomic data (Fig. 3.5), labeled here as ChrX and ChrY1 and ChrY2.

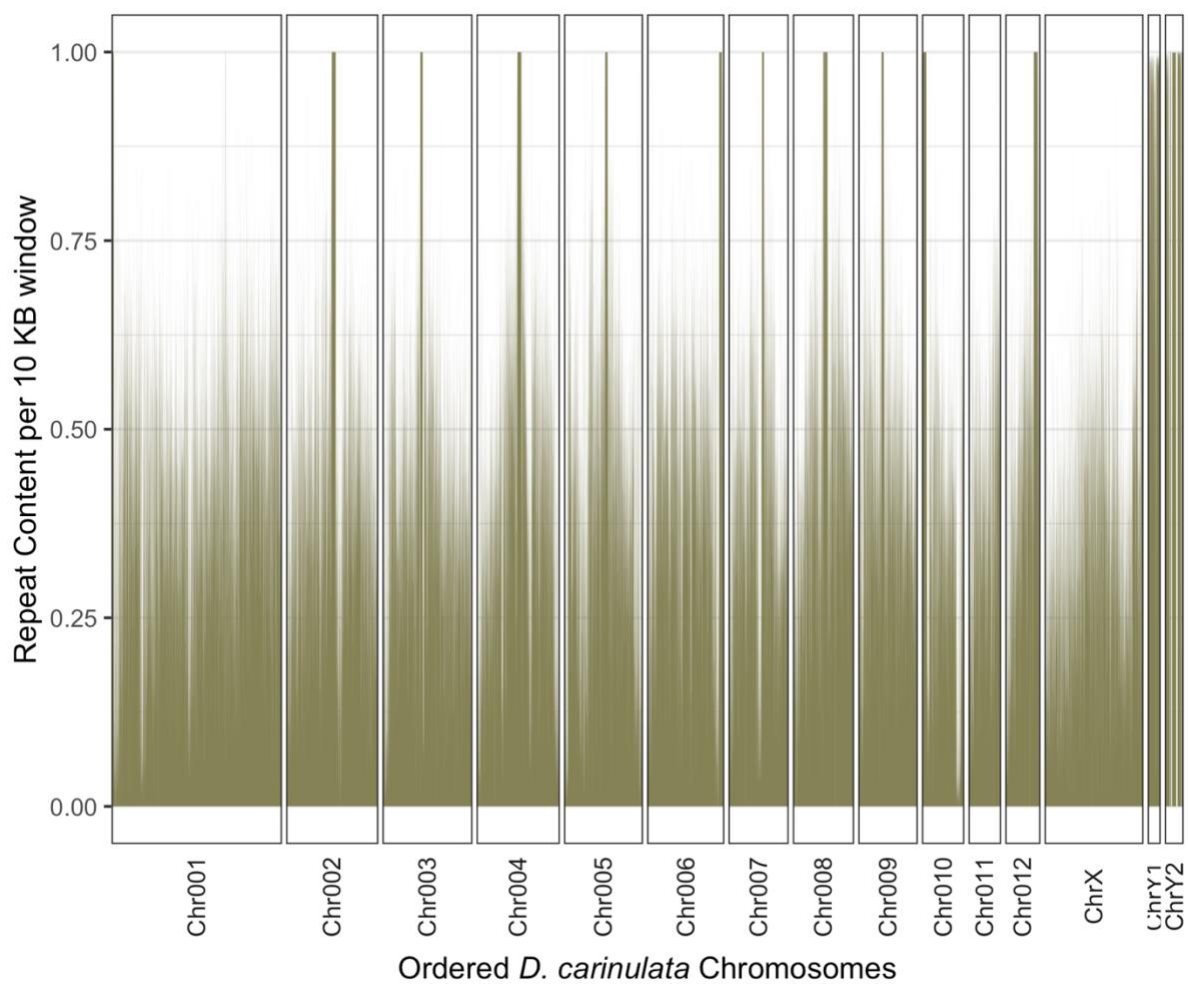


Figure 3.7 BUSCO synteny between the improved assembly produced here with HiFi and Hi-C data (top) versus the first scaffolded assembly of *D. carinulata* using Shotgun and 10X Chromium data (Stahlke et al. in prep) (bottom). The HiFi + Hi-C chromosome-length scaffolds are ordered by length (Chr001 - Chr012), followed by the X chromosome, then one scaffold (Scaff129) at the end. The Shotgun + Chromium assembly is ordered according to scaffold length. Blocks are sized according to the number of BUSCO genes mapped to that chromosome or scaffold. BUSCO paths are colored according to chromosome assignment in the new HiFi + Hi-C assembly.

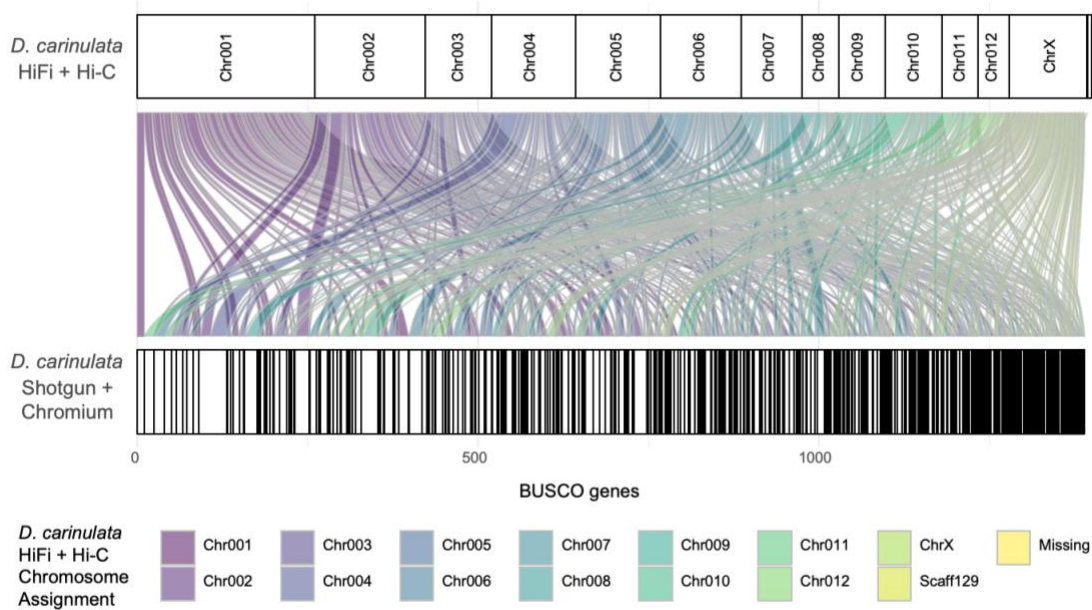


Figure 3.8 BUSCO syntenic across assemblies of *D. sublineata*, *D. elongata*, *D. carinata*, and *D. carinulata*, ordered from left to right by native range west to east, and colored according to the assigned *D. carinata* chromosome. The small number of BUSCOs that were not found ('Missing') are shown in the bottom row. Blocks are sized according to the number of BUSCO genes mapped to that chromosome or scaffold. Chromosomes are ordered by size within each species, followed by the X chromosome. *D. carinulata* was the only species with a twelfth chromosome-level scaffold.

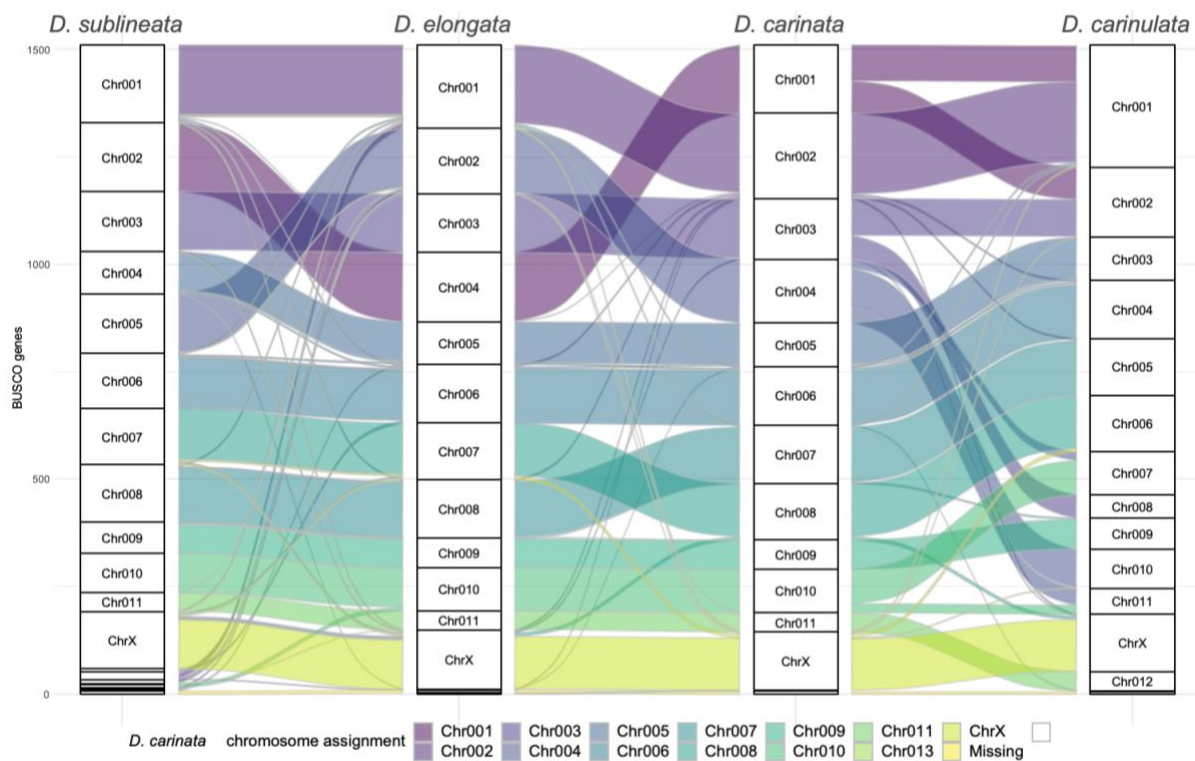
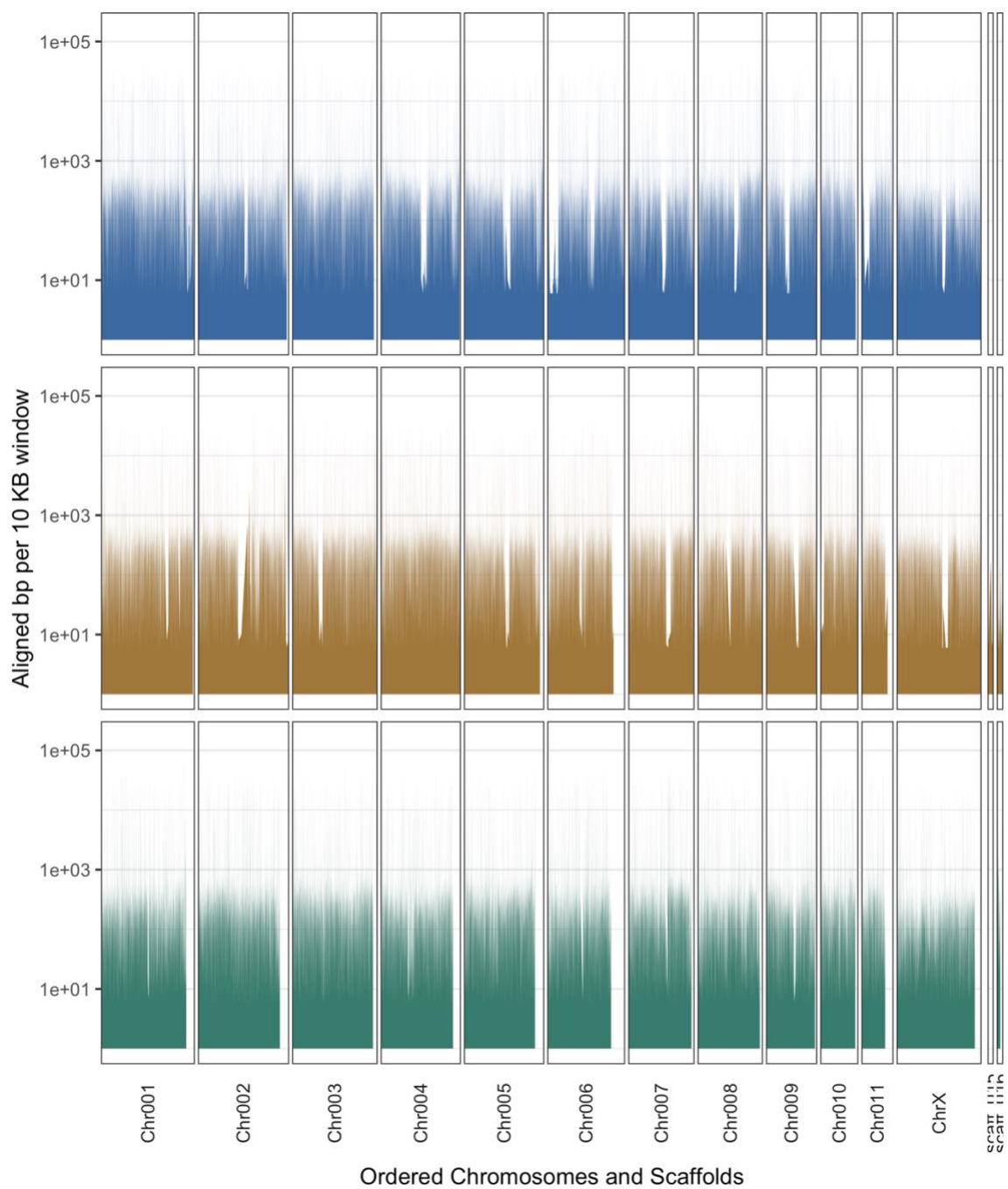


Figure 3.9 Coverage of reads from samples of *D. sublineata*, *D. elongata*, and *D. carinata* that did not align concordantly to the *D. carinulata* assembly. Chromosome-level scaffolds are ordered by size within each species, followed by the X chromosome



Literature Cited

- Aken, B. L., P. Achuthan, W. Akanni, M. R. Amode, F. Bernsdorff, J. Bhai, K. Billis, D. Carvalho-Silva, C. Cummins, and P. Clapham. 2016. Ensembl 2017. *Nucleic acids research* 45:D635-D642.
- Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* 202:389-400.
- Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17:81.
- Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.
- Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular biology and evolution* 18:1585-1592.
- Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 14:113-124.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one* 3:e3376.
- Baiz, M. D., P. K. Tucker, J. L. Mueller, and L. Cortés-Ortiz. 2020. X-linked signature of reproductive isolation in humans is mirrored in a howler monkey hybrid zone. *Journal of Heredity* 111:419-428.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19:455-477.
- Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth. 2011. BamTools. *Bioinformatics* 27:1691-1692.
- Barreiro, L. B. and L. Quintana-Murci. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* 11:17-30.
- Bay, R. A., N. Rose, R. Barrett, L. Bernatchez, C. K. Ghalambor, J. R. Lasky, R. B. Brem, S. R. Palumbi, and P. Ralph. 2017. Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. *Am Nat* 189:463-473.
- Bean, D. and T. Dudley. 2018. A synoptic review of *Tamarix* biocontrol in North America: tracking success in the midst of controversy. *BioControl* 63:361-376.
- Bean, D. W., P. Dalin, and T. L. Dudley. 2012. Evolution of critical day length for diapause induction enables range expansion of *Diorhabda carinulata*, a biological control agent against tamarisk (*Tamarix* spp.). *Evolutionary Applications* 5:511-523.
- Bean, D. W., T. L. Dudley, and J. C. Keller. 2007a. Seasonal timing of diapause induction limits the effective range of *Diorhabda elongata desertiicola* (Coleoptera: Chrysomelidae) as a biological control agent for tamarisk (*Tamarix* spp.). *Environmental entomology* 36:15-25.
- Bean, D. W., D. J. Kazmer, K. Gardner, D. C. Thompson, B. Reynolds, J. C. Keller, and J. F. Gaskin. 2013. Molecular Genetic and Hybridization Studies of *Diorhabda* spp. Released for Biological Control of Tamarix. *Invasive Plant Science and Management* 6:1-15.
- Bean, D. W., T. Wang, R. J. Bartelt, and B. W. Zilkowski. 2007b. Diapause in the leaf beetle *Diorhabda elongata* (Coleoptera: Chrysomelidae), a biological control agent for tamarisk (*Tamarix* spp.). *Environmental Entomology* 36:531-540.
- Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. *Trends in ecology & evolution* 22:148-155.

- Bielawski, J. P. and Z. Yang. 2005. Maximum likelihood methods for detecting adaptive protein evolution. Pp. 103-124. *Statistical methods in molecular evolution*. Springer.
- Bitume, E. V., D. Bean, A. R. Stahlke, and R. A. Hufbauer. 2017. Hybridization affects life-history traits and host specificity in *Diorhabda* spp. *Biological Control* 111:45-52.
- Blackmon, H. and J. P. Demuth. 2014. Estimating tempo and mode of Y chromosome turnover: explaining Y chromosome loss with the fragile Y hypothesis. *Genetics* 197:561-572.
- Blackmon, H., L. Ross, and D. Bachtrog. 2017. Sex determination, sex chromosomes, and karyotype evolution in insects. *Journal of Heredity* 108:78-93.
- Blommaert, J., S. Riss, B. Hecox-Lea, D. B. M. Welch, and C.-P. Stelzer. 2019. Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC genomics* 20:1-12.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
- Bouchemousse, S., L. Falquet, and H. Muller-Scharer. 2020. Genome Assembly of the Ragweed Leaf Beetle: A Step Forward to Better Predict Rapid Evolution of a Weed Biocontrol Agent to Environmental Novelty. *Genome Biol. Evol.* 12:1167-1173.
- Brandies, P., E. Peel, C. J. Hogg, and K. Belov. 2019. The Value of Reference Genomes in the Conservation of Threatened Species. *Genes* 10:846.
- Braschi, B., P. Denny, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates, and E. Bruford. 2019. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic acids research* 47:D786-D792.
- Brunson, J. C. and Q. D. Read. 2020. Ggalluvial: Alluvial Plots in ggplot2.
- Brüniche-Olsen, A., J. J. Austin, M. E. Jones, B. R. Holland, and C. P. Burridge. 2016. Detecting selection on temporal and spatial scales: a genomic time-series assessment of selective responses to devil facial tumor disease. *PloS one* 11.
- Brüniche-Olsen, A., M. E. Jones, J. J. Austin, C. P. Burridge, and B. R. Holland. 2014. Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. *Biology letters* 10:20140619.
- Brüniche-Olsen, A., M. E. Jones, C. P. Burridge, E. P. Murchison, B. R. Holland, and J. J. Austin. 2018. Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. *Journal of biogeography* 45:963-976.
- Burton, O. J., B. L. Phillips, and J. M. J. Travis. 2010. Trade-offs and the evolution of life-histories during range expansion. *Ecology Letters* 13:1210-1220.
- Bushnell, B. 2014. BBTools software package. URL <http://sourceforge.net/projects/bbmap> 578:579.
- Cai, S., X. Cheng, Y. Liu, Z. Lin, W. Zeng, C. Yang, L. Liu, O. A. Chukwuebuka, and W. Li. 2018. EYA1 promotes tumor angiogenesis by activating the PI3K pathway in colorectal cancer. *Experimental cell research* 367:37-46.
- Calla, B., K. Noble, R. M. Johnson, K. K. O. Walden, M. A. Schuler, H. M. Robertson, and M. R. Berenbaum. 2017. Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. *Molecular ecology* 26:6021-6035.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, and K. Bealer. T. Madden, 2009: BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10:1-9.
- Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour* 15:855-867.
- Caniglia, R., M. Galaverni, E. Velli, F. Mattucci, A. Canu, M. Apollonio, N. Mucci, M. Scandura, and E. Fabbri. 2020. A standardized approach to empirically define reliable assignment

- thresholds and appropriate management categories in deeply introgressed populations. *Scientific Reports* 10.
- Capinha, C., F. Essl, H. Seebens, D. Moser, and H. M. Pereira. 2015. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348:1248-1251.
- Carruthers, R. I., C. J. DeLoach, J. C. Herr, G. L. Anderson, and A. E. Knutson. 2008. Salt cedar areawide pest management in the western USA. *Areawide pest management theory and implementation*. CAB International, Wallingford:271-299.
- Casanova, J.-L. and L. Abel. 2018. Human genetics of infectious diseases: Unique insights into immunological redundancy. Pp. 1-12. Elsevier.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22:3124-3140.
- Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist* 130:113-146.
- Chen, C., L. Xiao, H. M. He, J. Xu, and F. S. Xue. 2014. A genetic analysis of diapause in crosses of a southern and a northern strain of the cabbage beetle *Colaphellus bowringi* (Coleoptera: Chrysomelidae). *Bulletin of entomological research* 104:586.
- Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170-175.
- Childers, A. K., B. Coates, S. Geib, M. F. Poelchau, C. P. Childers, B. E. Scheffler, and K. Hackett. 2019. The Ag100Pest Initiative: USDA-ARS's Contribution to the Earth BioGenome Project. *Entomology 2019*. ESA.
- Christmas, M. J., J. C. Jones, A. Olsson, O. Wallerman, I. Bunikis, M. Kierczak, V. Peona, K. M. Whitley, T. Larva, and A. Suh. 2021. Genetic barriers to historical gene flow between cryptic species of alpine bumblebees revealed by comparative population genomics. *Molecular Biology and Evolution*.
- Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, and B. Wilczynski. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422-1423.
- Comeault, A. A., A. Serrato-Capuchina, D. A. Turissini, P. J. McLaughlin, J. R. David, and D. R. Matute. 2017. A nonrandom subset of olfactory genes is associated with host preference in the fruit fly *Drosophila orena*. *Evolution Letters* 1:73-85.
- Coss, A. A., R. J. Bartelt, B. W. Zilkowski, D. W. Bean, and R. J. Petroski. 2005. The aggregation pheromone of *Diorhabda elongata*, a biological control agent of saltcedar (*Tamarix* spp.): identification of two behaviorally active components. *Journal of Chemical Ecology* 31:657-670.
- Cosse, A. A., R. J. Bartelt, B. W. Zilkowski, D. W. Bean, and R. J. Petroski. 2005. The aggregation pheromone of *Diorhabda elongata*, a biological control agent of saltcedar (*Tamarix* spp.): Identification of two behaviorally active components. *J. Chem. Ecol.* 31:657-670.
- Czekanski-Moir, J. E. and R. J. Rundell. 2019. The Ecology of Nonecological Speciation and Nonadaptive Radiations. *Trends Ecol. Evol.* 34:400-415.
- Dalin, P., D. W. Bean, T. L. Dudley, V. A. Carney, D. Eberts, K. T. Gardner, E. Hebertson, E. N. Jones, D. J. Kazmer, and G. J. Michels Jr. 2010. Seasonal adaptations to day length in ecotypes of *Diorhabda* spp. (Coleoptera: Chrysomelidae) inform selection of agents against saltcedars (*Tamarix* spp.). *Environmental entomology* 39:1666-1675.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. Genomes Project Analysis. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156-2158.
- De Castro, F. and B. Bolker. 2005. Mechanisms of disease-induced extinction. *Ecology Letters* 8:117-126.
- Deakin, J. E., H. S. Bender, A.-M. Pearse, W. Rens, P. C. M. O'Brien, M. A. Ferguson-Smith, Y. Cheng, K. Morris, R. Taylor, and A. Stuart. 2012. Genomic restructuring in the Tasmanian

- devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. *PLoS genetics* 8:e1002483.
- DeBach, P. and D. Rosen. 1991. *Biological Control by Natural Enemies*. Cambridge University Press, Cambridge.
- DeLoach, C. J., R. I. Carruthers, T. L. Dudley, D. Eberts, D. J. Kazmer, A. E. Knutson, D. W. Bean, J. Knight, P. A. Lewis, and L. R. Milbrath. 2004. First results for control of saltcedar (*Tamarix* spp.) in the open field in the western United States. Pp. 505.
- DeLoach, C. J., P. A. Lewis, J. C. Herr, R. I. Carruthers, J. L. Tracy, and J. Johnson. 2003. Host specificity of the leaf beetle, *Diorhabda elongata deserticola* (Coleoptera: Chrysomelidae) from Asia, a biological control agent for saltcedars (*Tamarix*: Tamaricaceae) in the Western United States. *Biological Control* 27:117-147.
- Di Tomaso, J. M. 1998. Impact, biology, and ecology of saltcedar (*Tamarix* spp.) in the southwestern United States. *Weed Technology* 12:326-336.
- Dias-Alves, T., J. Mairal, and M. G. B. Blum. 2018. Loter: a software package to infer local ancestry for a wide range of species. *Molecular biology and evolution* 35:2318-2326.
- Dlugosch, K. M., S. R. Anderson, J. Braasch, F. A. Cang, and H. D. Gillette. 2015. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology* 24:2095-2111.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular ecology resources* 14:209-214.
- Dowle, E. J., R. R. Bracewell, M. E. Pfrender, K. E. Mock, B. J. Bentz, and G. J. Ragland. 2017. Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology* 26:6071-6084.
- Drus, G. M., T. L. Dudley, M. L. Brooks, and J. R. Matchett. 2013. The effect of leaf beetle herbivory on the fire behaviour of tamarisk (*Tamarix ramosissima* Lebed.). *Int. J. Wildland Fire* 22:446-458.
- Dudchenko, O., S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, and A. P. Aiden. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92-95.
- Durand, N. C., J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, and E. L. Aiden. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* 3:99-101.
- Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21:3439-3440.
- Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* 4:1184.
- Edgar, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5:113.
- Edgar, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
- Epstein, B., M. Jones, R. Hamede, S. Hendricks, H. McCallum, E. P. Murchison, B. Schönfeld, C. Wiench, P. Hohenlohe, and A. Storfer. 2016. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature communications* 7:12684.
- Estoup, A. and T. Guillemaud. 2010. Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* 19:4113-4130.
- Estoup, A., V. Ravigne, R. Hufbauer, R. Vitalis, M. Gautier, and B. Facon. 2016. Is There a Genetic Paradox of Biological Invasion? Pp. 51-72 in D. J. Futuyma, ed. *Annual Review of Ecology, Evolution, and Systematics*, Vol 47. Annual Reviews, Palo Alto.

- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2611-2620.
- Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* 40:481-501.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.
- Fang, W. H., Q. Wang, H. M. Li, M. Ahmed, P. Kumar, and S. Kumar. 2014. PAX 3 in neuroblastoma: oncogenic potential, chemosensitivity and signalling pathways. *Journal of cellular and molecular medicine* 18:38-48.
- Faria, R. and A. Navarro. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in ecology & evolution* 25:660-669.
- Faust, G. G. and I. M. Hall. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30:2503-2505.
- Fauvergue, X., E. Vercken, T. Malausa, and R. A. Hufbauer. 2012. The biology of small, introduced populations, with special reference to biological control. *Evolutionary Applications* 5:424-443.
- Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences* 100:10314-10319.
- Feron, R. 2021. SexGenomicsToolkit/sgtr: 1.1.4. Zenodo.
- Feron, R., Q. Pan, M. Wen, B. Imarazene, E. Jouanno, J. Anderson, A. Herpin, L. Journot, H. Parrinello, and C. Klopp. 2020. RADSex: a computational workflow to study sex determination using Restriction Site-Associated DNA Sequencing data.
- Fisher, R. A. and E. B. Ford. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* 1:143-174.
- Fišer, C., C. T. Robinson, and F. Malard. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* 27:613-635.
- Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* 117:9451-9457.
- Fraik, A. K., M. J. Margres, B. Epstein, S. Barbosa, M. Jones, S. Hendricks, B. Schönfeld, A. R. Stahlke, A. Veillet, R. Hamede, H. McCallum, E. Lopez-Contreras, S. J. Kallinen, P. A. Hohenlohe, J. L. Kelley, and A. Storfer. 2020. Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *Evolution* n/a.
- Francis, R. M. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* 17:27-32.
- Francois, O., H. Martins, K. Caye, and S. D. Schoville. 2016. Controlling false discoveries in genome scans for selection. *Mol Ecol* 25:454-469.
- Frick, K. E. 1970. RAGWORT FLRA BEETLE ESTABLISHED FOR BIOLOGICAL CONTROL OF TANSY RAGWORT IN NORTHERN CALIFORNIA. *California Agriculture* 24:12-&.
- Frith, M. C. 2011. Gentle masking of low-complexity sequences improves homology search. *PLoS One* 6:e28819.
- Frith, M. C. and R. Kawaguchi. 2015. Split-alignment of genomes finds orthologies more accurately. *Genome biology* 16:1-17.
- Fu, S., C. Chen, L. Xiao, H. He, and F. Xue. 2015. Inheritance of diapause in crosses between the northernmost and the southernmost strains of the Asian corn borer *Ostrinia furnacalis*. *PLoS One* 10:e0118186.

- Funk, D. J., K. E. Filchak, and J. L. Feder. 2002. Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetics of mate choice: From sexual selection to sexual isolation*:251-267.
- Gaffke, A. M., S. E. Sing, T. L. Dudley, D. W. Bean, J. A. Russak, A. Mafra-Neto, P. A. Grieco, R. K. Peterson, and D. K. Weaver. 2018. Semiochemicals to enhance herbivory by *Diorhabda carinulata* aggregations in saltcedar (*Tamarix* spp.) infestations. *Pest Manag Sci* 74:1494-1503.
- Gandon, S. and Y. Michalakis. 2000. Evolution of parasite virulence against qualitative or quantitative host resistance. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267:985-990.
- Gaskin, J. F. and B. A. Schaal. 2002. Hybrid *Tamarix* widespread in U.S. invasion and undetected in native Asian range. *Proceedings of the National Academy of Sciences of the United States of America* 99:11256-11259.
- Godec, J., Y. Tan, A. Liberzon, P. Tamayo, S. Bhattacharya, A. J. Butte, J. P. Mesirov, and W. N. Haining. 2016. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. *Immunity* 44:194-206.
- Goldman-Huertas, B., R. F. Mitchell, R. T. Lapoint, C. P. Faucher, J. G. Hildebrand, and N. K. Whiteman. 2015. Evolution of herbivory in *Drosophilidae* linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proceedings of the National Academy of Sciences* 112:3026-3031.
- Gompert, Z. 2016. Bayesian inference of selection in a heterogeneous environment from genetic time-series data. *Mol Ecol* 25:121-134.
- Griffin, B. P., J. L. Chandler, J. C. Andersen, N. P. Havill, and J. S. Elkinton. 2020. The Reliability of Genitalia Morphology to Monitor the Spread of the Invasive Winter Moth (Lepidoptera: Geometridae) in Eastern North America. *Environmental Entomology*.
- Grueber, C. E., E. Peel, B. Wright, C. J. Hogg, and K. Belov. 2019. A Tasmanian devil breeding program to support wild recovery. *Reproduction, Fertility and Development* 31:1296-1304.
- Guerrero, R. F. and M. W. Hahn. 2018. Quantifying the risk of hemiplasy in phylogenetic inference. *Proceedings of the National Academy of Sciences* 115:12787-12792.
- Guiler, E. 1992. *The Tasmanian devil*. St. David's Park Pub.
- Hagen, R. H. and J. M. Scriber. 1989. Sex-linked diapause, color, and allozyme loci in *Papilio glaucus*: linkage analysis and significance in a hybrid zone. *Journal of Heredity* 80:179-185.
- Haller, B. C. and P. W. Messer. 2019. SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Molecular biology and evolution* 36:632-637.
- Hamede, R. K., H. McCallum, and M. Jones. 2008. Seasonal, demographic and density-related patterns of contact between Tasmanian devils (*Sarcophilus harrisii*): Implications for transmission of devil facial tumour disease. *Austral Ecology* 33:614-622.
- Hamede, R. K., H. McCallum, and M. Jones. 2013. Biting injuries and transmission of Tasmanian devil facial tumour disease. *Journal of Animal Ecology* 82:182-190.
- Hamede, R. K., A. M. Pearse, K. Swift, L. A. Barmuta, E. P. Murchison, and M. E. Jones. 2015. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. *Proc Biol Sci* 282.
- Hamilton, D. G., M. E. Jones, E. Z. Cameron, H. McCallum, A. Storfer, P. A. Hohenlohe, and R. K. Hamede. 2019. Rate of intersexual interactions affects injury likelihood in Tasmanian devil contact networks. *Behavioral Ecology* 30:1087-1095.
- Hawkins, C. E., C. Baars, H. Hesterman, G. J. Hocking, M. E. Jones, B. Lazenby, D. Mann, N. Mooney, D. Pemberton, and S. Pyecroft. 2006. Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation* 131:307-324.

- Hendricks, S., B. Epstein, B. Schönfeld, C. Wiench, R. Hamede, M. Jones, A. Storfer, and P. Hohenlohe. 2017. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). *Conservation Genetics* 18:977-982.
- Herrera, A. M., D. D. Dahlsten, N. Tomic-Carruthers, and R. I. Carruthers. 2005. Estimating temperature-dependent developmental rates of *Diorhabda elongata* (Coleoptera: Chrysomelidae), a biological control agent of saltcedar (*Tamarix* spp.). *Environmental entomology* 34:775-784.
- Hogg, C. J., A. V. Lee, C. Srb, and C. Hibbard. 2017. Metapopulation management of an Endangered species with limited genetic diversity in the presence of disease: the Tasmanian devil *Sarcophilus harrisii*. *International Zoo Yearbook* 51:137-153.
- Hohenlohe, P. A., H. I. McCallum, M. E. Jones, M. F. Lawrance, R. K. Hamede, and A. Storfer. 2019. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. *Conservation Genetics* 20:81-87.
- Hopper, J. V., K. F. McCue, P. D. Pratt, P. Duchesne, E. D. Grosholz, and R. A. Hufbauer. 2019. Into the weeds: matching importation history to genetic consequences and pathways in two widely used biological control agents. *Evolutionary applications* 12:773-790.
- Hubert, J.-N., T. Zerjal, and F. Hospital. 2018. Cancer-and behavior-related genes are targeted by selection in the Tasmanian devil (*Sarcophilus harrisii*). *PloS one* 13:e0201838.
- Hudgeons, J. L., A. E. Knutson, C. J. DeLoach, K. M. Heinz, W. A. McGinty, and J. L. Tracy. 2007. Establishment and Biological Success of *Diorhabda elongata elongata* 1 on Invasive *Tamarix* in Texas. *Southwestern Entomologist* 32:157-168.
- Hufbauer, R. A. 2008. Biological invasions: Paradox lost and paradise gained. *Current Biology* 18:R246-R247.
- Hultine, K. R., D. W. Bean, T. L. Dudley, and C. A. Gehring. 2015. Species introductions and their cascading impacts on biotic interactions in desert riparian ecosystems. *Integrative and comparative biology* 55:587-601.
- Hultine, K. R., J. Belnap, C. van Riper Iii, J. R. Ehleringer, P. E. Dennison, M. E. Lee, P. L. Nagler, K. A. Snyder, S. M. Uselman, and J. B. West. 2010. Tamarisk biocontrol in the western United States: ecological and societal implications. *Frontiers in Ecology and the Environment* 8:467-474.
- Hutchison, D. W. and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898-1914.
- Ikten, C., S. R. Skoda, T. E. Hunt, J. Molina-Ochoa, and J. E. Foster. 2011. Genetic variation and inheritance of diapause induction in two distinct voltine ecotypes of *Ostrinia nubilalis* (Lepidoptera: Crambidae). *Annals of the Entomological Society of America* 104:567-575.
- James, S., G. Jennings, Y. M. Kwon, M. Stammnitz, A. Fraik, A. Storfer, S. Comte, D. Pemberton, S. Fox, and B. Brown. 2019. Tracing the rise of malignant cell lines: Distribution, epidemiology and evolutionary interactions of two transmissible cancers in Tasmanian devils. *Evolutionary applications* 12:1772-1780.
- Ji, W., L. Wang, and A. E. Knutson. 2017. Detection of the spatiotemporal patterns of beetle-induced tamarisk (*Tamarix* spp.) defoliation along the Lower Rio Grande using Landsat TM images. *Remote Sensing of Environment* 193:76-85.
- Johnson, R. N., D. O'Meally, Z. Chen, G. J. Etherington, S. Y. W. Ho, W. J. Nash, C. E. Grueber, Y. Cheng, C. M. Whittington, and S. Dennison. 2018. Adaptation and conservation insights from the koala genome. *Nature genetics* 50:1102.
- Johnston, J. S., A. Bernardini, and C. E. Hjelmen. 2019. Genome size estimation and quantitative cytogenetics in insects. Pp. 15-26. *Insect genomics*. Springer.
- Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. *Nature* 451:990-993.

- Jorde, P. E. and N. Ryman. 2007. Unbiased estimator for genetic drift and effective population size. *Genetics* 177:927-935.
- Joshi, N. and J. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.
- Kardos, M. and A. B. A. Shafer. 2018. The Peril of Gene-Targeted Conservation. *Trends Ecol Evol* 33:827-839.
- Kelly, J. K. and K. A. Hughes. 2019. Pervasive linked selection and intermediate-frequency alleles are implicated in an evolve-and-resequencing experiment of *Drosophila simulans*. *Genetics* 211:943-961.
- Kim, D., B. Langmead, and S. L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12:357-360.
- Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* 37:907-915.
- Kingan, S. B., J. Urban, C. C. Lambert, P. Baybayan, A. K. Childers, B. Coates, B. Scheffler, K. Hackett, J. Korfach, and S. M. Geib. 2019. A high-quality genome assembly from a single, field-collected spotted lanternfly (*Lycorma delicatula*) using the PacBio Sequel II system. *GigaScience* 8:giz122.
- Knutson, A. E., C. J. DeLoach, J. L. Tracy, and C. W. Randal. 2012. Field Evaluation of *Diorhabda elongata* and *D. carinata* (Coleoptera: Chrysomelidae) for Biological Control of Saltcedars (*Tamarix* spp.) in Northwest Texas. *Southwestern Entomologist* 37:91-102.
- Knutson, A. E., J. L. Tracy, C. Ritzi, P. J. Moran, T. Royer, and C. J. DeLoach. 2019. Establishment, Hybridization, Dispersal, Impact, and Decline of *Diorhabda* spp. (Coleoptera: Chrysomelidae) Released for Biological Control of Tamarisk in Texas and New Mexico. *Environ. Entomol.* 48:1297-1316.
- Kolbe, J. J., R. E. Glor, L. R. Schettino, A. C. Lara, A. Larson, and J. B. Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177-181.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356.
- Kosiol, C., T. Vinař, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante, R. Nielsen, and A. Siepel. 2008. Patterns of positive selection in six mammalian genomes. *PLoS genetics* 4.
- Kovaka, S., A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, and M. Pertea. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* 20:1-13.
- Lafferty, K. D. 2009. The ecology of climate change and infectious diseases. *Ecology* 90:888-900.
- Lakha, W., I. Panteleeva, S. Squazzo, R. Saxena, J. Kroonen, S. Siembieda, M. Tilmes, and J. Hagopian. 2016. DNA fragmentation and quality control analysis using Diagenode shearing systems and Fragment Analyzer. *Nature Methods* 13:iii-iv.
- Landguth, E. L., B. R. Forester, A. J. Eckert, A. J. Shirk, M. Menon, A. Whipple, C. C. Day, and S. A. Cushman. 2020. Modelling multilocus selection in an individual-based, spatially-explicit landscape genetics framework. *Molecular ecology resources* 20:605-615.
- Langmead, B. and S. L. Salzberg. 2012a. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Langmead, B. and S. L. Salzberg. 2012b. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9:357.
- Lasne, C., C. M. Sgrò, and T. Connallon. 2017. The relative contributions of the X chromosome and autosomes to local adaptation. *Genetics* 205:1285-1304.
- Lazenby, B. T., M. W. Tobler, W. E. Brown, C. E. Hawkins, G. J. Hocking, F. Hume, S. Huxtable, P. Iles, M. E. Jones, and C. Lawrence. 2018. Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. *Journal of Applied Ecology* 55:1368-1379.

- Lee, S. R., Y. S. Jo, C. H. Park, J. M. Friedman, and M. S. Olson. 2018. Population genomic analysis suggests strong influence of river network on spatial distribution of genetic variation in invasive saltcedar across the southwestern United States. *Molecular ecology* 27:636-646.
- Leung, K., E. Ras, K. B. Ferguson, S. Ariens, D. Babendreier, P. Bijma, K. Bourtzis, J. Brodeur, M. A. Bruins, A. Centurion, S. R. Chattington, M. Chinchilla-Ramirez, M. Dicke, N. E. Fatouros, J. Gonzalez-Cabrera, T. V. M. Groot, T. Haye, M. Knapp, P. Koskinioti, S. Le Hesran, M. Lyrakis, A. Paspati, M. Perez-Hedo, W. N. Plouvier, C. Schlotterer, J. M. Stahl, A. Thiel, A. Urbaneja, L. van de Zande, E. C. Verhulst, L. E. M. Vet, S. Visser, J. H. Werren, S. W. Xia, B. J. Zwaan, S. Magalhaes, L. W. Beukeboom, and B. A. Pannebakker. 2020. Next-generation biological control: the need for integrating genetics and genomics. *Biological Reviews* 95:1838-1854.
- Levy Karin, Eli, Milot Mirdita, and Johannes Söding. 2020. MetaEuk-Sensitive, High-Throughput Gene Discovery, and Annotation for Large-Scale Eukaryotic Metagenomics. *Microbiome* 8: 48.
- Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, R. Durbin, S. V. Edwards, F. Forest, and M. T. P. Gilbert. 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences* 115:4325-4333.
- Lewis, P. A., C. J. DeLoach, A. E. Knutson, J. L. Tracy, and T. O. Robbins. 2003. Biology of *Diorhabda elongata deserticola* (Coleoptera: Chrysomelidae), an Asian leaf beetle for biological control of saltcedars (*Tamarix* spp.) in the United States. *Biological Control* 27:101-116.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of molecular evolution* 36:96-99.
- Li, Y., H. Park, T. E. Smith, and N. A. Moran. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular biology and evolution* 36:2143-2156.
- Liberzon, A., C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo. 2015. The molecular signatures database hallmark gene set collection. *Cell systems* 1:417-425.
- Lommen, S. T. E., P. W. de Jong, and B. A. Pannebakker. 2017. It is time to bridge the gap between exploring and exploiting: prospects for utilizing intraspecific genetic variation to optimize arthropods for augmentative pest control - a review. *Entomologia Experimentalis Et Applicata* 162:108-123.
- Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* 89:238-247.
- Ma, Y., X. Ding, S. Qanbari, S. Weigend, Q. Zhang, and H. Simianer. 2015. Properties of different selection signature statistics and a new strategy for combining them. *Heredity* 115:426.
- Margres, M. J., M. E. Jones, B. Epstein, D. H. Kerlin, S. Comte, S. Fox, A. K. Fraik, S. A. Hendricks, S. Huxtable, and S. Lachish. 2018a. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. *Molecular ecology* 27:4189-4199.
- Margres, M. J., M. Ruiz-Aravena, R. Hamede, K. Chawla, A. H. Patton, M. F. Lawrance, A. K. Fraik, A. R. Stahlke, B. W. Davis, and E. A. Ostrander. 2020. Spontaneous tumor regression in Tasmanian devils associated with RASL11A activation. *Genetics*.
- Margres, M. J., M. Ruiz-Aravena, R. Hamede, M. E. Jones, M. F. Lawrance, S. A. Hendricks, A. Patton, B. W. Davis, E. A. Ostrander, and H. McCallum. 2018b. The genomic basis of tumor regression in Tasmanian devils (*Sarcophilus harrisii*). *Genome biology and evolution* 10:3012-3025.

- Marsico, T. D., J. W. Burt, E. K. Espeland, G. W. Gilchrist, M. A. Jamieson, L. Lindström, G. K. Roderick, S. Swope, M. Szűcs, and N. D. Tsutsui. 2010. PERSPECTIVE: Underutilized resources for studying the evolution of invasive species during their introduction, establishment, and lag phases. *Evolutionary Applications* 3:203-219.
- Maruki, T. and M. Lynch. 2017. Genotype Calling from Population-Genomic Sequencing Data. G3 (Bethesda) 7:1393-1404.
- Marçais, G. and C. Kingsford. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764-770.
- Massey Jr, F. J. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* 46:68-78.
- Mathieson, I. and G. McVean. 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* 193:973-984.
- McCallum, H. and A. Dobson. 1995. Detecting disease and parasite threats to endangered species and ecosystems. *Trends in ecology & evolution* 10:190-194.
- McCallum, H., M. Jones, C. Hawkins, R. Hamede, S. Lachish, D. L. Sinn, N. Beeton, and B. Lazenby. 2009. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. *Ecology* 90:3379-3392.
- McFadyen, R. E. C. 1998. Biological control of weeds. *Annu. Rev. Entomol.* 43:369-393.
- McFarlane, S. E. and J. M. Pemberton. 2019. Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends Ecol. Evol.* 34:315-326.
- McGinnis, S. and T. L. Madden. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research* 32:W20-W25.
- McKinney, G. J., W. A. Larson, L. W. Seeb, and J. E. Seeb. 2017. RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al.(2016). *Molecular ecology resources* 17:356-361.
- McKnight, D. T., L. Schwarzkopf, R. A. Alford, D. S. Bower, and K. R. Zenger. 2017. Effects of emerging infectious diseases on host population genetics: a review. *Conservation genetics* 18:1235-1245.
- McLennan, E. A., C. E. Grueber, P. Wise, K. Belov, and C. J. Hogg. 2020. Mixing genetically differentiated populations successfully boosts diversity of an endangered carnivore. *Animal Conservation*.
- Meek, M. H. and W. A. Larson. 2019. The future is now: amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular ecology resources*.
- Meinhardt, K. A. and C. A. Gehring. 2012. Disrupting mycorrhizal mutualisms: a potential mechanism by which exotic tamarisk outcompetes native cottonwoods. *Ecological Applications* 22:532-549.
- Meirmans, P. G. 2019. Subsampling reveals that unbalanced sampling affects STRUCTURE results in a multi-species dataset. *Heredity* 122:276-287.
- Meisel, R. P. and T. Connallon. 2013. The faster-X effect: integrating theory and data. *Trends in genetics* 29:537-544.
- Metzger, M. J., C. Reinisch, J. Sherry, and S. P. Goff. 2015. Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* 161:255-263.
- Meurs, E. F., J. Galabru, G. N. Barber, M. G. Katze, and A. G. Hovanessian. 1993. Tumor suppressor function of the interferon-induced double-stranded RNA-activated protein kinase. *Proceedings of the National Academy of Sciences* 90:232-236.
- Mi, H., A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas. 2018. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research* 47:D419-D426.
- Michels, G. J., T. A. Royer, E. N. Jones, R. A. Lange, E. D. Bynum, D. C. Ruthven, J. L. Tracy, and J. B. Bible. 2013. New Establishment and County Records for *Diorhabda* spp. (Coleoptera:

- Chrysomelidae) and *Coniatus splendidulus* (Coleoptera: Curculionidae) in the Texas Panhandle and Western Oklahoma. *Southwestern Entomologist* 38:173-181.
- Mikheenko, A., A. Prjibelski, V. Saveliev, D. Antipov, and A. Gurevich. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34: i142–i150.
- Mikkelsen, T. S., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, and A. Heger. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167-177.
- Miller, I. F. and C. J. E. Metcalf. 2019. Evolving resistance to pathogens. *Science* 363:1277-1278.
- Mitchell, K. J., R. C. Pratt, L. N. Watson, G. C. Gibb, B. Llamas, M. Kasper, J. Edson, B. Hopwood, D. Male, and K. N. Armstrong. 2014. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Molecular biology and evolution* 31:2322-2330.
- Moll, P., M. Ante, A. Seitz, and T. Reda. 2014. QuantSeq 3' mRNA sequencing for RNA quantification. *Nature methods* 11:972.
- Muller-Scharer, H., S. Bouchemousse, M. Litto, P. B. McEvoy, G. K. Roderick, and Y. Sun. 2020. How to better predict long-term benefits and risks in weed biocontrol: an evolutionary perspective. *Curr. Opin. Insect Sci.* 38:84-91.
- Murchison, E. P., O. B. Schulz-Trieglaff, Z. Ning, L. B. Alexandrov, M. J. Bauer, B. Fu, M. Hims, Z. Ding, S. Ivakhno, and C. Stewart. 2012. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* 148:780-791.
- Murchison, E. P., C. Tovar, A. Hsu, H. S. Bender, P. Kheradpour, C. A. Rebbeck, D. Obendorf, C. Conlan, M. Bahlo, and C. A. Blizzard. 2010. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. *Science* 327:84-87.
- Müller-Schärer, H., S. T. E. Lommen, M. Rossinelli, M. Bonini, M. Boriani, G. Bosio, and U. Schaffner. 2014. *Ophraella communa*, the ragweed leaf beetle, has successfully landed in Europe: fortunate coincidence or threat? *Weed research* 54:109-119.
- Nagler, P. L., E. P. Glenn, C. S. Jarnevich, and P. B. Shafroth. 2011. Distribution and Abundance of Saltcedar and Russian Olive in the Western United States. *Crit. Rev. Plant Sci.* 30:508-523.
- Nagler, P. L., S. Pearlstein, E. P. Glenn, T. B. Brown, H. L. Bateman, D. W. Bean, and K. R. Hultine. 2014. Rapid dispersal of saltcedar (*Tamarix* spp.) biocontrol beetles (*Diorhabda carinulata*) on a desert river detected by phenocams, MODIS imagery and ground observations. *Remote Sens. Environ.* 140:206-219.
- Nelson, W. J. and R. Nusse. 2004. Convergence of Wnt, β -catenin, and cadherin pathways. *Science* 303:1483-1487.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences* 98:12084-12088.
- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry H. Stevens, E. Szoecs, and H. Wagner. 2019. *vegan: Community Ecology Package*.
- Ozsoy, A. Z., A. R. Stahlke, L. Jamison, and M. J. Johnson. 2019. Genetic Identification and Hybrid Analysis of Tamarisk Leaf Beetle (*Diorhabda* spp.) and Tamarisk Weevil (*Coniatus* spp.) along the Rio Grande River NM watershed. Army Corps of Engineers.
- Ozsoy, A. Z., A. R. Stahlke, and M. J. Johnson. 2018. Genetic Identification and Hybrid Analysis of Tamarisk Leaf Beetle (*Diorhabda* spp.) and Tamarisk Weevil (*Coniatus* spp.) along the Rio Grande River NM watershed. Army Corps of Engineers.
- Ozsoy, A. Z., A. R. Stahlke, and M. J. Johnson. 2021. Genetic Identification and Hybrid Analysis of Tamarisk Leaf Beetles (*Diorhabda* spp.) along the Rio Grande River NM watershed. Tetra Tech, Albuquerque, NM.
- Padial, J. M., A. Miralles, I. De la Riva, and M. Vences. 2010. The integrative future of taxonomy. *Frontiers in zoology* 7:1-14.

- Patchett, A. L., A. S. Flies, A. B. Lyons, and G. M. Woods. 2020. Curse of the devil: molecular insights into the emergence of transmissible cancers in the Tasmanian devil (*Sarcophilus harrisii*). *Cellular and Molecular Life Sciences*.
- Patton, A. H., M. F. Lawrance, M. J. Margres, C. P. Kozakiewicz, R. Hamede, M. Ruiz-Aravena, D. G. Hamilton, S. Comte, L. E. Ricci, and R. L. Taylor. 2020. A transmissible cancer shifts from emergence to endemism in Tasmanian devils. *Science* 370.
- Patton, A. H., M. J. Margres, A. R. Stahlke, S. Hendricks, K. Lewallen, R. K. Hamede, M. Ruiz-Aravena, O. Ryder, H. I. McCallum, and M. E. Jones. 2019. Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian Devils. *Molecular biology and evolution* 36:2906-2921.
- Pearse, A. M. and K. Swift. 2006. Allograft theory: transmission of devil facial-tumour disease. *Nature* 439:549.
- Peck, S. J., S. A. Michael, G. Knowles, A. Davis, and D. Pemberton. 2019. Causes of mortality and severe morbidity requiring euthanasia in captive Tasmanian devils (*Sarcophilus harrisii*) in Tasmania. *Australian veterinary journal* 97:89-92.
- Perte, G. and M. Perte. 2020. GFF utilities: GffRead and GffCompare. *F1000Research* 9.
- Peter, B. M. and M. Slatkin. 2013. Detecting range expansions from genetic data. *Evolution* 67:3274-3289.
- Petersen, M., D. Armisen, R. A. Gibbs, L. Hering, A. Khila, G. Mayer, S. Richards, O. Niehuis, and B. Misof. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC evolutionary biology* 19:1-15.
- Petit, R. J. and L. Excoffier. 2009. Gene flow and species delimitation. *Trends in Ecology & evolution* 24:386-393.
- Pratt, P. D., J. C. Herr, R. I. Carruthers, M. J. Pitcairn, B. Viellgas, and M. B. Kelley. 2019. Release, establishment and realised geographic distribution of *Diorhabda carinulata* and *D. elongata* (Coleoptera: Chrysomelidae) in California, USA. *Biocontrol Science and Technology* 29:686-705.
- Presgraves, D. C. 2008. Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics* 24:336-343.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Pruisscher, P., S. Nylin, K. Gotthard, and C. W. Wheat. 2018. Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. *Molecular ecology* 27:3613-3626.
- Pye, R., R. Hamede, H. V. Siddle, A. Caldwell, G. W. Knowles, K. Swift, A. Kreiss, M. E. Jones, A. B. Lyons, and G. M. Woods. 2016a. Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. *Biol Lett* 12.
- Pye, R. J., D. Pemberton, C. Tovar, J. M. C. Tubio, K. A. Dun, S. Fox, J. Darby, D. Hayes, G. W. Knowles, and A. Kreiss. 2016b. A second transmissible cancer in Tasmanian devils. *Proceedings of the National Academy of Sciences* 113:374-379.
- Quinlan, A. R. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Current protocols in bioinformatics* 47:11-12.
- Rao, S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, and E. S. Lander. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665-1680.
- Reid, K., J. Carlos Garza, S. R. Gephard, A. Caccone, D. M. Post, and E. P. Palkovacs. 2020. Restoration-mediated secondary contact leads to introgression of alewife ecotypes separated by a colonial-era dam. *Evolutionary applications* 13:652-664.
- Renfree, M. B., A. T. Papenfuss, J. E. Deakin, J. Lindsay, T. Heider, K. Belov, W. Rens, P. D. Waters, E. A. Pharo, and G. Shaw. 2011. Genome sequence of an Australian kangaroo,

- Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome biology* 12:1-26.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. *Trends in genetics* 16:276-277.
- Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends in ecology & evolution* 16:351-358.
- Rieseberg, L. H., N. C. Ellstrand, and M. Arnold. 1993. What can molecular and morphological markers tell us about plant hybridization? *Critical reviews in plant sciences* 12:213-241.
- Rius, M. and J. A. Darling. 2014. How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol. Evol.* 29:233-242.
- Robich, R. M., J. P. Rinehart, L. J. Kitchen, and D. L. Denlinger. 2007. Diapause-specific gene expression in the northern house mosquito, *Culex pipiens* L., identified by suppressive subtractive hybridization. *Journal of insect physiology* 53:235-245.
- Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019a. Stacks 2: Analytical Methods for Paired-end Sequencing Improve RADseq-based Population Genomics. *bioRxiv:615385*.
- Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019b. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28:4737-4754.
- Roderick, G. K. and M. Navajas. 2003. Genes in new environments: Genetics and evolution in biological control. *Nature Reviews Genetics* 4:889-899.
- Room, P. M., K. L. S. Harley, I. W. Forno, and D. P. A. Sands. 1981. SUCCESSFUL BIOLOGICAL-CONTROL OF THE FLOATING WEED SALVINIA. *Nature* 294:78-80.
- Rundle, H. D. and P. Nosil. 2005. Ecological speciation. *Ecology letters* 8:336-352.
- Sands, D. P. A. and K. L. S. Harley. 1980. Importance of geographic variation in agents selected for biological control of weeds. Pp. 81-90. *Proceedings of the V International Symposium on Biological Control of Weeds*. CSIRO, Melbourne, Australia.
- Sasibhushan, S., K. M. Ponnuvel, and N. B. Vijayaprakash. 2012. Diapause specific gene expression in the eggs of multivoltine silkworm *Bombyx mori*, identified by suppressive subtractive hybridization. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 161:371-379.
- Schild, D. R., E. S. C. Scordato, C. C. R. Smith, J. K. Carter, S. I. Cherkaoui, S. Gombobaatar, S. Hajib, S. Hanane, A. K. Hund, and K. Koyama. 2021. Sex-linked genetic diversity and differentiation in a globally distributed avian species complex. *Molecular Ecology*.
- Scholz, F. W. and M. A. Stephens. 1987. K-sample Anderson-Darling tests. *Journal of the American Statistical Association* 82:918-924.
- Schoville, S. D., Y. H. Chen, M. N. Andersson, J. B. Benoit, A. Bhandari, J. H. Bowsher, K. Brevik, K. Cappelle, M.-J. M. Chen, and A. K. Childers. 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific reports* 8:1-18.
- Schrader, L. and J. Schmitz. 2019. The impact of transposable elements in adaptive evolution. *Molecular Ecology* 28:1537-1549.
- Schwarzlander, M., H. L. Hinz, R. L. Winston, and M. D. Day. 2018. Biological control of weeds: an analysis of introductions, rates of establishment and estimates of success, worldwide. *Biocontrol* 63:319-331.
- Sethuraman, A., F. J. Janzen, D. W. Weisrock, and J. J. Obrycki. 2020. Insights from Population Genomics to Enhance and Sustain Biological Control of Insect Pests. *Insects* 11.
- Shaffer, J. P. 1995. Multiple hypothesis testing. *Annual review of psychology* 46:561-584.
- Shafroth, P. B., J. R. Cleverly, T. L. Dudley, J. P. Taylor, C. Van Riper, E. P. Weeks, and J. N. Stuart. 2005. Control of *Tamarix* in the Western United States: Implications for water salvage, wildlife use, and riparian restoration. *Environ. Manage.* 35:231-246.
- Shen, L. and M. Sinai. 2020. GeneOverlap: Test and visualize gene overlaps. version. R package.

- Shultz, A. J. and T. B. Sackton. 2019. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife* 8:e41815.
- Siddle, H. V., A. Kreiss, M. D. B. Eldridge, E. Noonan, C. J. Clarke, S. Pyecroft, G. M. Woods, and K. Belov. 2007. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences* 104:16221-16226.
- Siddle, H. V., A. Kreiss, C. Tovar, C. K. Yuen, Y. Cheng, K. Belov, K. Swift, A.-M. Pearse, R. Hamede, and M. E. Jones. 2013. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proceedings of the National Academy of Sciences* 110:5103-5108.
- Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Simmonds, F. J. 1963. Genetics and Biological Control. *The Canadian Entomologist* 95:561-567.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Sinnott, R. W. 1984. VIRTUES OF THE HAVERSINE. *Sky and Telescope* 68:159-159.
- Slatkin, M. and L. Excoffier. 2012. Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* 191:171-181.
- Smit, A. F. A., R. Hubley, and P. Green. 2015. RepeatMasker Open-4.0. 2013–2015.
- Smith, K. F., D. F. Sax, and K. D. Lafferty. 2006. Evidence for the role of infectious disease in species extinction and endangerment. *Conservation biology* 20:1349-1357.
- Smith, L., M. Cristofaro, M.-C. Bon, A. De Biase, R. Petanović, and B. Vidović. 2018. The importance of cryptic species and subspecific populations in classic biological control of weeds: a North American perspective. *Biocontrol* 63:417-425.
- Sogge, M. K., S. J. Sferra, and E. H. Paxton. 2008. Tamarix as habitat for birds: implications for riparian restoration in the southwestern United States. *Restoration Ecology* 16:146-154.
- Sota, T. and K. Kubota. 1998. Genital lock-and-key as a selective agent against hybridization. *Evolution* 52:1507-1513.
- Stahlke, A., D. Bell, T. Dhendup, B. Kern, S. Pannoni, Z. Robinson, J. Strait, S. Smith, B. K. Hand, P. A. Hohenlohe, and G. Luikart. 2020. Population genomics training for the next generation of conservation geneticists: ConGen 2018 Workshop. *J. Hered.*
- Stahlke, A. R., A. Z. Ozsoy, D. W. Bean, and P. A. Hohenlohe. 2019. Mitochondrial genome sequences of *Diorhabda carinata* and *Diorhabda carinulata*, two beetle species introduced to North America for biological control. *Microbiology resource announcements* 8.
- Stammnitz, M. R., T. H. H. Coorens, K. C. Gori, D. Hayes, B. Fu, J. Wang, D. E. Martin-Herranz, L. B. Alexandrov, A. Baez-Ortega, and S. Barthorpe. 2018. The origins and vulnerabilities of two transmissible cancers in Tasmanian devils. *Cancer Cell* 33:607-619.
- Stapley, J., A. W. Santure, and S. R. Dennis. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular ecology* 24:2241-2252.
- Storfer, A., C. P. Kozakiewicz, M. A. Beer, and A. E. Savage. 2020. Applications of Population Genomics for Understanding and Mitigating Wildlife Disease.
- Strakova, A. and E. P. Murchison. 2015. The cancer which survived: insights from the genome of an 11000 year-old cancer. *Curr Opin Genet Dev* 30:49-55.
- Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, and E. S. Lander. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102:15545-15550.
- Sullivan, R. T. 1981. Insect swarming and mating. *The Florida Entomologist* 64:44-65.

- Sun, Y., C. Beuchat, and H. Muller-Scharer. 2020. Is biocontrol efficacy rather driven by the plant or the antagonist genotypes? A conceptual bioassay approach. *Neobiota*:81-100.
- Szkiba, D., M. Kapun, A. von Haeseler, and M. Gallach. 2014. SNP2GO: functional analysis of genome-wide association studies. *Genetics* 197:285-289.
- Szucs, M., S. D. Eigenbrode, M. Schwarzlander, and U. Schaffner. 2012. Hybrid vigor in the biological control agent, *Longitarsus jacobaeae*. *Evol Appl* 5:489-497.
- Szöllősi, G. J., E. Tannier, V. Daubin, and B. Boussau. 2015. The inference of gene trees with species trees. *Systematic biology* 64:e42-e62.
- Szűcs, M., E. I. Clark, U. Schaffner, J. L. Littlefield, C. Hoover, and R. A. Hufbauer. 2021. The effects of intraspecific hybridization on the host specificity of a weed biocontrol agent. *Biological Control*:104585.
- Szűcs, M., U. Schaffner, W. J. Price, and M. Schwarzländer. 2012. Post-introduction evolution in the biological control agent *Longitarsus jacobaeae* (Coleoptera: Chrysomelidae). *Evolutionary Applications* 5:858-868.
- Szűcs, M., M. Schwarzlander, and J. F. Gaskin. 2011. Reevaluating establishment and potential hybridization of different biotypes of the biological control agent *Longitarsus jacobaeae* using molecular tools. *Biological Control* 58:44-52.
- Szűcs, M., E. Vercken, E. V. Bitume, and R. A. Hufbauer. 2019. The implications of rapid eco-evolutionary processes for biological control - a review. *Entomologia Experimentalis et Applicata*.
- Talevich, E., B. M. Invergo, P. J. A. Cock, and B. A. Chapman. 2012. Bio. Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC bioinformatics* 13:1-9.
- Talla, V., A. Suh, F. Kalsoom, V. Dincă, R. Vila, M. Friberg, C. Wiklund, and N. Backström. 2017. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (*Leptidea*) butterflies. *Genome Biology and Evolution* 9:2491-2505.
- Tracy, J. L. and T. O. Robbins. 2009. Taxonomic revision and biogeography of the *Tamarix*-feeding *Diorhabda elongata* (Brulle, 1832) species group (Coleoptera: Chrysomelidae: Galerucinae: Galerucini) and analysis of their potential in biological control of Tamarisk. *Zootaxa*:1-152.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 3:e285.
- Van Klinken, R. D. and O. R. Edwards. 2002. Is host-specificity of weed biological control agents likely to evolve rapidly following establishment? *Ecology Letters* 5:590-596.
- von Thaden, A., C. Nowak, A. Tiesmeyer, T. E. Reiners, P. C. Alves, L. A. Lyons, F. Mattucci, E. Randi, M. Cragolini, and J. Galián. 2020. Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources* 20.
- Vurture, G. W., F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, and M. C. Schatz. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202-2204.
- Waples, R. S. 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? *J. Hered.* 106:1-19.
- Warmuth, V. M. and H. Ellegren. 2019. Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Molecular ecology resources*.
- Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*.
- Wayne, R. K. and S. M. Jenks. 1991. Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf *Canis rufus*. *Nature* 351:565-568.
- Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe. 2017. Direct determination of diploid genome sequences. *Genome Res* 27:757-767.

- Wells, K., R. K. Hamede, M. E. Jones, P. A. Hohenlohe, A. Storfer, and H. I. McCallum. 2019. Individual and temporal variation in pathogen load predicts long-term impacts of an emerging infectious disease. *Ecology* 100:e02613.
- Wertheim, B., E.-J. A. van Baalen, M. Dicke, and L. E. M. Vet. 2005. Pheromone-mediated aggregation in nonsocial arthropods: an evolutionary ecological perspective. *Annu. Rev. Entomol.* 50:321-346.
- Wickham, H. 2016. *Ggplot2: Elegant graphics for data analysis*. Springer.
- Wickham, H., R. Francois, L. Henry, and K. Müller. 2016. *dplyr: A Grammar of Data Manipulation*: <https://CRAN.R-project.org/package=dplyr>.
- Williams, W. I., J. M. Friedman, J. F. Gaskin, and A. P. Norton. 2014. Hybridization of an invasive shrub affects tolerance and resistance to defoliation by a biological control agent. *Evolutionary Applications* 7:381-393.
- Wright, B., K. Morris, C. E. Grueber, C. E. Willet, R. Gooley, C. J. Hogg, D. O'Meally, R. Hamede, M. Jones, C. Wade, and K. Belov. 2015. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. *BMC Genomics* 16:791.
- Wright, B., C. E. Willet, R. Hamede, M. Jones, K. Belov, and C. M. Wade. 2017. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. *Sci Rep* 7:423.
- Wright, B. R., K. A. Farquharson, E. A. McLennan, K. Belov, C. J. Hogg, and C. E. Grueber. 2020. A demonstration of conservation genomics for threatened species management. *Molecular Ecology Resources*.
- Wu, K., Z. Li, S. Cai, L. Tian, K. Chen, J. Wang, J. Hu, Y. Sun, X. Li, and A. Ertel. 2013. EYA1 phosphatase function is essential to drive breast cancer cell proliferation through cyclin D1. *Cancer research* 73:4488-4499.
- Yamashita, O., K. Shiomi, Y. Ishida, N. Katagiri, and T. Niimi. 2001. Insights for future studies on embryonic diapause promoted by molecular analyses of diapause hormone and its action in *Bombyx mori*. Pp. 145-153. *Insect timing: Circadian rhythmicity to seasonality*. Elsevier.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13:555-556.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution* 15:568-573.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24:1586-1591.
- Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, and R. Bennett. 2020. Ensembl 2020. *Nucleic acids research* 48:D682-D688.
- Yeaman, S., A. C. Gerstein, K. A. Hodgins, and M. C. Whitlock. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS genetics* 14:e1007717.
- Yeo, S., L. Coombe, R. L. Warren, J. Chu, and I. Birol. 2018. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* 34:725-731.
- Yocum, G. D., M. J. Toutges, R. L. Roehrdanz, and P. J. Dihle. 2011. Insertion of miniature subterminal inverted repeat-like elements in diapause-regulated genes in the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae).
- Zavaleta, E. 2000. The economic value of controlling an invasive shrub. *Ambio* 29:462-467.
- Zhang, J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Molecular Biology and Evolution* 21:1332-1339.

Appendix A - Supplementary to Contemporary and historical selection in Tasmanian devils (*Sarcophilus harrisii*) support novel, polygenic response to transmissible cancer

Rapture Sequencing for Contemporary Selection

We multiplexed 672-868 individuals per lane of Illumina NextSeq, obtaining 2.4 billion 150 base pair (bp) paired-end reads. We later re-sequenced 2,379 individuals from those Rapture libraries on an Illumina HiSeq 4000 to increase coverage and confidence in genotype inference. Finally, we also incorporated data from standard RAD sequencing libraries on four additional individuals from two of the six populations, West Pencil Pine and Fentonbury (see Epstein et al. 2016 for details.; Hendricks et al. 2017).

For each individual, we merged all available reads from across sequencing efforts. Then reads were de-multiplexed and low-quality reads were removed using `process_radtags` in Stacks using the ‘--bestrad’ option which checks for the single restriction enzyme cut site on either read; this step also removed any reads without recognizable barcodes or cut sites (Rochette et al. 2019a). The Stacks `clonefilter` program was used to remove potential PCR duplicates (Andrews et al. 2016). Using `bowtie2` (Langmead and Salzberg 2012b), reads were aligned to the *S. harrisii* reference genome `Devil_ref v7.0` (Murchison et al. 2012), downloaded from Ensembl in June 2014. Population allele frequencies can often be estimated with greater accuracy and reduced bias than individual genotypes (Warmuth and Ellegren 2019). To do this, we used ANGSD v0.910 (Korneliussen et al. 2014). For each set of individuals, we calculated genotype likelihoods and estimated allele frequencies within regions using the settings in Table A.2. Regions on the X chromosome were excluded.

Spatial and temporal analysis of contemporary selection

First, we calculated allele frequency change after DFTD infection. Within the five locations for which we had sampling both before and after DFTD appearance, we estimated the magnitude and direction of allele frequency change at SNPs with data from at least 10 individuals at both time points and a minor allele frequency (MAF) ≥ 0.05 and a likelihood-ratio test p-value (for presence of a SNP from ANGSD) $\leq 10^{-6}$ in at least one of the time points. For this analysis, individuals were assigned to “before” or “after” time points based on the date of DNA collection. Table A.1 presents the first year of DFTD detection in each population and samples collected from individuals after those years were considered “after.” DFTD could have been present at very low frequency prior to detection in some of these populations, but we believe using these dates of detection still provides a good estimate of pre-DFTD allele frequencies. We performed an arcsine (Fisher’s angular) transformation on the estimated

allele frequencies to reduce bias induced by the allele frequency spectrum (Fisher and Ford 1947). The SNPs were ranked by the magnitude of change, and the fractional rank was used as a pseudo- p-value for the composite statistic (described in the Main Text).

Then, we identified SNPs with estimates of strong selection using two time-series approaches which account for a population structure: the method of Mathieson & McVean (Mathieson and McVean 2013) (hereafter mm), which allows the estimated selection coefficient to vary over space; and spatpg, which allows the selection coefficient to vary over time as well as space (Gompert 2016). Individuals were divided into two-year cohorts based on their year of birth, starting with 1997 and ending with 2012 (Table A.1). Only SNPs with MAF ≥ 0.05 , minor allele count ≥ 3 , and p-value from ANGSD $\leq 10^{-6}$ in at least five population / cohort combinations were tested. For mm, we assumed the same effective population sizes as Epstein et al. (2016), and otherwise assumed similar effective population sizes where previous estimates did not exist (Table A.3). For spatpg, variance effective population sizes were estimated within the program with a bounded prior between 25 and 40. We created input allele count files for both spatpg and mm by multiplying allele frequencies estimated in ANGSD and the number of individuals and rounded to the nearest whole number. Due to computational limits, the dataset was randomly divided into 18 separate spatpg runs. Following spatpg recommendations in the manual (Gompert 2016), we calculated a support value for each SNP by taking the proportion of the posterior distribution (i.e. proportion of MCMC steps) for which the regression coefficient β , was non-zero ($0 < \beta$ or $0 > \beta$, whichever was smaller), where β describes the association between allele change and presence/absence of DFTD within a population. We multiplied these support values by two and treated them as pseudo- p-values when calculating a genomic inflation factor and the composite statistic.

Following the recommendations in Francois and colleagues (Francois et al. 2016), we adjusted the p-values of each test to reduce false positives. If the distribution of p-values was not uniform, we divided the Z-scores by the inflation factor. The inflation factor was calculated as the ratio between the median Z-scores and the expected median Z-scores for a χ^2 distribution with one degree of freedom. The genomic inflation factor varied from 0.25 – 0.58 for mm analyses of each population. For spatpg, we found an inflation factor of 0.44.

Using the adjusted p-values and pseudo p-values from the individual analyses, we calculated the DCMS statistic (Ma et al. 2015). This statistic combines the p-values from different tests at each SNP while accounting for genome-wide correlation among statistics. For each SNP, we used a weight based on only the statistics with results at that SNP, and we only included SNPs with results from at

least eleven of the twelve individual tests (Δa_f for five populations, mm for all six populations, and $spatpg$).

We then divided DCMS by the number of defined tests to get a mean composite score and ranked SNPs by this mean score. Because DCMS is not defined when one of the p-values is one or zero, we replaced p-values of one with 0.99999, and values of zero (only occurred for $spatpg$) with 0.00005 before performing the calculation. In accordance with previous linkage disequilibrium estimates (Epstein et al. 2016), we identified candidate genes within 100 kb of top SNPs, using *bedtools* v2.26.0 (Quinlan 2014); and supplemented annotations of novel genes with the Ensembl Compara gene family pipeline (Yates et al. 2020).

Historical Selection

To test for historical selection we used the branch-site test of PAML (Phylogenetic Analysis by Maximum Likelihood; (Yang 1997, 2007), implemented in the Bio.Phylo toolkit (Talevich et al. 2012) of BioPython (Cock et al. 2009). The branch-site test estimates the ratio of nonsynonymous-synonymous mutation rates (dN/dS) among aligned codons using a phylogenetic tree to allow for the appropriate evolutionary model to be employed. In the neutral model, all site classes and branches are constrained to $dN/dS \leq 1$. In the alternative model, dN/dS of Site Class 2 is allowed to exceed 1 for only the foreground branch, while constraining the background branches to $dN/dS \leq 1$.

The devil, wallaby, and opossum orthologous genes and respective sequences were mined from the Ensembl database with BioMart (Durinck et al. 2005; Durinck et al. 2009). For the koala, we used the orthologs identified with *blastx* v2.2.27 (Camacho et al.) supplied by Johnson and colleagues (Johnson et al. 2018). If splicing variants were available, only the first (most common) variant was retained for downstream data preparation. Only 1-to-1 orthologs were retained (i.e., paralogs were excluded). Orthologous gene tables were then reduced to genes with at least three of four possible sequences present and the respective species were pruned from the greater phylogeny. For the koala, we used an open reading frame finder, *getorf* from EMBOSS, (Rice et al. 2000) to generate amino acid sequences. The peptide sequence alignments were generated with MUSCLE v3.8.31 (Edgar 2004a; Edgar 2004b), then used to guide alignments of nucleotides with *tranalign*.

Functional Enrichment of Genes Under Selection

For contemporary candidate SNPs, we used the SNP2GO package (Szkiba et al. 2014) in the R environment v3.4.3. We filtered the most recent Gene Transfer File (*Sarcophilus_harrisii.DEVIL7.0.100*) and Ensembl gene ID GO term associations downloaded from Ensembl May 7, 2020 to include only genes which were within 100 kb of targeted loci and account

for the biased subset of targeted genes. We limited enrichment analysis to GO terms with a minimum of one association in the reference set and allowed an extension window of 100 kb from a candidate SNP. For GO term enrichment analysis of historical selection, we used the PANTHER web-interface with HUGO gene names (Braschi et al. 2019) and the reference set defined as all genes that were tested in PAML (Mi et al. 2018).

We only compared the distributions among genes for which $dN/dS > 1$ and were statistically significant according to the likelihood ratio test. To account for the bias induced by targeted sequencing among contemporary candidates, we defined the reference set as all genes that could have been detected in both tests for a given overlapping or enrichment analysis

We capitalized on the wealth of ongoing research in devils and DFTD by comparing our contemporary and historical candidates to those previously identified using different datasets and analytical approaches (Epstein et al. 2016; Wright et al. 2017; Hubert et al. 2018; Margres et al. 2018a; Fraik et al. 2020) using Fisher's Exact Test implemented in the R package GeneOverlap (Shen and Sinai 2020). We then tested for overrepresentation of contemporary and historical candidates in gene sets of the molecular signatures database (MsigDB) (Subramanian et al. 2005). MSigDB contains several libraries of gene sets which allowed us to gain further insight to pathways that may be under selection in devils. We compared our contemporary and historical candidate gene lists to gene sets from the MsigDB Hallmark, Curated, Computational, Oncogenic Signatures, and Immunologic Signatures (Liberzon et al. 2015; Godec et al. 2016). We built 2x2 contingency tables for each set of genes under positive selection and in each of the tested gene sets of MsigDB. Despite limitations, this overrepresentation method is straight-forward and flexible for non-model organisms and targeted sequencing. To identify intersecting gene sets, we first converted all Ensembl gene IDs of interest to HUGO annotations with Biomart. Second, we created appropriate background sets (of length N) by intersecting each MsigDB gene set with the respective list of all genes that were tested for selection. Lastly, from these contingency tables we computed overlaps, using the hypergeometric distribution ($\text{dhyper}(c(0:x), m, n, x, \text{log} = \text{FALSE})$), where $c(0:x)$ is a vector of quantiles representing the number of genes both under selection and found in an MsigDB gene set; m is the number of genes in the candidate gene list, n is the number of genes in the candidate gene list but not in the MSigDB gene set, and x is the number of candidate genes in the MsigDB gene set list. After accounting for multiple testing with the Benjamini-Hochberg correction (Shaffer 1995), we considered the overrepresentation result statistically significant with adjusted p-values < 0.05 and background gene sets greater than or equal to ten genes. Finally, we performed a permutation test to establish a null expectation for the rate of shared gene sets between contemporary and historical selection and

compared the resultant empirical null to our observed proportion of shared gene sets. To do this, we randomly selected the same number of candidate genes with HUGO annotations (112) from the list of all HUGO annotated contemporary candidates (3,920) 1,000 times, with replacement, and performed gene set overlap analysis as above with only the known gene sets significantly overlapping with the historical candidates.

Table A.1 Number and year of sampling across six localities.

Population	Year	Number of
wukalina/Mt. William	2004	11
wukalina/Mt. William	2005	28
wukalina/Mt. William	2006	20
wukalina/Mt. William	2007	21
wukalina/Mt. William	2008	17
wukalina/Mt. William	2009	14
wukalina/Mt. William	2010	8
wukalina/Mt. William	2011	15
wukalina/Mt. William	2012	4
wukalina/Mt. William	2013	10
wukalina/Mt. William	2014	8
Freycinet	1999	107
Freycinet	2000	122
Freycinet	2001	71
Freycinet	2002	65
Freycinet	2003	38
Freycinet	2004	60
Freycinet	2005	54
Freycinet	2006	27
Freycinet	2007	32
Freycinet	2008	21
Freycinet	2009	7
Freycinet	2010	10
Freycinet	2011	10
Freycinet	2012	18
Freycinet	2013	12
Freycinet	2014	26
Forestier	2004	131
Forestier	2005	46
Forestier	2006	168
Forestier	2007	98
Forestier	2008	93
Forestier	2009	107

Forestier	2010	13
Forestier	2012	26
Forestier	2013	1
Fentonbury	2004	47
Fentonbury	2005	52
Fentonbury	2006	58
Fentonbury	2007	63
Fentonbury	2008	37
Fentonbury	2009	11
West Pencil Pine	2006	52
West Pencil Pine	2007	71
West Pencil Pine	2008	33
West Pencil Pine	2009	57
West Pencil Pine	2010	33
West Pencil Pine	2011	84
West Pencil Pine	2012	62
West Pencil Pine	2013	43
West Pencil Pine	2014	13
Narawntapu	1999	33
Narawntapu	2003	9
Narawntapu	2004	46
Narawntapu	2005	27
Narawntapu	2006	64
Narawntapu	2007	45
Narawntapu	2008	63
Narawntapu	2009	30
Narawntapu	2010	27
Narawntapu	2011	15
Narawntapu	2012	22

Table A.2 ANGSD genotype calling settings.

Option	Setting	Description
-	40	Minimum read mapping quality
-minQ	25	Minimum base Phred score
-baq	2	Perform extended base quality adjustment around indels
-GL	2	Use the (old) GATK genotype likelihood model
-doMaf	2	Estimate allele frequencies assuming one known allele
-	4	Use the reference allele as the known allele

Table A.3 Estimates of effective population size.

Population	N_e
Fentonbury	35
Forestier	35
Freycinet	34
Narawntapu	37
West Pencil Pine	26

Table A.4 Number of shared genes within 10000 bp of the top 1% of SNPs for the DCMS list of contemporary candidates and each intermediate test. The total number of genes in each list is found in the diagonal. Populations are abbreviated as follows: FEN = Fentonbury, FOR = Forestier, FREY = Freycinet, wuk = wukalina/Mt. William, NAR= Narwantapu, and WPP = West Pencil Pine.

	DCMS	FEN	FEN	FOR	FOR	FREY	FREY	wuk	NAR	NAR	spatpg	WPP	WPP
DCMS	247	14	60	34	67	28	32	41	17	55	39	14	26
FEN	-	165	20	24	31	15	8	19	26	21	9	4	16
FEN	-	-	279	33	22	23	22	8	42	38	19	18	20
FOR	-	-	-	301	30	20	28	28	25	48	17	34	22
FOR	-	-	-	-	283	14	14	38	28	26	21	28	28
FREY	-	-	-	-	-	299	19	8	29	12	31	24	42
FREY	-	-	-	-	-	-	289	16	16	53	13	16	26
wuk	-	-	-	-	-	-	-	295	21	48	39	8	22
NAR	-	-	-	-	-	-	-	-	282	24	21	46	42
NAR	-	-	-	-	-	-	-	-	-	314	36	21	30
spatpg	-	-	-	-	-	-	-	-	-	-	284	15	21
WPP	-	-	-	-	-	-	-	-	-	-	-	329	29
WPP	-	-	-	-	-	-	-	-	-	-	-	-	232

Table A.5 Annotated Tasmanian devil gene IDs of within 1000 bp of the top 1% of composite SNPs; i.e., candidates for contemporary selection, provided at:
https://github.com/Astahlke/contemporary_historical_sel_devils/blob/master/contemporary/angsd_2019-01-18/next/composite_stat/2019-2-22/results/annotation_top1.0/composite.snps.everything.top.genes.100000bp.txt

Figure A.1 Mean coverage of individuals across populations at targeted loci.

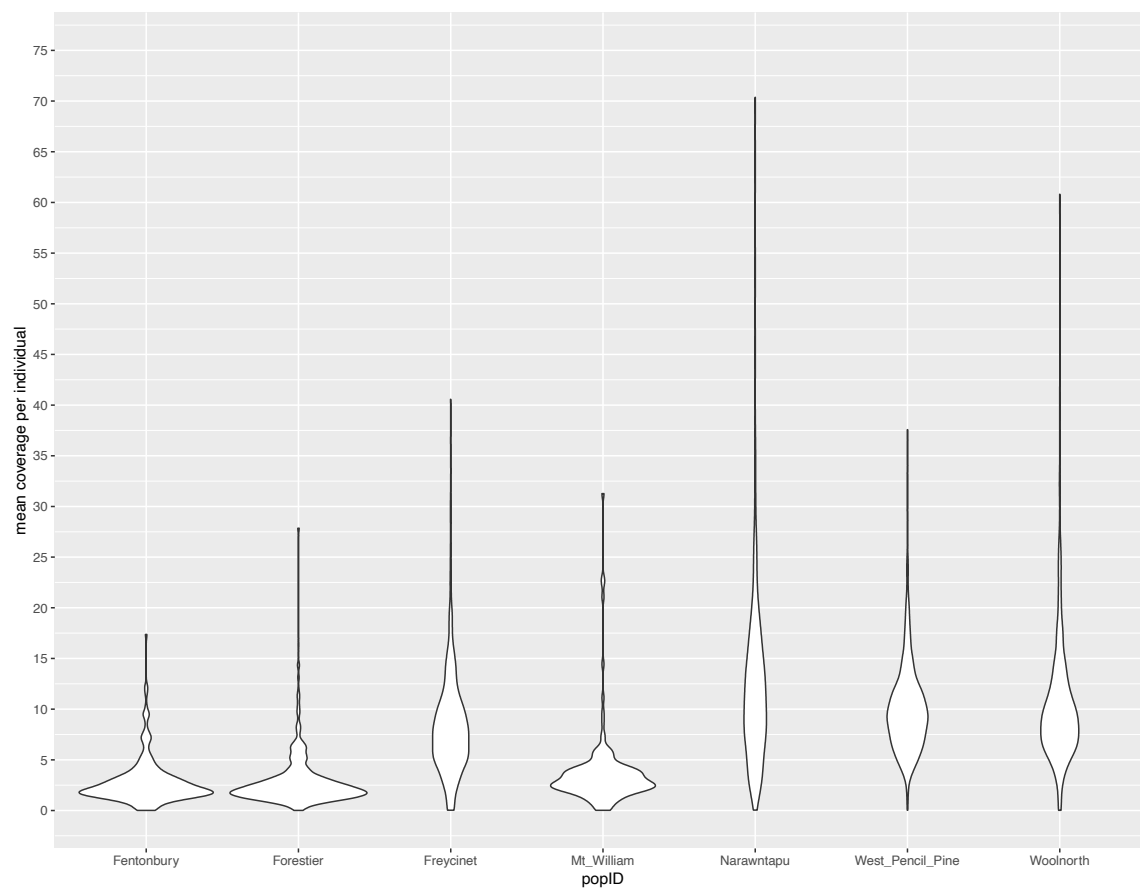


Figure A.2 The folded allele frequency spectra for each population before DFTD became prevalent. wuaklina/Mt. William is not presented because it was first sampled in 2004, eight years after DFTD was first described at that locality.

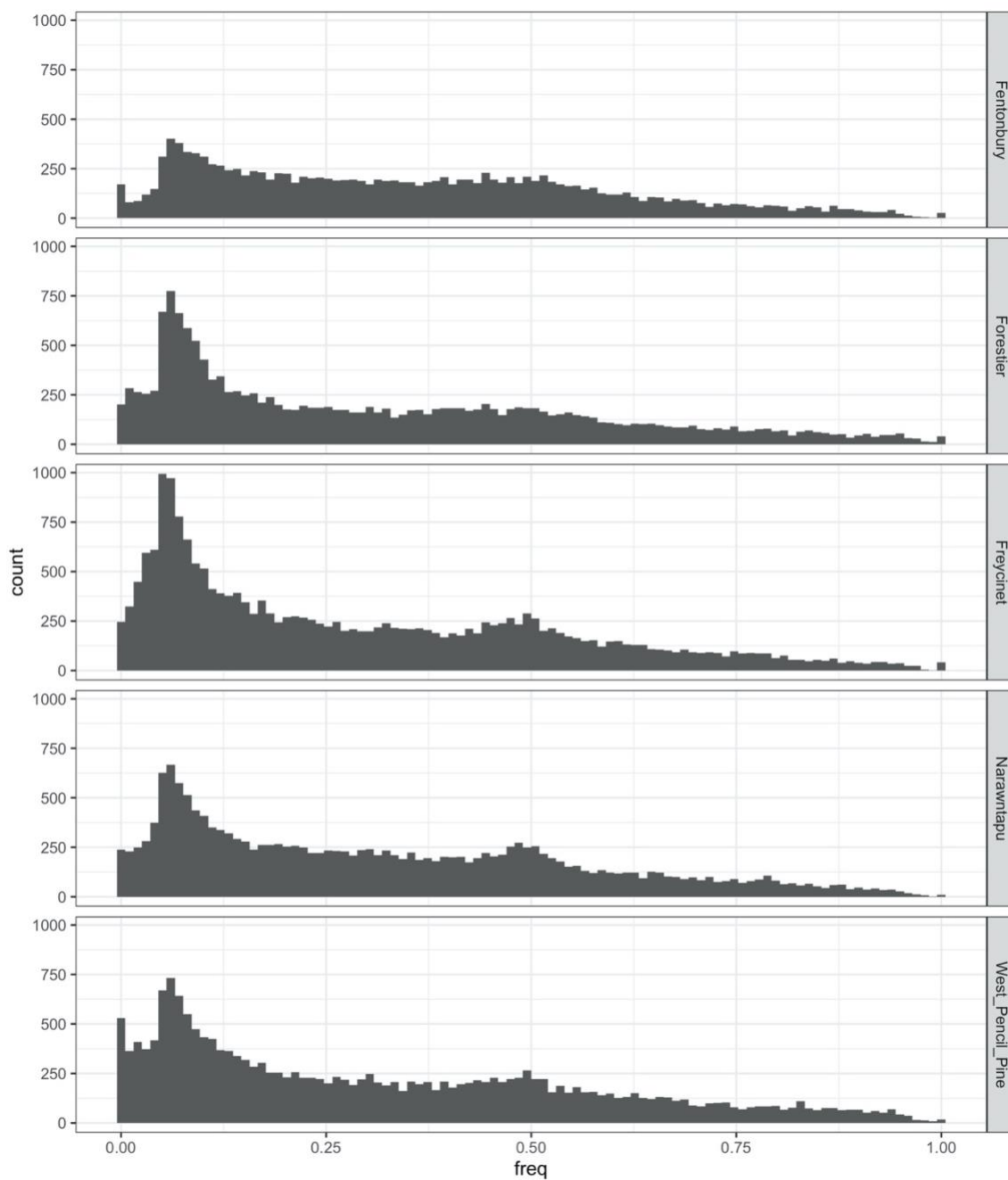


Figure A.3 Un-adjusted p-values for all SNPs of each population analyzed with mm (Mathieson and McVean 2013).

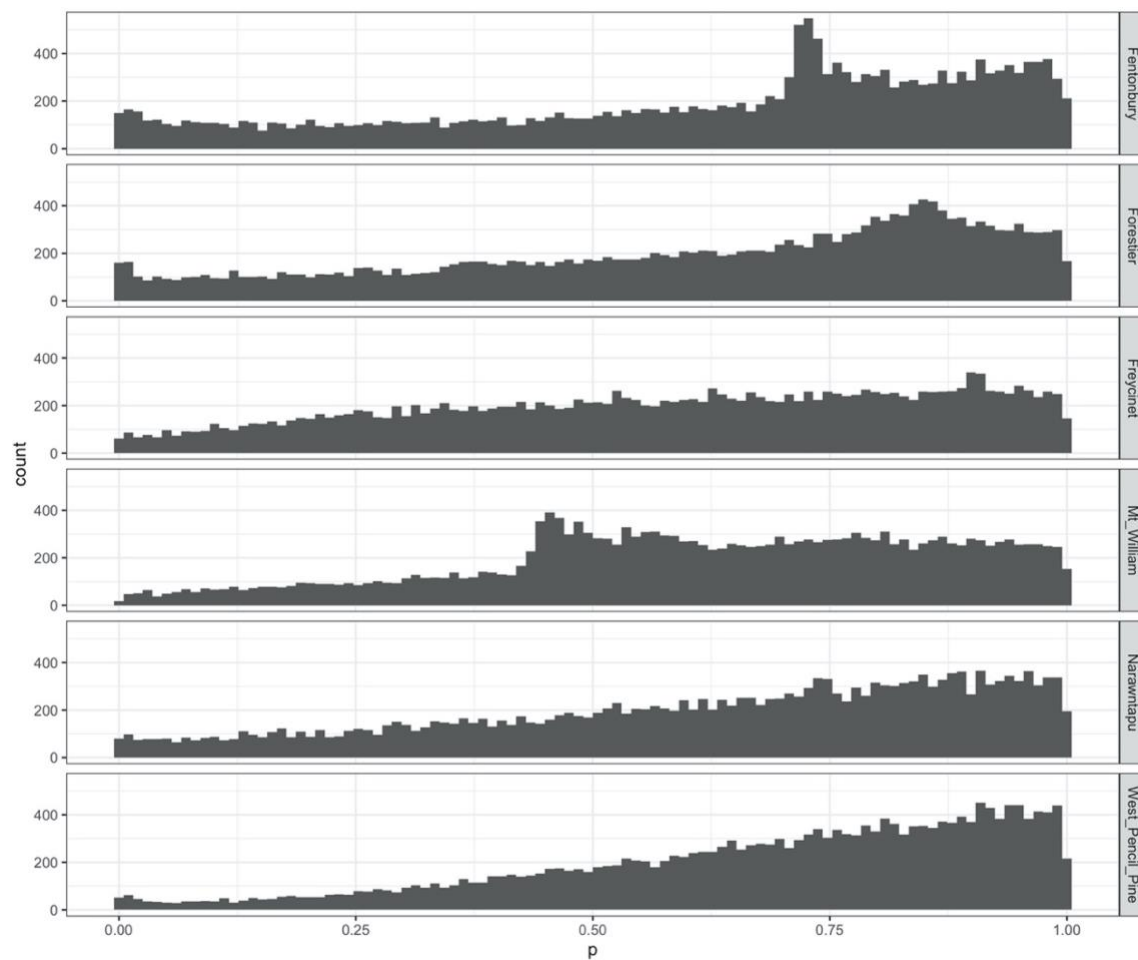


Figure A.4 Adjusted p-values for all SNPs of each population analysed with mm (Mathieson and McVean 2013).

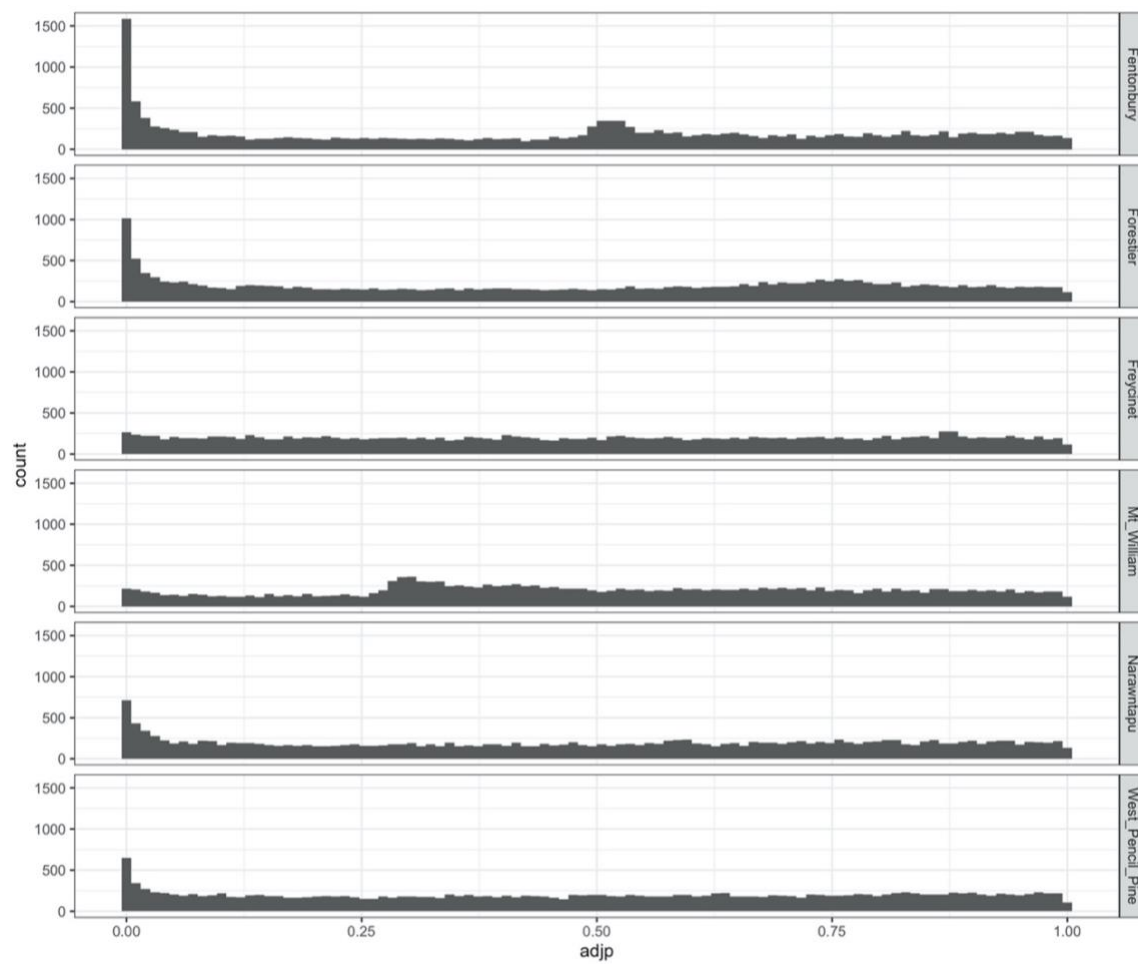


Figure A.5 Un-adjusted p-values for all SNPs across all populations analysed with spatpg (Gompert 2016).

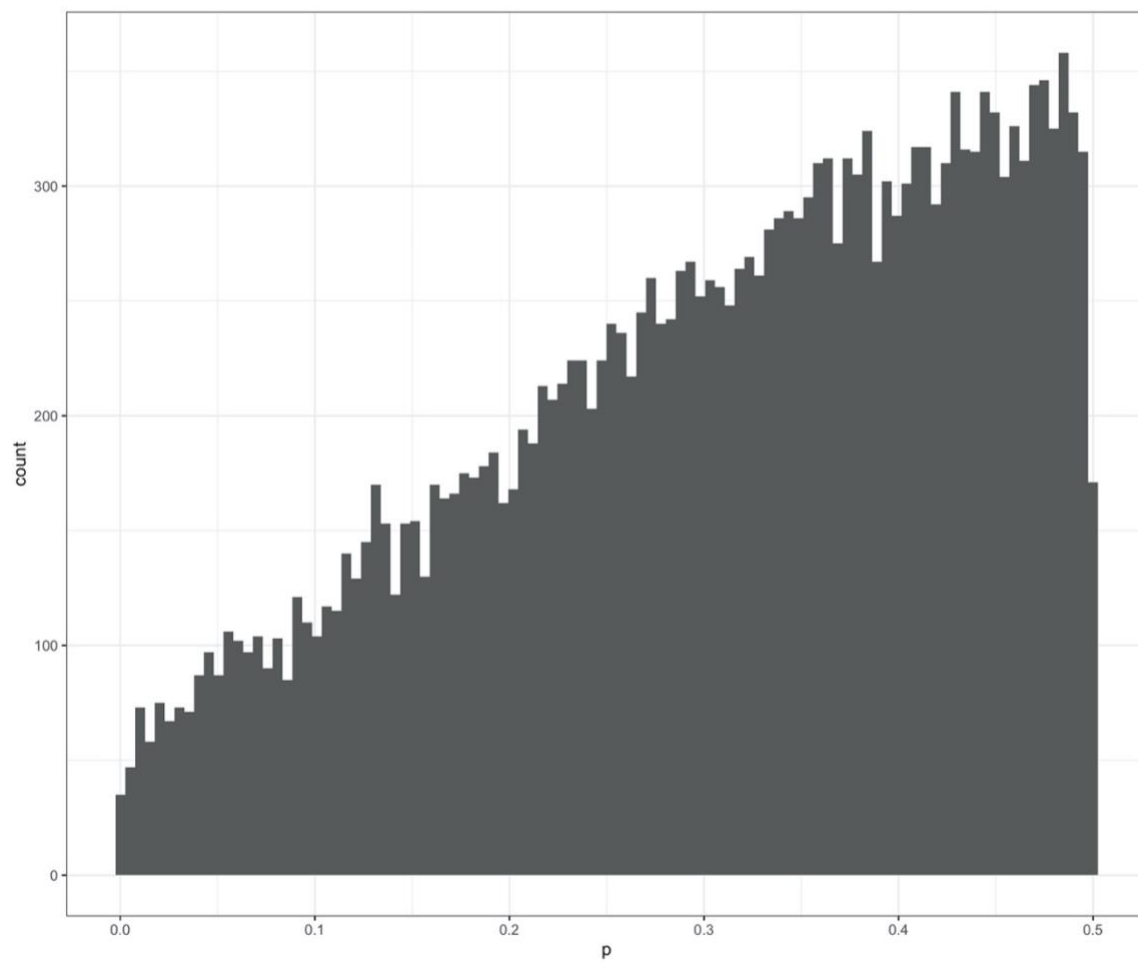


Figure A.6 Adjusted p-values (Francois et al. 2016) for all SNPs across all populations analysed with spatpg (Mathieson and McVean 2013; Gompert 2016).

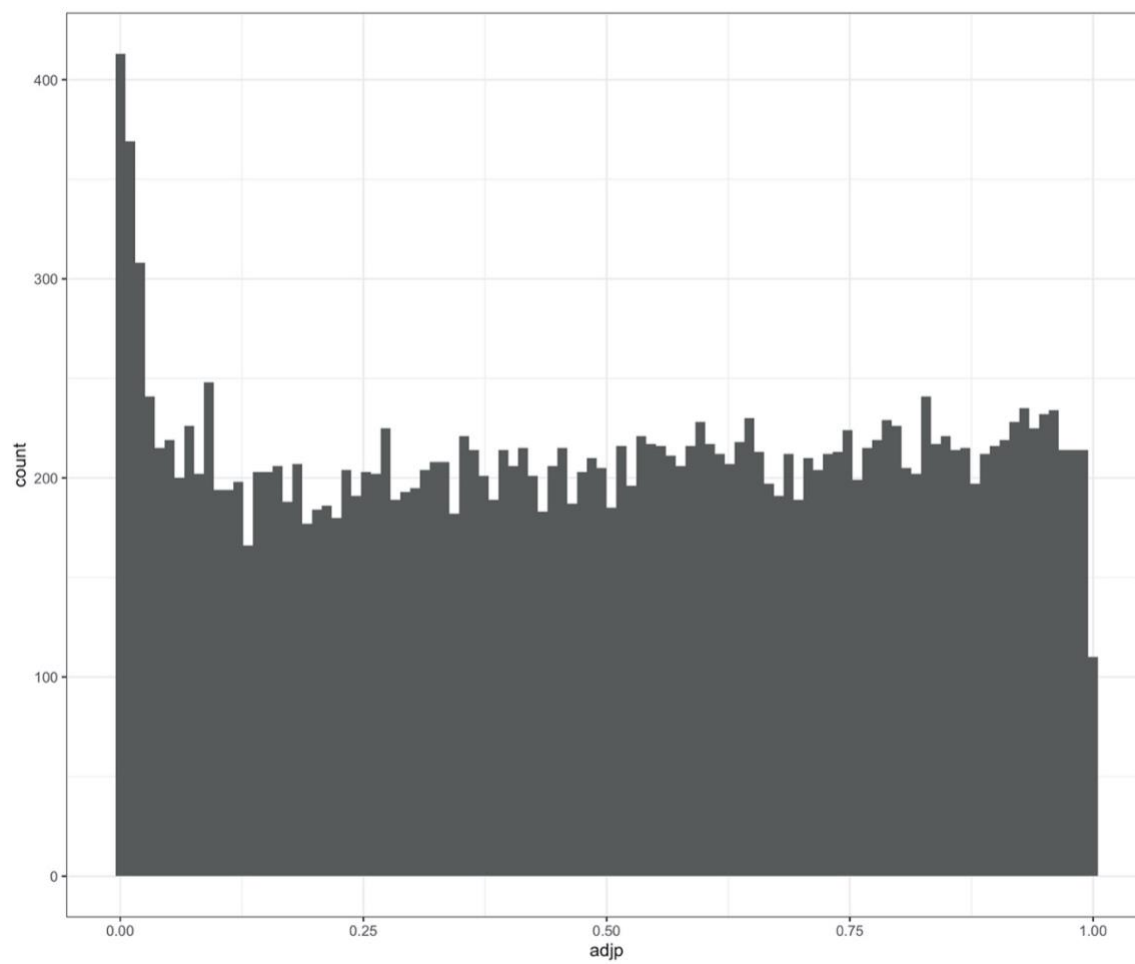


Figure A.7 Allele frequency change (Δaf) for each population separately. SNPs in the top 1% are indicated by more opaque points. The threshold line for the top 1% within each population is indicated by a dashed line.

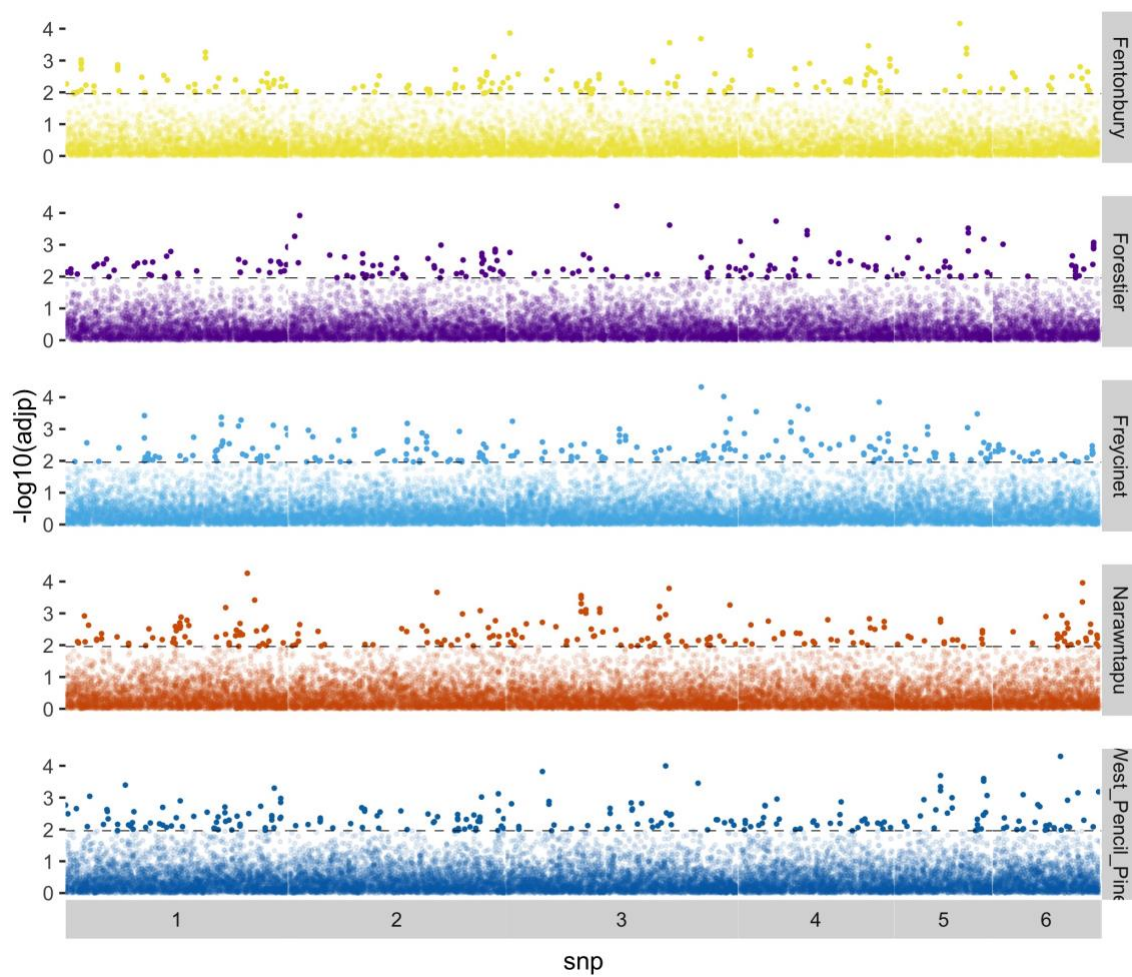


Figure A.8 Signatures of selection as detected by mm for each population separately. SNPs in the top 1% are indicated by more opaque points. The threshold line for the top 1% within each population is indicated by a dashed line.

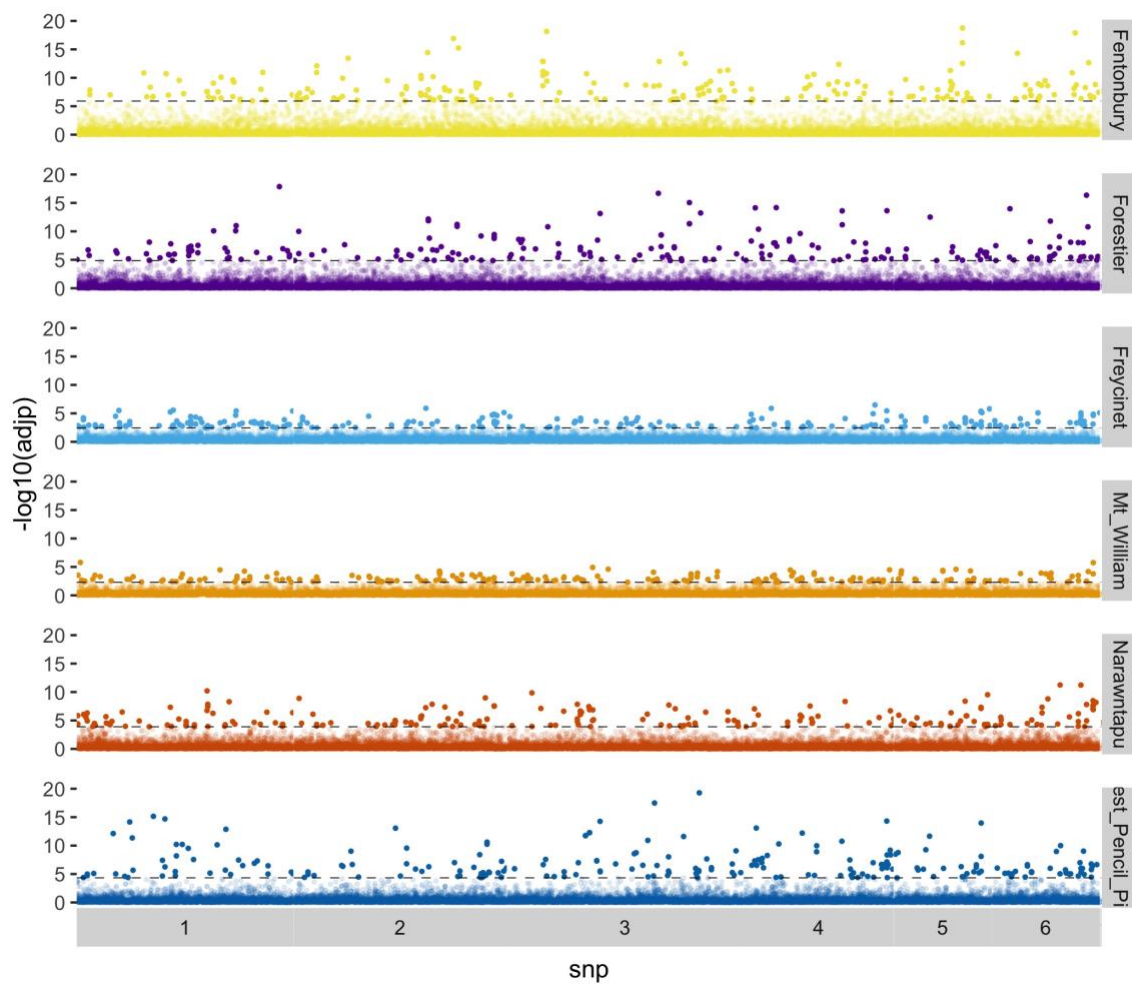


Figure A.9 Binned initial allele frequency distributions for DCMS candidates and non-candidates across all populations with samples before DFTD became prevalent.

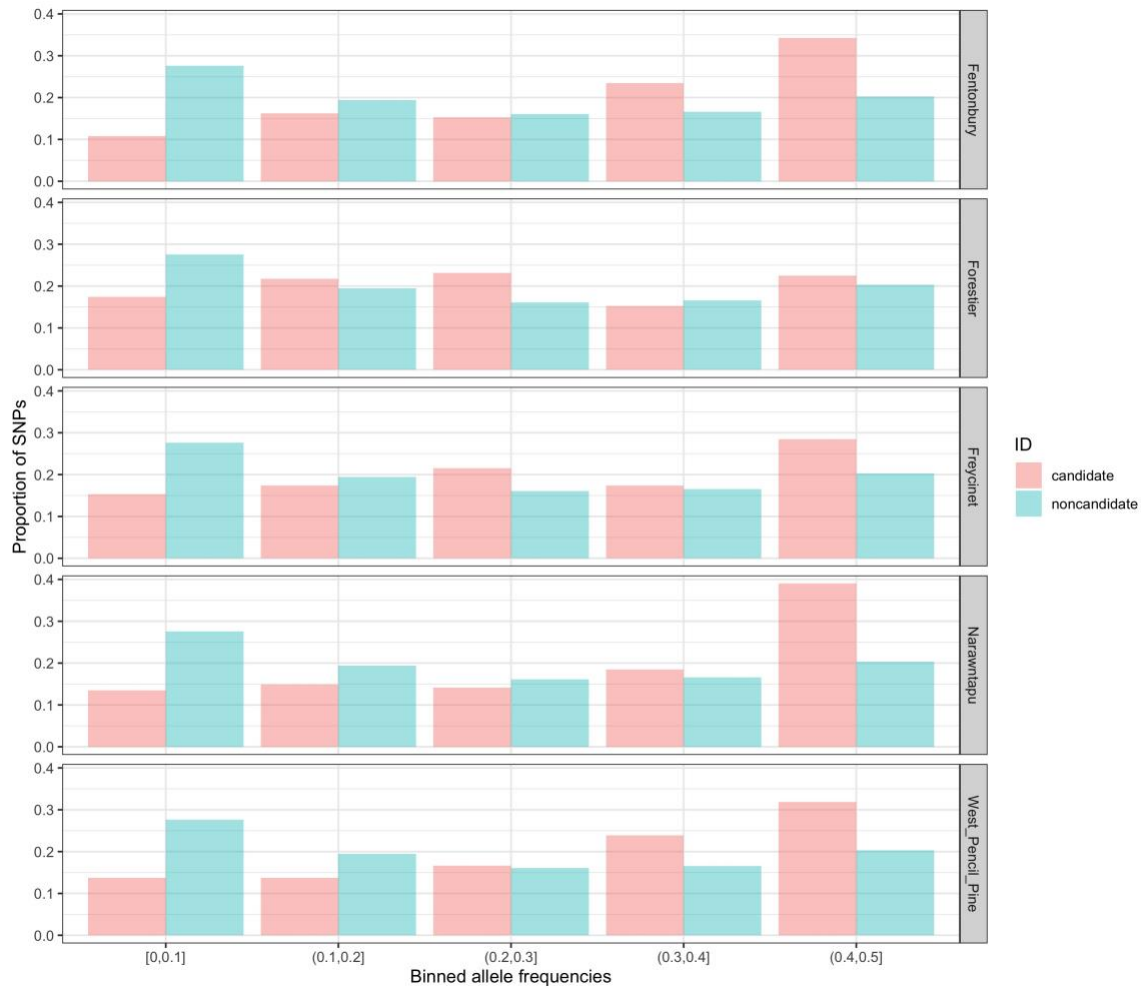
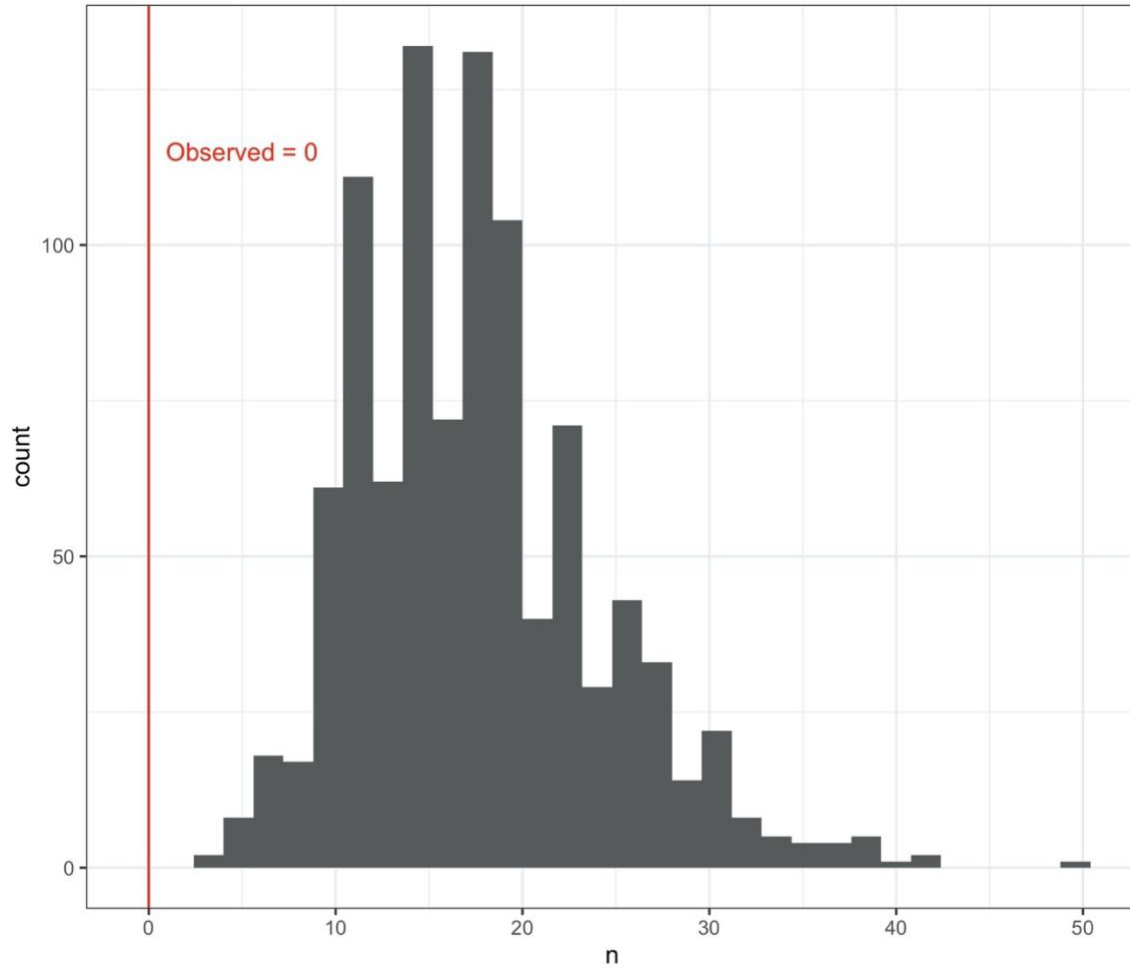


Figure A.11 Histogram of shared MSigDB (Subramanian et al. 2005) gene set overlaps between contemporary and historical candidates, i.e. the signature of ongoing selection, in a permutation test of 1000 draws with replacement. We observed no shared gene set overlaps between contemporary and historical candidates. Of the 1000 permutations, 100% had more overlapping sets than observed.



**Appendix B - Supplementary to Hybridization and range expansion in
tamarisk beetles (*Diorhabda* spp.) introduced to North America for
classical biological control**

Table B.1 Sampling locality identification codes (Locality ID), site names, latitude and longitude coordinates, dates and sample sizes for each collection effort. Locality IDs include the country or state sampled and symbols for relevant groups as follows: * = native range, Δ = source population, ∇ =release sites, \bullet = hybrid zone, \blacksquare = *D. carinulata* expansion front.

Locality ID	Site Name	Latitude	Longitude	Date	Sample
<i>D. carinulata</i> sites					
46CH* Δ	Bitun, China	47.30	87.75	7/3/16	24
1WY ∇	Lovell	44.86	-108.18	9/3/14	17
34UT ∇	Delta	39.23	-112.93	10/15/14	16
2CO ∇	Fountain Creek	38.34	-104.61	9/2/14	16
4CO \blacksquare	Adobe Reservoir	38.26	-103.25	9/4/14	16
5CO \blacksquare	SE corner	37.70	-103.42	8/21/14	19
11CO \blacksquare	Wilkinson	37.34	-104.16	9/18/14	11
32UT \blacksquare	St. George	37.07	-113.58	10/15/14	10
31NV \blacksquare	Virgin River,	36.69	-114.26	10/15/14	19
33NV \blacksquare	Lake Mead,	36.38	-114.40	10/15/14	12
12AZ \blacksquare	Big Bend	35.12	-114.64	9/18/14	20
Laboratory cultures					
CARINA_LAB	<i>D. carinata</i> Lab			1/7/18	20
SUB_LAB	<i>D. sublineata</i> Lab			1/7/18	7
Native range <i>D. elongata</i>					
43GR* Δ	Posidi, Greece	39.97	23.37	7/6/15	9
44GR*	Delta Aksiou, Greece	40.55	22.74	7/6/15	12
41CR*	Plakias, Crete, Greece	35.19	24.40	7/4/15	11
37CR*	Rethimno, Crete,	35.37	24.47	7/2/15	6
39CR*	Panaramnos, Crete,	35.42	24.68	7/4/15	16
38CR* Δ	Sfkaki, Crete, GR	35.42	24.69	7/4/15	17
Suspected Hybrid Zone					
6KS \bullet	W Finney County	37.99	-101.08	8/8/14	15

8OK•	Guymon	36.70	-101.55	9/16/14	18
28NM•	Tucumcari Lake	35.19	-103.69	10/6/14	20
26NM•	Lake Sumner	34.15	-104.48	9/29/14	16
27NM•	Roswell E	33.40	-104.41	9/29/14	15
15TX•	NE Post	33.32	-101.26	9/26/14	15
19TX•	Aspermont	33.17	-100.24	9/27/14	18
21NM• ^v	Artesia Wildlife	32.98	-104.44	10/7/14	17
16TX• ^v	Lake JB Thomas	32.61	-101.22	9/26/14	17
20NM•	Malaga	32.22	-104.08	9/27/14	21
18TX•	Orla	31.49	-103.48	9/27/14	8
52TX•	Tornillo	31.44	-106.09	8/3/17	14
50TX•	Presidio Hwy	29.34	-104.07	8/3/17	25
51TX•	CO Canyon Boat	29.34	-104.06	8/3/17	27
47TX•	Rio Grande Village	29.18	-102.96	8/3/17	14
49TX• ^v	Santa Elana	29.16	-103.60	8/3/17	17

B.2 Available data on releases of the different *Diorhabda* species by year and state. Note: This table does not reflect every release since not all data was accessible at time of publication. Provided at:
<https://github.com/Astahlke/DiorhabdaPopulationStructure/blob/master/info/Supp%20tab%203%20Diorhabda%20release%20table.xlsx>

B.3 Detailed Structure results for one repetition of $K = 4$ (presented in Figure 2.2), including sample ID, locality ID, inferred cluster assignments, and 90% credible intervals, available at https://github.com/Astahlke/DiorhabdaPopulationStructure/blob/master/structure_analysis/ancestry_confidenceintervals.csv.

Table B.4 Estimates of π (nucleotide diversity), F_{IS} (inbreeding), and number of private alleles for hybrid populations, pure populations, and native populations of *Diorhabda* (spp.) by sample size (N). Superscript letters indicate significant differences at $p < 0.05$.

	N	Mean (95% CI)	P
π			
Hybrid	7	0.0863 (0.063, 0.11) ^a	
Pure	21	0.0335 (0.015, 0.062) ^b	0.002
Native	7	0.0381 (0.014, 0.062) ^b	
F_{IS}			
Hybrid	7	0.157 (0.098, 0.217) ^a	
Pure	21	0.0454 (0.011, 0.080) ^b	0.004
Native	7	0.027 (-0.033, 0.0868) ^b	
Private alleles			
Hybrid	7	4.71 (-2.63, 12.1) ^a	
Pure	21	6.24 (2.00, 10.5) ^a	0.019
Native	7	17.86 (10.52, 25.2) ^b	

Figure B.1 Sampling sites and original source population locations from the native range of *D. elongata*.

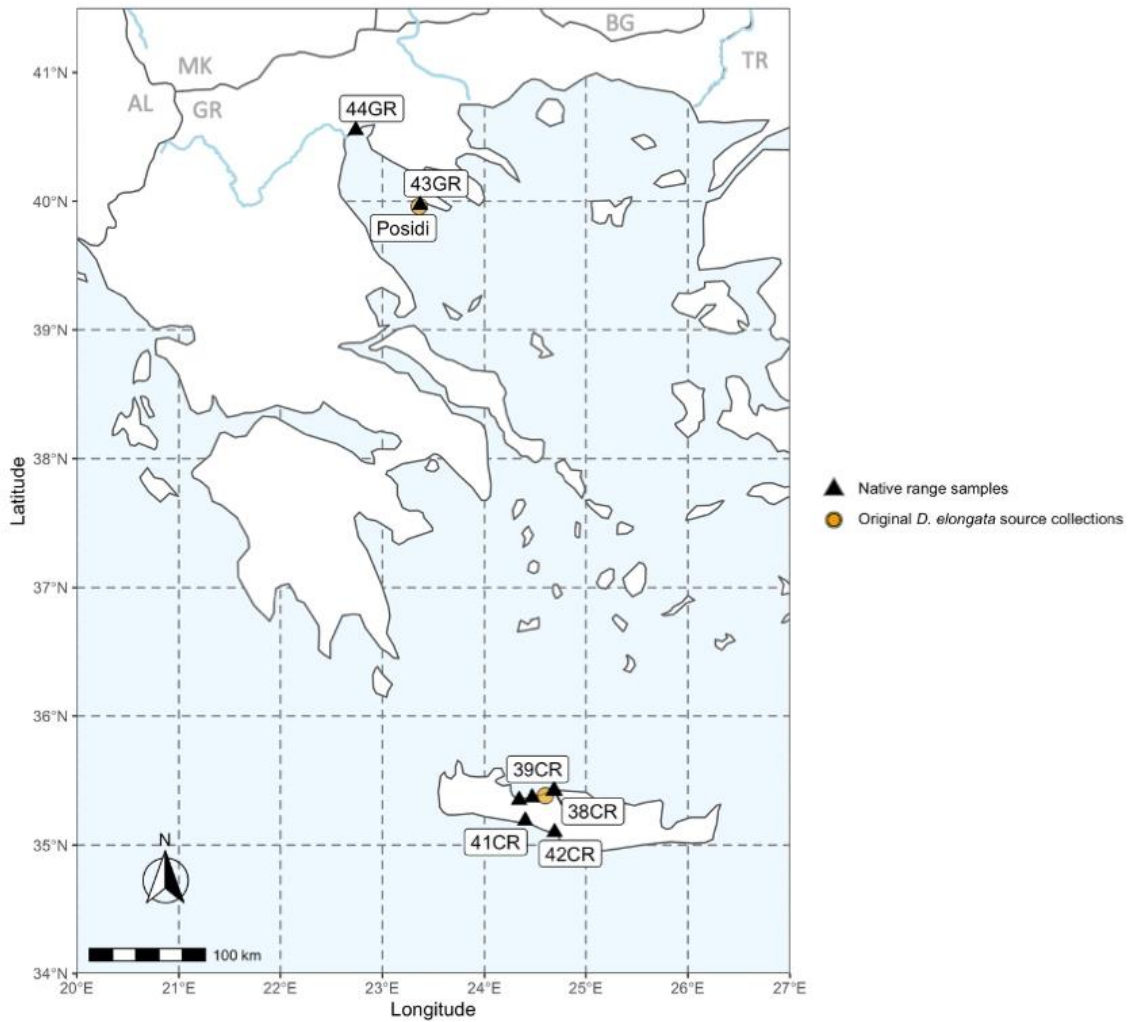


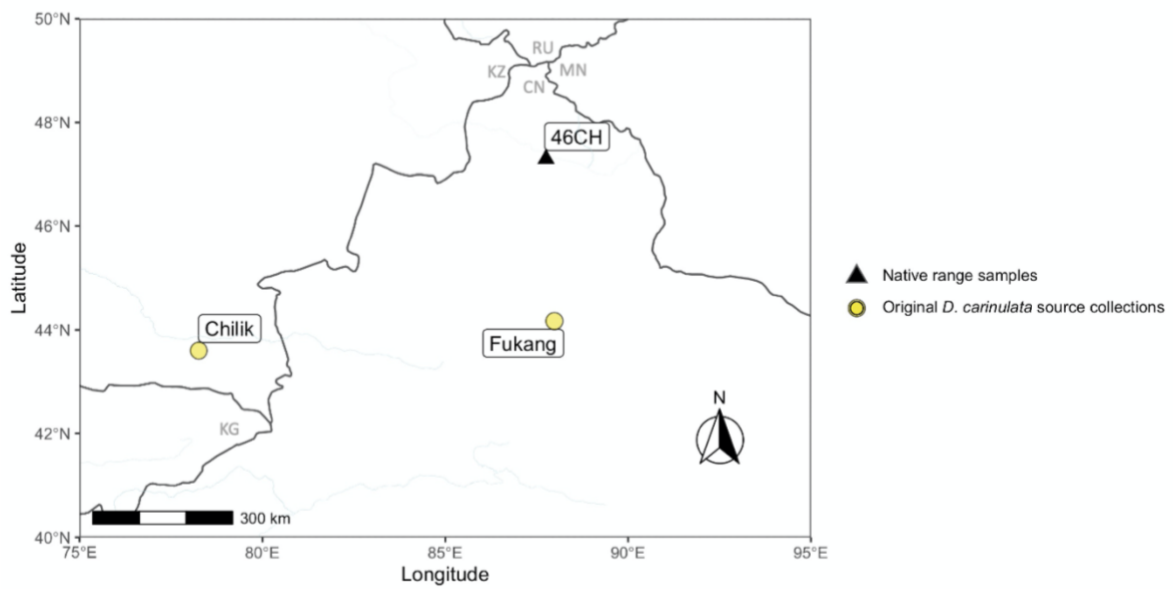
Figure B.2 Sampling sites and original source population locations from the native range of *D. carinulata*.

Figure B.3 Percent of paired reads that mapped concordantly to the *D. carinulata* genome per sample, according to group. The mean of 72.78% is plotted as a horizontal line.

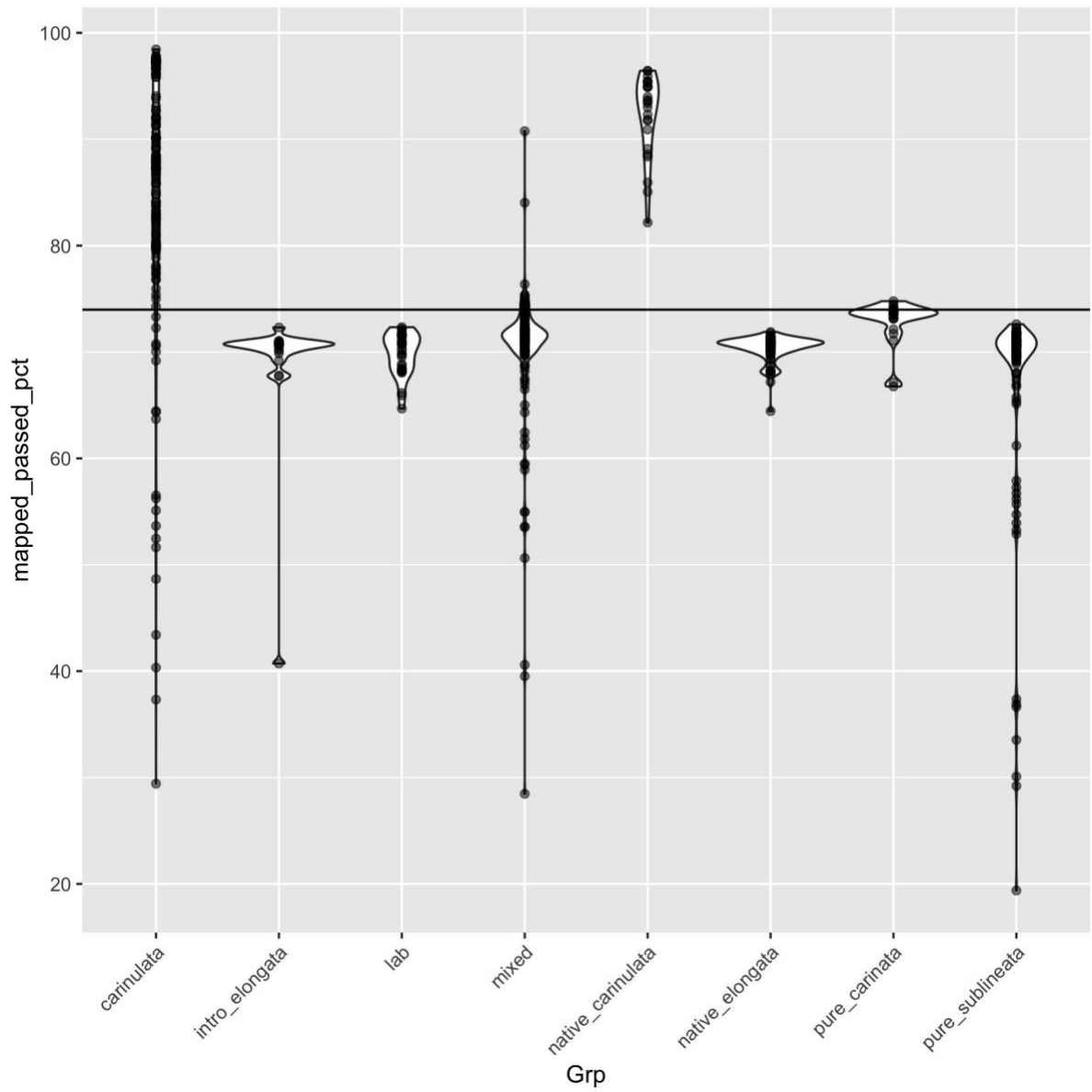


Figure B.4 Change in likelihood values from Structure results for the global SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.

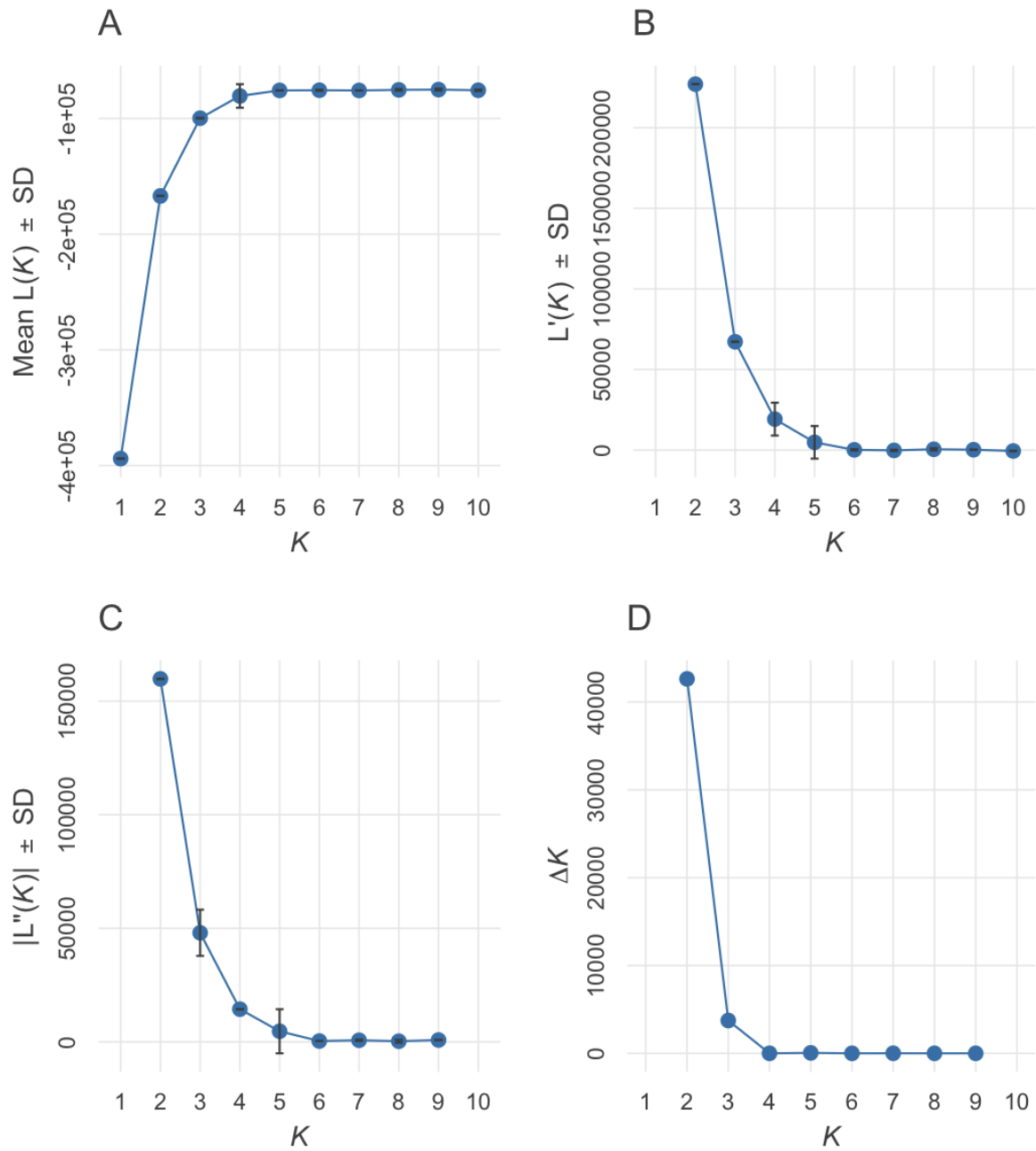


Figure B.5 Modal structure results for the global SNP dataset for K=2, K=3, and K=4, from top to bottom.

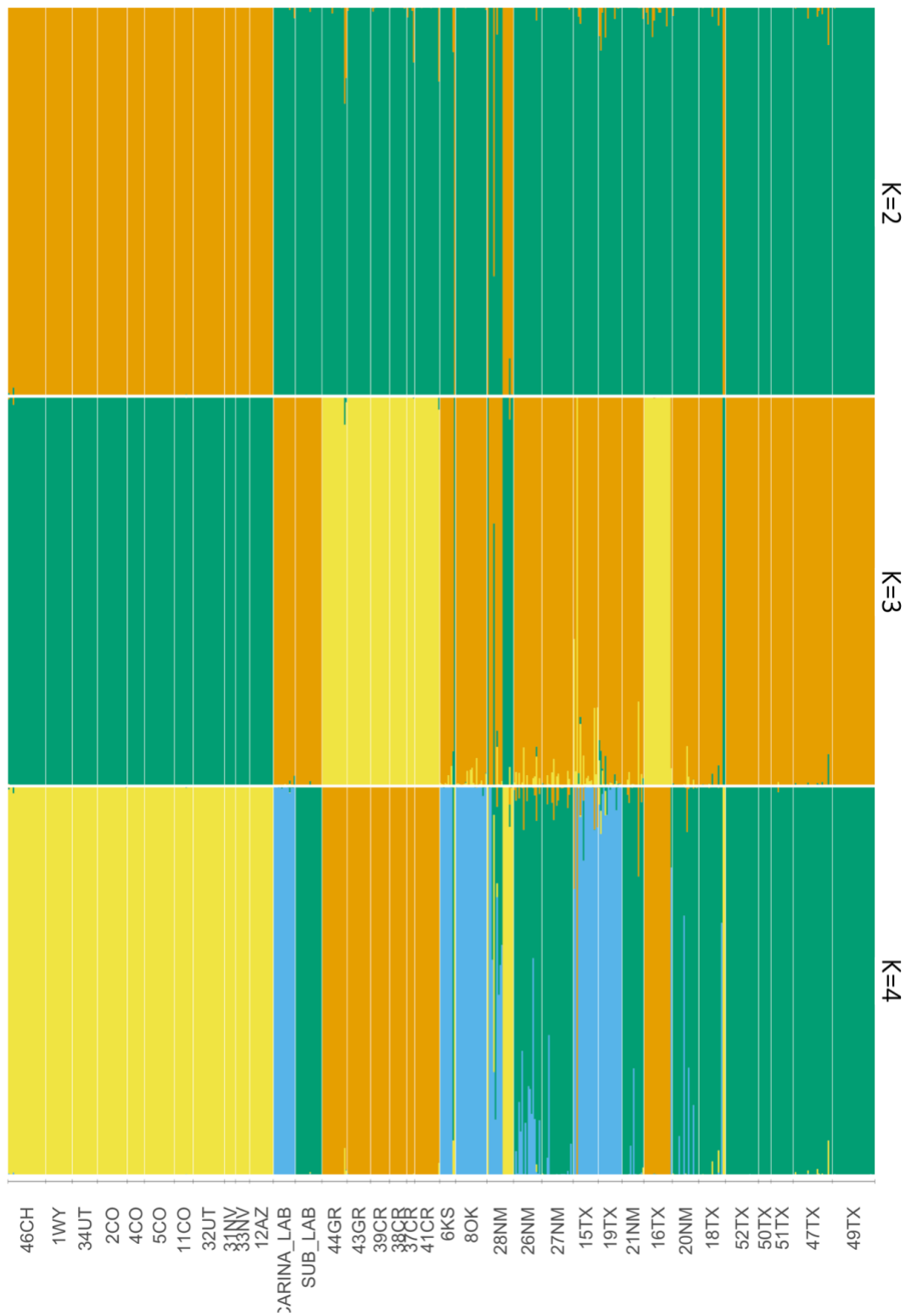


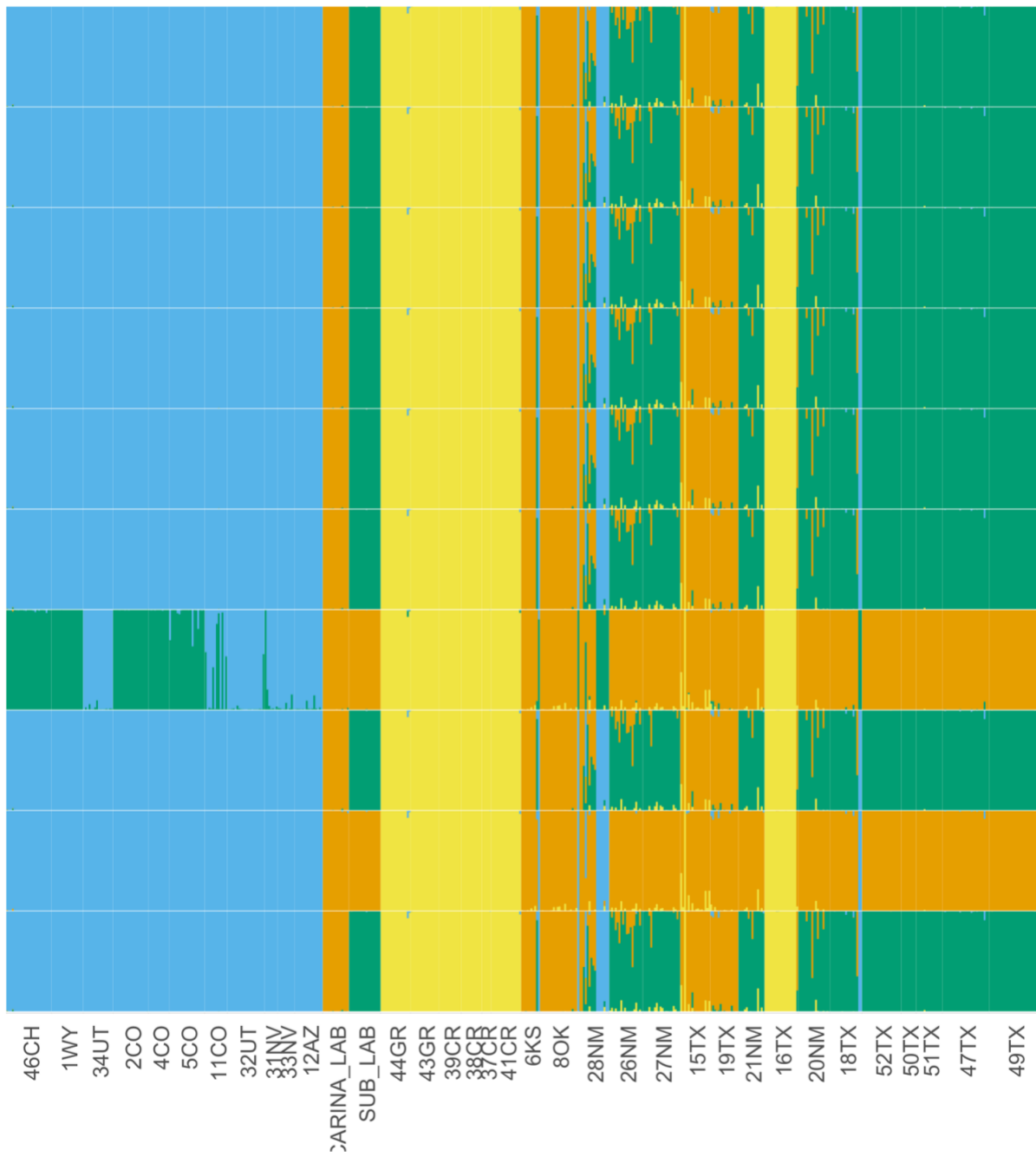
Figure B.6 Structure results for the global SNP dataset across all ten repetitions of $K=4$.

Figure B.7 Histograms of individual q-values (as a hybrid index) from modal $K=4$ Structure (Pritchard et al 2000) results for bi-parental hybrid individuals. Each histogram is symmetric about the 0.50 line and colors indicate the assigned ancestry for the respective cluster. Ordered leading with most abundant pair: A) *D. carinata* x *D. sublineata* hybrids ($n=24$) have both extreme (near 0 and 1) and intermediate values suggesting a lack of barriers. B) *D. elongata* x *D. carinata* hybrids ($n=3$) and C) *D. elongata* x *D. sublineata* hybrids ($n=3$) both have more extreme q-values with one likely back-crossed hybrid (0.25, 0.75).

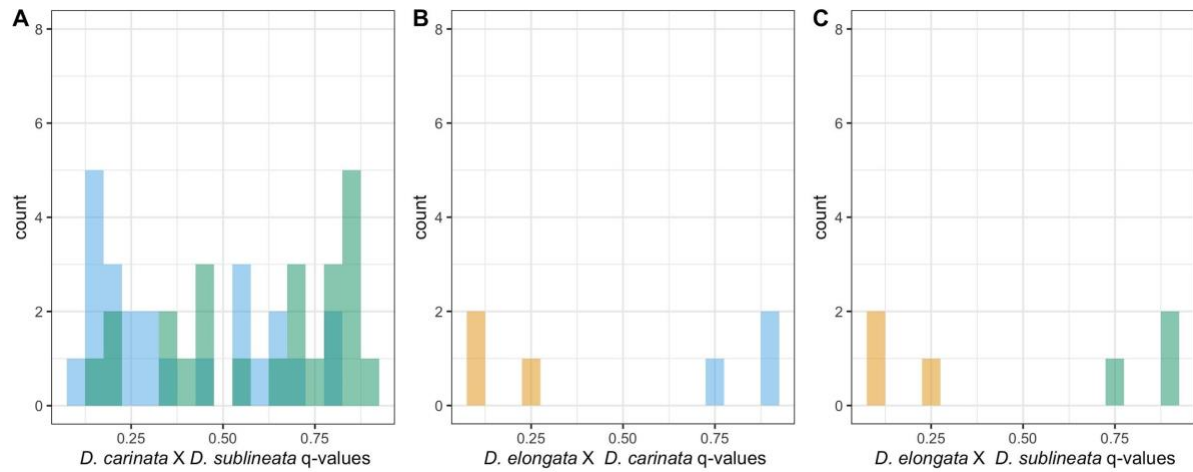


Figure B.8 Change in likelihood values from Structure results for the *D. carinulata* SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.

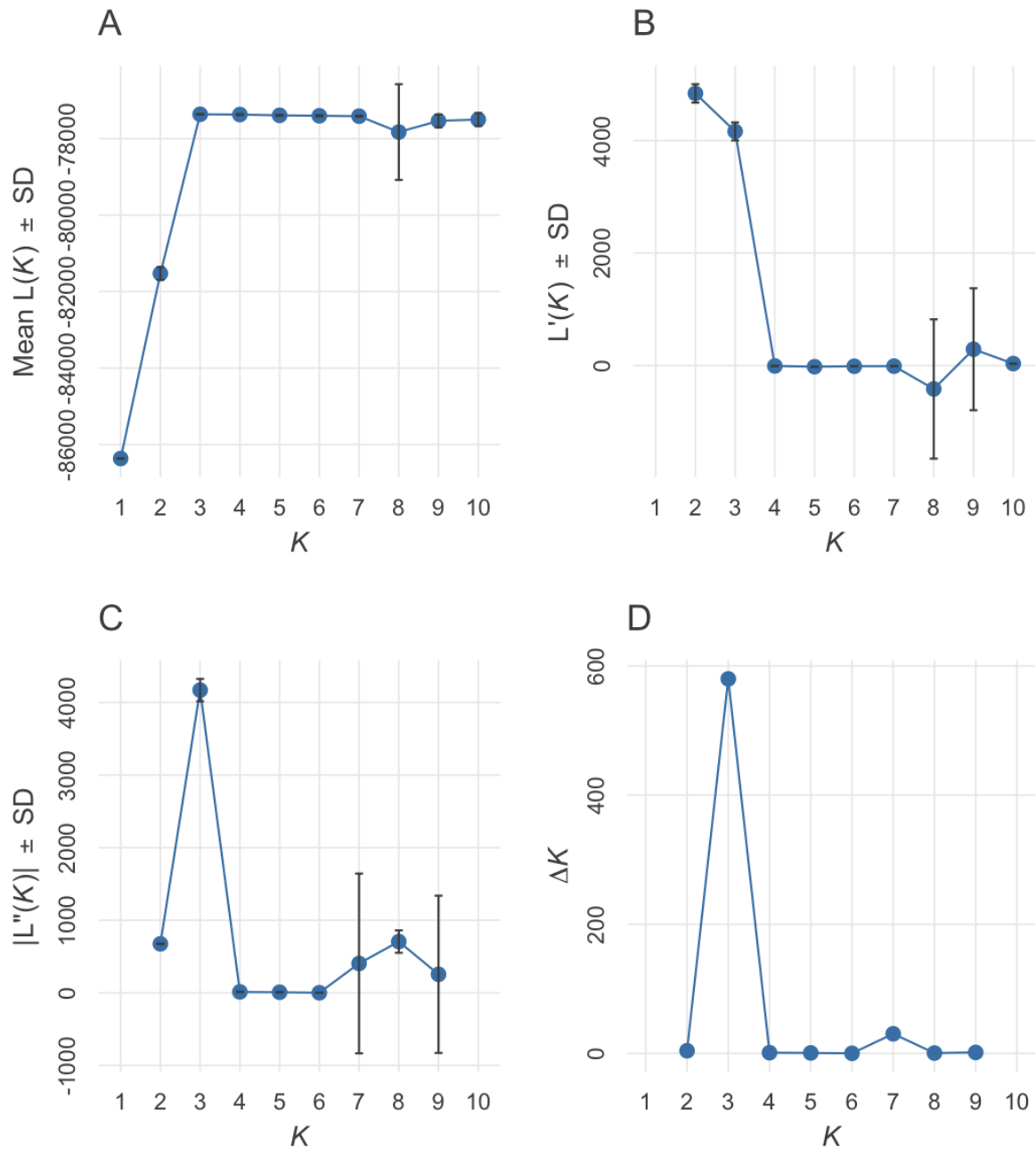


Figure B.9 Change in likelihood values from Structure results for the *D. elongata* SNP dataset, visualized according to four methods (A-D) for values of $K = 1$ to $K = 10$.

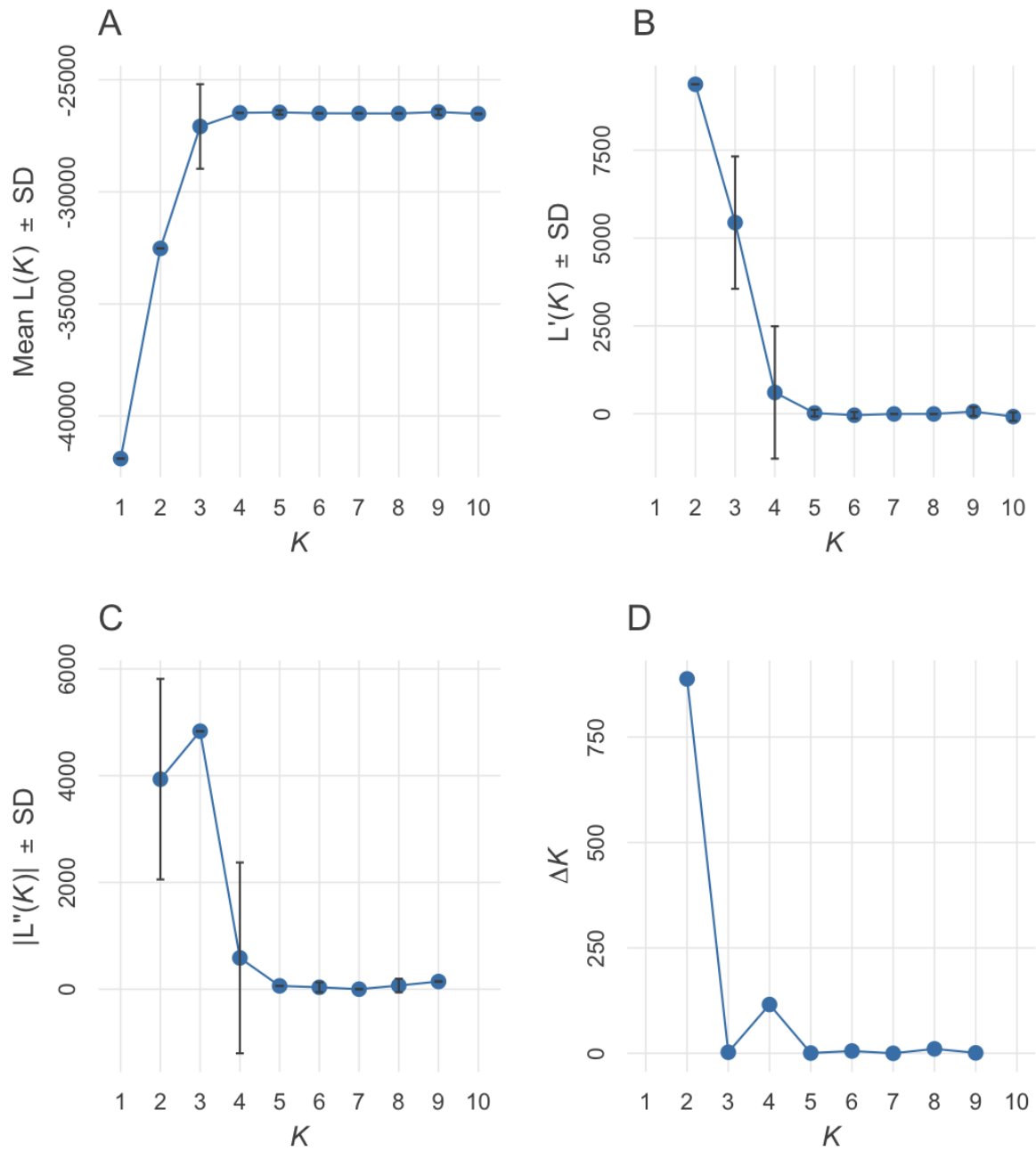


Figure B.10 Estimated ancestry according to latitude for individuals from sites indicated by color and shape of A) *D. sublineata* from the global set of Structure results and B) *D. carinulata* Fukang ecotype from the *D. carinulata* substructure results. The linear model for each distribution is plotted as a red line with standard error in grey shading.

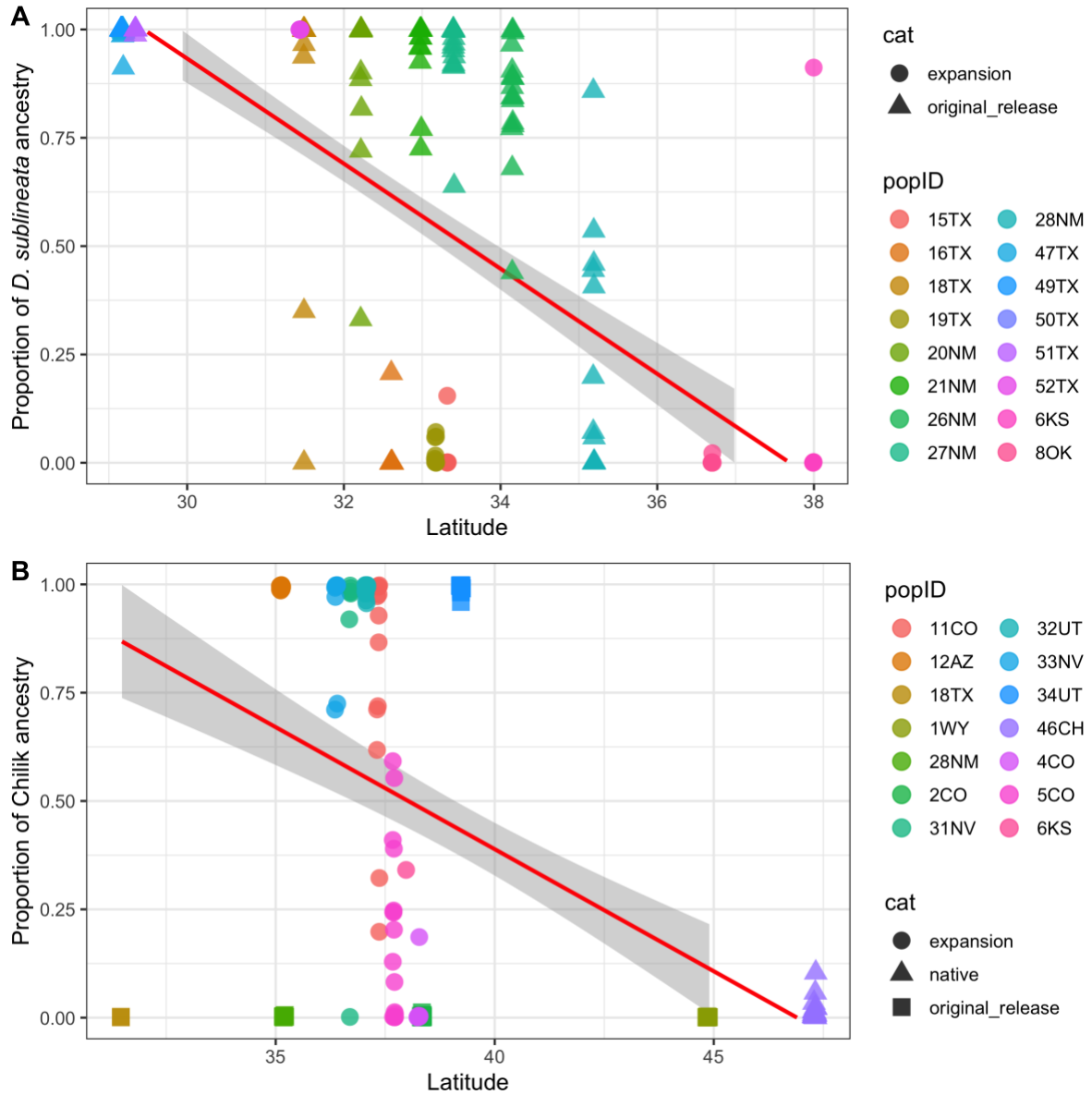


Figure B.11 Pairwise- F_{ST} compared to haversine distance between respective sites indicate patterns of isolation-by-distance associated with population structure for *D. elongata* collected from the native range in Greece (Mantel's $r = 0.086$, $p = 0.492$). Crossed-squares indicate pair-wise comparisons within ecotype, solid squares across ecotype. A linear model for each distribution projected behind points as a red line with standard error in grey to aid in visualization.

