Whole genome sequencing reveals the structure of environment associated divergence in a broadly distributed montane bumble bee, *Bombus vancouverensis*

Sam D. Heraghty*, Sarthok Rasique Rahman*, Jason M. Jackson*, Jeffery D. Lozier*

*Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA

Corresponding author: Sam D. Heraghty

Mailing address: Box 870344, University of Alabama, Tuscaloosa, AL, 35487

Phone: 205-348-2754

Email: sdheraghty@crimson.ua.edu

Abstract

Broadly distributed species experience divergent abiotic conditions across their ranges that may drive local adaptation. Montane systems where populations are distributed across both latitudinal and elevational gradients are especially likely to produce local adaptation due to spatial variation in multiple abiotic factors, including temperature, oxygen availability, and air density. We use whole genome resequencing to evaluate the landscape genomics of *Bombus vancouverensis* Cresson, a common montane bumble bee that is distributed throughout the western part of North America. Combined statistical approaches revealed several large windows of outlier SNPs with unusual levels of differentiation across the region and indicated that isothermality and elevation were the environmental features most strongly associated with these variants. Genes found within these regions had diverse biological functions, but included neuromuscular function, ion homeostasis, oxidative stress, and hypoxia that could be associated with tolerance of temperature, desiccation, or high elevation conditions. The whole genome sequencing approach revealed outliers occurred in genome regions with elevated linkage disequilibrium, elevated mean F_{ST} and low intrapopulation nucleotide diversity. Other kinds of structural variations were not widely associated with environmental predictors but did broadly match geographic separation. Results are consistent with other studies suggesting that regions of low recombination may harbor adaptive variation in bumble bees within as well as between species and refine our understanding of candidate genes that could be further investigated as possible targets of selection across the *B. vancouverensis* range.

Key words: Elevation, Local adaptation, Whole genome sequencing, Environmental association analysis, Population divergence

Introduction:

A major focus in evolutionary biology is understanding the genetic changes associated with species adaptation to abiotic conditions throughout their ranges (Manel et al. 2010, Orr 2005, Dillon and Lozier 2019). In broadly distributed species, populations must contend with different climatic stressors produced by large scale abiotic gradients (Savolainen et al. 2013, Cayuela et al. 2021), which may require unique adaptations that produce signatures of divergent selection within the genome (Hoban et al. 2016, Ahrens et al. 2018). Recent technological and statistical advances have created opportunities for identifying such signatures in wild populations of non-model organisms (Ellegren 2014, Ahrens et al. 2018, Luo et al. 2021) and the ever-increasing availability of species-specific reference genomes has made it possible to begin addressing questions about the genome structure of putative adaptations using whole genome data (Fuentes-Pardo and Ruzzante 2017, Taylor et al. 2021).

Species that occur in landscapes with substantial variation in environmental conditions provide opportunities for investigating environmentally associated genomic divergence that can indicate local adaptation (Joost et al. 2007, Eckert et al. 2010, Jackson et al. 2020, Yadav et al. 2020, Lim et al. 2021). Montane systems are an excellent example of a complex landscape where latitude and altitude together can produce changes in abiotic conditions over both large and small spatial scales (Keller et al. 2013, Rahbek et al. 2019). Some variables (e.g., air pressure, atmospheric oxygen, and temperature) will shift sharply across elevations (Dillon 2006, Cheviron and Brumfield 2012), while others, such as temperature, can vary with both elevation and more gradually with latitude. Such changes in environmental conditions and can impose strong selective pressures that may require physiological adaptations, and montane species with broad altitudinal and latitudinal ranges offer unique opportunities to sample multiple spatial-

environmental gradients to identify the genomic signatures of such adaptations (Dillon 2006, Cheviron and Brumfield 2012, Jackson et al. 2018, 2020, Montejo-Kovacevich et al. 2019, Pimsler et al. 2020).

In this study we employ whole genome resequencing (WGR) to evaluate genome-wide patterns of divergence and potential signatures of environmental adaptation in *Bombus vancouverensis* Cresson (Ghisbain et al. 2020), a bumble bee (Hymenoptera: *Bombus*) species that is abundant in mountainous regions of western North America. Bumble bees are a widely distributed genus of pollinators that consists of approximately 250 species globally (Cameron et al. 2007, Cameron and Sadd 2020). The genus is common in many biomes but has an evolutionary history associated with mountainous regions, which have played an important role in their diversification (Hines 2008, Williams et al. 2018, Lee et al. 2019, Orr et al. 2020). Several key traits may facilitate adaptation to complex montane landscapes in bumble bees. For example, bumble bees have evolved traits to deal with climate extremes, including insulating hairs (pile) and facultative thermoregulation mechanisms (e.g., "shivering" of flight muscles) for maintaining activity in cold temperatures, while also having the capacity to shunt excess heat from the thorax to abdomen to prevent overheating (Heinrich and Kammer 1973, Heinrich 1976, Heinrich 2004, Woodard 2017). Recent work suggests that intraspecific populations vary in thermal tolerance across species ranges, especially cold tolerance (Pimsler et al. 2020, Martinet et al. 2020), and gene expression under thermal stress suggests that populations may exhibit molecular variation that can facilitate responses to temperature extremes (Pimsler et al. 2020). Beyond temperature, bumble bees have traits that may facilitate life across elevations, including for flight at reduced air density and oxygen levels, such as metabolic changes, wing beat

93 kinematics and reduced body size or wing loading (Dillon 2006, Dillon and Dudley 2014, Liu et
94 al. 2020, Lozier et al. 2021).

95 This study expands on recent population genomics work (Jackson et al. 2018, 2020) that
96 used RAD-tag sequencing to characterize spatial-environmental drivers of gene flow and identify
97 selection in *B. vancouverensis* across the Sierra-Cascade region of the western United States.
98 Like many North American *Bombus* species (e.g., Lozier et al. 2011), *B. vancouverensis* is
99 characterized by weak population structure (e.g., mean F_{ST} is < 0.05) (Lozier et al. 2013, Jackson
100 et al. 2018), although genetic differentiation does increase with distance and with bioclimatic
101 resistance (Jackson et al. 2018). Environmental Association Analysis (EAA) with these RAD-tag
102 data has revealed unusual associations with bioclimatic variables for over 100 single nucleotide
103 polymorphisms (SNPs) in genes that were consistent with signatures of selection across the
104 genome. Morphological analyses have also found intraspecific body size clines across the region,
105 with notable reductions in mass and wing loading in bees from the southern High Sierras portion
106 of the *B. vancouverensis* range (Lozier et al. 2021), further suggesting the possibility of local
107 adaptation.

108 An important caveat to prior population genomic analyses in *B. vancouverensis* is that
109 linkage disequilibrium (LD) is weak in *Bombus* (Stolle et al. 2011, Sadd et al. 2015). In such
110 situations, RAD-tag data may incompletely survey the genome (Lowry et al. 2017) and a WGR
111 approach could prove beneficial for identifying localized signals of selection (Fuentes-Pardo and
112 Ruzzante 2017). Jackson et al. (2020) also relied on mapping reads to the genome of a related
113 species, *Bombus impatiens*, but an annotated *B. vancouverensis* genome has recently been
114 published (Heraghty et al. 2020), which may improve the ability to detect species-specific
115 regions harboring locally adaptive variants. Whole genome data also enable examination of other

aspects of genomic architecture that are relevant in understanding patterns of adaptation and evolution. Structural variants (SV's) are a variety of different mutations (e.g. inversions, deletions, insertions, etc.) that can provide insight in population structure (Dorant et al. 2020, Cayuela et al. 2021) as well as adaptation (Joron et al. 2011, Wellenreuther et al. 2019). SV's have been found to play a role in elevation adaptation in the European Honey Bee *Apis mellifera* (Wallberg et al. 2017) and could also be relevant in *Bombus.* In addition, spatial patterns of linkage disequilibrium and nucleotide diversity in the genome can give insight into processes of both adaptation and speciation that may not be visible without whole genome data (Christmas et al. 2021). Regions of elevated F_{ST}, decreased nucleotide diversity and increased linkage disequilibrium, often referred to as "Islands of Divergence", can be produced following divergence between both populations and closely related species, and may be especially interesting when observed in the face of gene flow among populations within species (Ellegren et al. 2012, Papadopulos et al. 2019, Christmas et al. 2021).

We use WGR of *B. vancouverensis* sampled across a latitudinal montane gradient in the Sierra-Cascades region of western North America to examine patterns of differentiation across the genome and refine possible targets of environmentally-associated selection that were previously suggested from reduced representation sequencing. We combine landscape genomics approaches that employ F_{ST} outlier detection with Environmental Association Analysis (EAA) models to identify locally adapted genetic markers candidate SNPs (Hoban et al. 2016, Storfer et al. 2018, De la Torre et al. 2019, Hartke et al. 2020, Lim et al. 2021). The objectives of this study are to sample across broad latitudinal and altitudinal gradients to identify genomic variants (SNPs and SVs) that are associated with environmental variation, especially those that may contribute to local adaptation to key environmental variables such as temperature and elevation.

We also aim to evaluate patterns of F_{ST}, nucleotide diversity, and linkage throughout the genome that may help better characterize the nature of regions exhibiting unusual interpopulation differentiation. Overall results provide new data on how complex landscapes drive changes in genomic variants, diversity, and architecture in montane species.

Materials and Methods

Sampling, DNA Extraction, and Whole Genome Resequencing

Bombus vancouverensis workers (diploid females) were selected for whole-genome resequencing from previously collected samples (Jackson et al. 2018, 2020) representing populations from environmental extremes across elevation and latitude in the California, Oregon, and Washington portion of the species range (36°N – 48°N latitude, 49m - 2900m elevation)(Figure 1, Supp. Table S1). Detailed characterization of sampling localities was presented in (Jackson et al. 2018, 2020). We attempted to include bees from locations that were at a relatively high and relatively low elevation across latitudes, but *B. vancouverensis* is generally restricted to its highest elevation sites in the southern part of the species range in the High Sierras, while the lowest elevation sites are in the north. Flying bees were collected at each site using sweep nets and placed in 100% ethanol on dry ice. Samples were ultimately stored in ethanol at -80°C.

DNA was extracted from thoracic muscle tissue using the Qiagen DNeasy Blood and Tissue kit (Hilden, N.R.W., Germany). Genomic DNA libraries were prepared using the NEBnext Ultra II FS DNA kit (Ipswich, MA, USA) with subsequent 150 bp paired-end sequencing using an Illumina NovaSeq 6000 (one lane at HudsonAlpha Institute for Biotechnology, Huntsville, AL, and another one at Psomagen, Rockville, MD).

Read Mapping and Variant Calling

Raw reads were processed using BBDuk v37.32 (Bushnell 2020) to remove adaptors, trim low quality bases, and remove short reads (ktrim=r k=23 mink=11 hdist=1 tpe tbo ftm=5 qtrim=rl trimq=10 minlen=25). Trimmed reads were evaluated for quality using FastQC v0.11.5 (Andrews 2010). Reads were then mapped to the *B. vancouverensis* reference genome (NCBI RefSeq ID: GCF_011952275.1) (Heraghty et al. 2020) using the BWA-MEM algorithm of BWA v0.7.15-r1140 (Li and Durbin 2009) and generated alignment (i.e., SAM) files were converted to the binary (BAM) format using SAMtools v1.10 (Li et al. 2009). Picard tools v2.20.4 (Broad Institute 2019) was used to sort, mark duplicates, and index the binary alignment (BAM) files. Single nucleotide polymorphisms (SNPs) were called using freebayes v1.3.2 (Garrison and Marth 2012). An initial round of variant filtering was conducted on the variant calling file produced by freebayes using VCFtools v0.1.13 (Danecek et al. 2011) with the following flags: --remove-indels --min-alleles 2 --max-alleles 2 --minQ 20 --minDP 4 --max-missing 0.75. After visual inspection of the data, we removed an additional small number of SNPs (n=45,872) with unusually high coverage (>2x mean coverage) or excess heterozygosity (--hardy flag in vcfttools) that could indicate repeat regions or paralogous sequences. A final round of variant filtering was then performed to focus on SNPs from intact scaffolds (>100 kb in size) with minor allele frequencies (MAF) $\geq$ 0.05 to remove the influence of low frequency alleles and SNPs in regions that may have assembly artifacts. SNPs were annotated using SNPeff v4.3 (Cingolani et al. 2012) and missing data was imputed for some analyses that required a complete matrix using the phasing function in Beagle v5.1 (Browning et al. 2018).

Structural variants (SVs) were identified using DELLY v0.8.1 (Rausch et al. 2012) from the binary alignment (i.e.,BAM) files described above. DELLY was run with the -all parameter which enables detection of deletions, duplications, inversions and transversions. The DELLY output was converted (from a bcf to vcf format) using BCFtools v1.10.2 (Li et al. 2009) then filtered to only retain SVs that were supported by multiple high quality reads (PASS) and by split reads (PRECISE). As for SNPs, we only evaluated SVs on the genome longer than 100 kb.

Environmental Variable Selection and Genome Environment Association Analyses

Environmental conditions at each site were charactered using the 19 Bioclim variables at 0.5 arc minute resolution from the WorldClim v2 database (Fick and Hijmans 2017) via the raster v3.3-14 R package (Hijams 2021). To select the variables providing unique information for analysis, we performed item clustering using the iclust function with default settings in the *psych* v2.0.12 R package (Revelle 2020). A single variable was then selected from each cluster to be included in the subsequent modeling approaches. Although elevation is correlated with environmental conditions, elevation has its own associated stressors, such as reductions in air density and oxygen availability (Dillon 2006, Cheviron and Brumfield 2012, Lim et al. 2021). Given the hypothesized importance for bumble bees generally (Dillon et al. 2006) and *B. vancouverensis* specifically (Lozier et al. 2021), elevation was also included as a variable for Environment Association Analysis (EAA) to identify genomic regions that have undetected unique associations from any other bioclimatic variables.

Environmental association analyses to identify loci with putative signatures of local adaptation to abiotic variables were performed using latent factor mixed modeling (LFMM2 in LEA v3.0.0 R package) (Gain and François 2021). LFMM2 implements a least-squares approach

to identify SNPs with a significant association with the environmental variables of interest after controlling for population structure. Utilization of this approach is particularly advantageous in that it is faster than earlier versions of the software (Caye et al. 2019) and is more conservative in terms of false positives (Luo et al. 2021). The optimal number of population clusters (k) for population structure control was determined with sMNF (implemented in LEA v3.0.0 R package) (Frichot and François 2015) as the k represents the smallest cross-entropy. To account for multiple testing, the q-value v2.20.0 R package (Storey, Bass, Dabney, & Robinson, 2020) was used to transform the p-values produced by LFMM2 into q-values. Significant environmentally-associated loci were considered at a threshold of $q \leq 0.05$.

Employing multiple models for methodological cross-validation is a standard practice in studies to detect local adaptation to further reduce the potential for false positives (De la Torre et al. 2019, Hartke et al. 2020, Jackson et al. 2020). As a second approach to confirm results from LFMM2, we used OutFLANK v0.2 (Whitlock and Lotterhos 2015), implemented using the default settings in the SambaR R package/wrapper v1.00 (de Jong et al. 2021). We selected OutFLANK as a complementary approach because this method does not depend on associations with environmental variables like LFMM2 approach, but rather detects SNPs using an F_{ST} outlier approach. Individuals were pooled into populations by sampling coordinates and loci with heterozygosity ≥ 0.1 were used for the OutFLANK analysis (default). The p-values produced by OutFLANK were corrected for multiple testing by converting to q-values using default threshold settings to select outliers ($q \leq 0.01$). Cross-validated loci were then identified as those detected as significant from both LFMM2 and OutFLANK approaches.

To analyze the of structural variations detected in the *B. vancouverensis* genome, SV categories (deletions, duplications, inversions, and transversions) were separated and analyzed

using the RDA function implemented in vegan v2.5-7 (Oksanen et al. 2020). RDA (Redundancy Analysis) is a robust and flexible approach for a variety of questions in landscape genomics and is particularly advantageous for analysis of SVs because of the relative low numbers of detected SVs (Capblancq and Forester 2021). We use (Redundancy Analysis) RDA here because SVs also represent a novel data type for *B. vancouverensis*, and RDA provides the opportunity to simultaneously examine unusual differentiation at individual outlier regions alongside general patterns of population structure among individuals to compare with prior knowledge of overall population genetic structure in this species. Overall model significance was assessed using the anova.cca function in R (from the vegan v2.5.7 package) and each axis was tested for significance to identify which axes represented non-random variation (Legendre et al. 2011). Significant axes were then evaluated for outlier SVs based on axis loading exceeding 4 standard deviations (Forester et al. 2018).

Gene Ontology Enrichment Analysis

Gene Ontology enrichment analysis utilized recently generated species-specific annotations for *B. vancouverensis* downloaded from Hymenoptera Genome Database (Walsh et al. 2021) (Bombus_vancouverensis_HGD_go_annotation.gaf.gz ; last updated 6/29/2021). Genes without the annotation information in the annotation file were not considered in our analysis. The GoFuncR v1.8.0 (Grote 2020) R package was used for gene set enrichment analysis (GSEA) using the go_enrich function. A custom annotation database was created from the downloaded gene annotation file (in .gaf format) following GoFuncR manual guidelines (Grote 2020). Genes with cross-validated outlier SNPs (n = 44) were considered as candidate genes for Gene Ontology (GO) enrichment testing against all other genes as the background set (n = 9,432).

253 Statistically significant GO terms ($p < 0.01$) were retained and then the Revigo web server
254 (Supek et al. 2011) was used to summarize redundant GO terms using medium stringency filter.
255
256 **Patterns of Chromosomal Diversity in Outlier Regions**
257 To examine the patterns of diversity and differentiation in environmentally associated
258 candidate regions relative to genome-wide patterns, fixation index (F_{ST}) for each site was
259 calculated using the Weir and Cockerham method as implemented in the SNPRelate R package
260 v1.22.0 (Weir and Cockerham 1984, Zheng et al. 2012) and per-site nucleotide diversity (π) for
261 each population was calculated with the –site-pi flag in VCFtools v0.1.13 (Danecek et al. 2011)
262 and averaged to obtain a mean within-population nucleotide diversity (π). For both statistics,
263 averages were calculated across 5kb widows using the GenomicRanges R package v1.40.0
264 (Lawrence et al. 2013). Visualization of scaffolds with outlier regions and their associated F_{ST}
265 and π were generated using the Gviz v1.34.1 R package (Ivanek 2016). Linkage disequilibrium
266 (LD) was calculated for the entire genome and the major candidate outlier regions using
267 PopLDdecay v3.41 (Zhang et al. 2019) with the max-distance between loci set to 300kb. To
268 visualize differences in LD patterns between the outlier-dense and outlier-free regions, we also
269 plotted LD for ten randomly selected scaffolds that did not contain outlier loci (Supp. Table S2).
270 Previous studies examining the genomics of divergence between bumble bee species,
271 including that of the *B. bifarius-B. vancouverensis* species complex, have found that many
272 putatively adaptive regions of the genome (i.e., "islands of divergence") between sister species
273 are maintained in repetitive and low recombination regions along the chromosome (Ghisbain et
274 al. 2020, Christmas et al. 2021). First, we downloaded the repeat masker output file from the *B.*
275 *vancouverensis* RefSeq assembly (NCBI RefSeq ID: GCF_011952275.1) (Heraghty et al. 2020)

and calculated the number of repeats in same 5kb windows used for the F_{ST} calculations above. Linear regressions (lm function in R) were used to test for a relationship between repeat region density and F_{ST} in each window (genome-wide and for focal scaffolds). Second, although the *B. vancouverensis* assembly is sufficiently intact for most analyses, it does not have chromosome level integrity. We thus took advantage of the high degree of synteny in *Bombus* genomes (Sadd et al. 2015, Sun et al. 2020) to determine probable locations of major candidate regions with strong cross-validated environmental associations (i.e., scaffolds with multiple outlier regions, or with genes containing many outlier SNPs) using the chromosome-level assembly for *B. terrestris* (RefSeq ID: GCF_000214255.1). For each candidate region, we determined the orthologous *B. terrestris* chromosome using BLASTn (Altschul et al. 1990) and aligned the highly divergent outlier-dense *B. vancouverensis* scaffolds to the *B. terrestris* chromosomes using MAUVE algorithm implemented as a plugin for Geneious v 2021.0.3.

Results

Data Summary

We sequenced 122 female workers from 19 localities to an average estimated coverage of ~40x (32,997,359 ± 16,271,517 SD read pairs per library), and 33,439,776 SNPs were called using Freebayes. The final filtered data set included 1,369,356 SNPs (MAF ≥ 5%, depth > 4x), with a mean sequencing coverage of 25.3 reads per SNP per individual and a mean of 0.34% missing data per SNP per individual.

For environmental variables, item clustering analysis identified 4 clusters from the total set of WorldClim Bioclim variables (Supp. Figure S1). We retained the following variables for

298 analysis: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3),
299 and annual precipitation (BIO12), as well as elevation.
300
301 **Environmental Association and F_{ST} Outlier Analyses**
302 Two population clusters were identified by sMNF using the minimum cross-entropy
303 approach and we specified $k = 2$ for population structure control in LFMM2 (same as Jackson et
304 al. 2020). Across all variables, LFMM2 detected a total of 774 unique SNPs that were
305 significantly associated with one of the environmental variables, representing 154 unique genes
306 and 66 intergenic regions. Isothermality (BIO3, 646 SNPs) and elevation (340 SNPs) had the
307 largest number of significant SNP associations ($q \leq 0.05$). BIO1 (annual mean temperature),
308 BIO2 (mean diurnal range) and BIO12 (annual precipitation) were associated with 0, 15, and 46
309 SNPs, respectively. The OutFLANK F_{ST} outlier approach produced 1,274 outlier SNPs (with
310 default SambaR settings, $q \leq 0.01$). There were 551 cross-validated SNPs (in 44 genes and 24
311 intergenic regions) shared between the two methods (Table 1, Supp. Table S3); as above, these
312 were most frequently associated with isothermality and elevation (Table 1).
313 The densest clusters of cross-validated loci were found on scaffolds NW_022881829.1
314 and NW_022881902.1 (Fig 2). The most notable group of cross-validated SNPs in the genome
315 were located on NW_022881829.1 within the adjacent genes LOC117157569 (*Sax-3-like*,
316 homologous to *dpr20* in *D. melanogaster*, n = 36 SNPs) and LOC117157568 (synaptogenesis
317 *protein syg-2-like*, *side-VI* in *D. melanogaster*, n = 66 SNPs) and their intergenic region (n = 364
318 SNPs) (Table 1, Fig. 2). Two additional dense clusters of cross-validated SNPs were grouped
319 into two regions on scaffold NW_02991902.1. In the first of these clusters (~350 - 680 kb region
320 of NW_02991902.1) outliers were present across 11 genes, mostly with one or two SNPs each.

The gene with the largest number of SNPs in this cluster is LOC117161116 (uncharacterized, homologous to *CG13138* in *D. melanogaster*, n = 15 SNPs). The second cluster (~1,250 kb – 1,470 kb region of NW_02991902.1) comprised several genes with multiple statistically significant SNPs. The genes with the largest number of SNPs in this cluster were LOC117161100 (*plasma membrane calcium-transporting ATPase 3*, homologous to *PMCA* in *D. melanogaster*, n = 36 SNPs), LOC117161180 (*xanthine dehydrogenase/oxidase-like*, homologous to *AOX3* in *D. melanogaster*, n = 36 SNPs), and LOC117161157 (uncharacterized in *B. vancouverensis* but homologous to *beat-IIIc* in *D. melanogaster*, n = 5 SNPs). A fourth notable cluster of statistically significant SNPs was detected on a separate scaffold (NW_022881786.1) in LOC117156535 (*Multidrug resistance-associated protein 4-like*, best BLAST homology to *CG5789* and *Mrp4* in *D. melanogaster*, n = 6 SNPs), including five non-synonymous polymorphisms. This region is particularly interesting as the largest set of SNPs that were significantly associated with elevation alone and not with any other variables. The remaining significant SNPs (32 genic, 27 intergenic) were distributed more sparsely, with relatively few SNPs per gene (Table 1).

Gene Ontology (GO) enrichment analysis of Cross-validated Loci

Gene Ontology enrichment analysis was used to explore general functions of outlier genes. Our analysis returned an initial list of 87 GO terms ($p < 0.01$), which was subsequently reduced to 51 biological terms, 15 cellular terms and 19 molecular terms (Supp. Table S4) using Revigo summarization web tool. Some outlier loci were excluded from GO analysis due to the genes having no GO annotations (Walsh et al. 2021). There were several notable trends which generally reflected the functions of genes in the outlier dense regions. For example, several terms

were related to ion transport, including GO terms like "calcium ion transport" (GO:0006816) and "P-type calcium transporter activity involved in regulation of presynaptic cytosolic calcium ion concentration" (GO:1905056) (reflecting genes like, *plasma membrane calcium-transporting ATPase-3*, LOC117161100). "ATPase-coupled transmembrane transporter activity" (GO: 0042626) is associated with genes such as the elevation-specific gene *multidrug resistance protein 4* (LOC117156535). There were also several terms involved with neuron and synapse function with terms such as "synaptic target recognition" (GO:0008039) (e.g., associated with LOC117161157, a *beatIIIc* homolog). Several terms were related to cardiac function including annotations; for example, "regulation of cardiac muscle tissue development" (GO:0055024) and "cardiac myofibril assembly" (GO:0055003).

Analysis of Structural Variants (SVs)

Our structural variant detection analysis using DELLY detected 7,419 deletions, 226 duplications, and 6,303 inversions. The overall RDA models for all three SV types were significant and contained significant axes, representing non-random variation (Legendre et al. 2011) (Supp. Table S5). Twenty-two deletion outliers and 3 inversion outliers were identified, all on axis 1 for their respective model (axis loading ≥ 4 standard deviations). Although the duplication RDA model was significant, no outliers were identified. Most SVs were less than 1 kb and spanned over at least one gene (Table 2). Clustering patterns for individuals in the RDA model generally reflected geographic relationships, with samples from nearby localities loading near one another on the RDA graph (Fig 3). The deletions model best captures the geographic relationship between samples with relatively clear clustering by state (CA,OR, and WA) (Fig 3). Mirroring the SNP analysis, the strongest predictive variables were isothermality (16 of 19

deletions, 2 of 3 inversions) and elevation (3 of 19 deletions, 1 of 3 inversions). The highest density of outliers was found on scaffold NW_022881829.1 (5 deletions) and occurred in same region of high outlier SNP density as SNP outliers discussed above. One other notable SV was a deletion in LOC117160713 (*Chaoptin-like*, homologous to *CG42346* in *D. melanogaster*) which also contained an outlier SNP (Table 2). There were several SVs located on scaffold NW_022881881.1 (3 deletions and 1 inversion), but visual inspection suggested that some of the detected SV outliers may represent artifacts, highlighting the general challenges to structural variant calling. For example, two of the SVs on NW_022881881.1 were far larger than average and spanned over numerous genes; one 385kb deletion falls in a region with a large number of small repeats (average length 51.3 bp, n = 184) which may complicate calling SVs (Mahmoud et al. 2019).

Patterns of Chromosomal Diversity in Outlier Regions

The three largest environmentally associated regions containing clusters of cross-validated SNPs (NW_02881786.1, NW_02881829.1, and NW_02881902.1) exhibited markedly increased F_{ST} and decreased π (Fig 4). Average F_{ST} across the genome was low (global F_{ST} = 0.02) but increased to values > 0.4 in localized regions around outliers. Nucleotide diversity shows the inverse pattern, with average values between 0.002 and 0.004 but dropping to less than 0.001 in windows of environmentally associated regions (Fig 4). Linkage disequilibrium (LD) was also elevated in these regions relative to the genome wide average (Fig 5). The increased LD was apparent when averaged across scaffolds with environmental associations vs the overall genome average, as well as for individual comparisons of the two main outlier-containing scaffolds against a set of randomly selected comparable scaffolds with no outlier regions (Fig 5).

In both sets of comparisons, the mean r^2 measure of LD is higher for the environmentally associated regions, with mean r^2 for the genome-wide average and random scaffolds dropping to zero within a shorter distance (~25 kb), whereas the scaffolds with strong environment associations showed elevated r^2 as far as 300 kb (Fig 5). Scaffold NW_02881829.1, which has the largest outlier region, also had the highest mean r^2 in the genome (Fig 5).

The elevated LD in putative outlier regions is consistent with patterns observed for interspecies divergence islands identified in the *B. vancouverensis – B. bifarius* species complex (Christmas et al. 2021). To localize approximate chromosomal locations of scaffolds with strong genome environment associations, we aligned the scaffolds to homologous regions of the *B. terrestris* linkage groups (near chromosome level assembly). Two regions of interest (NW_02881829.1 and NW_02881786.1) were mapped to linkage group LG B11 of *B. terrestris* (Fig 5). NW_02881829.1, which contained the densest cluster of SNPs in the genome maps nearly on the end of LG11 (~0-1Mb), suggesting these genes probably also lie near the chromosome end in *B. vancouverensis*. NW_02881786.1, which contains one of the top genes associated only with elevation (LOC117156535, *multidrug resistance-associated protein 4-like*) maps to the center of LG B11 (between ~6-8Mb position along LG B11) but does not appear to be located near the putative centromere (see Christmas et al. 2021). There was a very weak albeit significant relationship between repeat density and F_{ST} in 5 kb windows across the genome ($R^2 < 0.001$, $F_{1, 48354} = 36.92$, $p \leq 0.001$), however, not all outlier-dense scaffolds showed this pattern. NW_022881829.1 (located near a telomere) did show the positive relationship between F_{ST} and repeat density ($R^2 = 0.0636$, $F_{1, 143} = 9.718$, $p = 0.0022$), but neither NW_022881786.1 ($R^2 = 0.0018$, $F_{1, 361} = 0.6564$, $p = 0.42$) or NW_022881902.1 ($R^2 = 0.0019$, $F_{1, 293} = 1.56$, $p = 0.212$)

showed a relationship, suggesting that not all highly differentiated SNPs were in unusually repeat-rich regions (Fig 5).

Discussion

Using a whole-genome sequencing dataset for the montane bumble bee *B. vancouverensis*, we discovered strong associations with environmental variables and unusual patterns of diversity at several genomic regions based on environmental association analysis (EAA) of SNPs, structural variants (SVs), and genome-wide patterns of diversity and linkage. Several regions throughout the genome had large numbers of SNPs that were predominantly associated with isothermality and/or elevation after controlling for population structure and using two independent methods, and thus show potential signatures of local adaptation (Fig 2). These association peaks tended to fall in regions of the genome that had increased linkage (Fig 5), low π and elevated F_{ST} (Fig 4) compared to the rest of the genome. Several outlier structural variants were also found within these regions, and although SVs produced patterns of regional population structure consistent with prior SNP results (e.g., Jackson et al. 2018), there were relatively few SV outliers and more work is required to better characterize these mutation types. Many of the SNP outlier genes detected here have putative functions that would have adaptive value in montane regions, including neuromuscular development and function, ion transport, and hypoxia resistance. These putative functions would be relevant for *Bombus* that must fly and thermoregulate in landscapes characterized by strong elevational changes and associated abiotic variation.

Patterns of environmental association

Genome regions with unusual patterns of diversity detected by both LFMM2 and OutFLANK were predominantly associated with isothermality and elevation, consistent with prior RAD-Seq data in this species (Jackson et al. 2020). Some of the most notable outlier containing genes appear to play a vital role in neural development, particularly neuromuscular synapse formation and muscle formation, with overrepresented terms like "regulation of cardiac muscle tissue development" and "cardiac myofibril assembly" and "synaptic target recognition" (Supp. Table S4). For example, two genes LOC117157569 (*Sax-3-like*; *dpr20* in *D. melanogaster*) and LOC117157568 (*synaptogenesis protein syg-2-like*; *side-VI* in *D. melanogaster*) were found in the highly associated region of environmentally associated differentiation on NW_022881829.1; both genes code for immunoglobin domain proteins related to synapse formation, especially in muscle (Igsf et al. 2015, Cheng et al. 2019). Notably, another outlier, LOC117161157 on scaffold NW_022881902.1, is homologous to *beat-IIIc* in *D. melanogaster*, which together with *side-VI* belong to the beaten path-sidestep interaction networks involved in neuromuscular development in *D. melanogaster* (Li et al. 2017). All of these genes were previously identified in the RAD-Seq data (Jackson et al. 2020). Other cross-validated SNPs in genes that may be involved in synapse formation were detected outside of the major environmentally associated regions, including LOC117158593 (SNPs in *cilia- and flagella-associated protein 20-like*; *Bug22* in *D. melanogaster*) and LOC117153469 (E3 ubiquitin-protein ligase Nedd-4; *Nedd4* in *D melanogaster*) (Schnorrer et al. 2010, Zhong et al. 2011). Related to these neural and muscle function terms, GO analysis also provided support for the importance of ion homeostasis (e.g., GO term "calcium ion transport") (Supp. Table S4). For example, significant SNPs associated with isothermality and elevation were also identified in

458 *plasma membrane calcium-transporting ATPase 3*, which maintains calcium ion levels including
459 at the neuromuscular junction and could be relevant because of the important role calcium plays
460 in insect neuromuscular function (Iwamoto 2009) and cold tolerance (Seamus et al. 2018). The
461 redundancy analysis (RDA) of structural variants (SVs) also detected outliers in genes related to
462 muscle function that were not found in the SNP analysis, such as a deletion in LOC117165908
463 (*syntaxin-binding protein 5*; *tomosyn* in *D. melanogaster*) which has a role in motor axon
464 guidance and synaptogenesis (Kraut et al. 2001).

465 Neural and muscular function are expected to be highly relevant for bumble bees across
466 montane landscapes. Muscle function is crucial for thermoregulation to achieve flight in cold
467 temperatures for bumble bees (e.g., via shivering of thoracic muscles) (Heinrich 2004), and it is
468 possible that some candidate genes reflect selection related to stresses from varying thermal
469 conditions across the *B. vancouverensis* range. Furthermore, analyses in another species from
470 this region, *B. vosnesenskii*, found that lower critical thermal limits (CT_{MIN}) correlated strongly
471 with temperature in populations in replicate elevation transects (Pimsler et al. 2020). Although
472 such data is not available in *B. vancouverensis*, CT_{MIN} in bumble bees and other insects is likely
473 associated with physiological failure of the neuromuscular junction (Oyen et al. 2016, Overgaard
474 and MacMillan 2017), providing another possible mechanism by which selection may shape
475 variation across genes related to neural and muscle function across elevations. Muscle function
476 also may reflect stresses associated with bumble bee flight itself, which may face challenge in
477 high elevation populations as environmental conditions related to elevation have been noted to
478 alter various traits related to bumble bee flight such as wing loading and wing beat amplitude
479 (Dillon and Dudley 2014, Lozier et al. 2021).

The largest SNP cluster associated only with elevation and no other Bioclim variable was in the gene *multidrug resistance-associated protein 4-like* (*MDRP4*, LOC117156535) on one of our focal scaffolds NW_022881786.1. This gene contained six cross-validated SNPs including five nonsynonymous substitutions. This elevation-specific association is intriguing because in *D. melanogaster*, *multidrug resistance protein 4* is involved in the response to hypoxia and oxidative stress (Huang and Haddad 2007) as well as being associated with adaptation to elevation in alpine stoneflies (McCulloch et al. 2021). Hypoxia is a major challenge associated with elevation and can be associated with a variety of responses such as changes in tracheal physiology, metabolism, and activity levels (Harrison et al. 2018). Given that *MDRP4* is one of the few genes with outlier SNPs associated solely with elevation, this gene could be a strong candidate for future research on elevational adaptation in *B. vancouverensis*.

Finally, it is notable that we found very few SNPs associated with precipitation (n = 6), which is interesting given cross-validated SNPs were identified in several genes that had functions involving the cuticle, which was also reflected in the GO results with terms such as "cuticle hydrocarbon biosynthetic process" (Ferveur et al. 2018, Krupp et al. 2020). The cuticle plays an important role in desiccation tolerance by limiting the water loss in response to the environment (Ferveur et al. 2018). The lack of selective signal associated with precipitation suggests a different variable may be shaping cuticle development, including a link between desiccation and thermal tolerance (Sinclair et al. 2013, Nguyen et al. 2017, Manenti et al. 2018) (please see below for further discussion of issues with assigning causal roles to environmental variables).

The Influence of Genomic Architecture on Adaptive Signatures

Many of the genes with outlier SNPs were also discovered in a previous RAD-Seq study of *B. vancouverensis* (Jackson et al. 2020). This was somewhat surprising as bumble bees tend have weak linkage disequilibrium (Sadd et al. 2015, Sun et al. 2020) that should reduce the likelihood of sequencing SNPs that are or are linked to causal mutations using RAD-Seq (Fuentes-Pardo and Ruzzante 2017, Lowry et al. 2017). In this case, it appears that RAD-Seq data were in fact able to capture many of the same signatures of divergence detected here owing to the large size of the main outlier regions, with LD and interpopulation differentiation within these regions much larger than is typical in the remaining *B. vancouverensis* genome (Fig 5). The whole-genome resequencing approach did identify loci that were not detected in the RAD-Seq data, however, with the unique elevation-associated peak in *MDRP4* being particularly notable, as well as providing the potential to detect structural variants throughout the genome, although these appear more limited. Taken together, our results suggest that RAD-Seq may be a useful tool for capturing many genome-wide patterns of differentiation across large numbers of populations and individuals in bumble bees, however, utilization of whole genome data will likely be required to identify signatures of selection outside of high LD regions. At the same time, if high LD regions harbor many putative adaptive loci, identifying sites targeted by selection within blocks of linked SNPs will remain a challenge even within whole genome data.

Some of the results presented in our study also parallel recent genome resequencing experiments that have identified how structural genome divergence may be related to gene flow barriers in bumble bees, with such regions potentially harboring loci with adaptive significance during the speciation process (Christmas et al. 2021). In two bumble bee species complexes, one of which included *B. vancouverensis* with its sister species *B. bifarius*, islands of elevated

526 divergence between taxa were generally found in areas of low recombination. Our LD analyses
527 show similar patterns at the intraspecific level within *B. vancouverensis*, suggesting that genome
528 architecture may play a role in within-species divergence even under overall general pattern of
529 low genome-wide differentiation and ongoing gene flow (e.g., F_{ST} = ~0.02 here) and that many
530 potential signatures of adaptation may reside in regions of low recombination. Placement of
531 several scaffolds with dense outlier clusters in the *B. terrestris* genome indicates that some
532 highly divergent loci fall in regions that may experience low recombination, such as near the
533 ends of chromosome LG B11 (Fig 5). Although allopatric divergence in the interspecific
534 comparisons (Christmas et al. 2020) did not specifically focus on major signatures of divergence
535 near the ends of chromosomes, regions near telomeres can harbor islands of divergence in other
536 species (Ellegren et al. 2012). Further research spanning the continuum of divergence from
537 populations to species will provide important clues into the emergence of islands of divergence
538 in bumble bees and their potential roles in intraspecific adaptation and speciation (Christmas et
539 al. 2021).

540 As mentioned earlier, detected structural variants (SVs) were far more limited than SNPs
541 both in total number and number of outliers. The lack of SVs in general, as well as some of the
542 large detected SVs that we suspect are artifacts in repeat-rich or polymorphic regions, may be
543 explained by the challenges in detecting structural variants from short reads (Mahmoud et al.
544 2019), so a targeted study using long read sequencing approaches may be required to fully
545 understand the role of structural variants in relation to the adaptation to environmental variables.
546 While the lack of outlier SVs relative to outlier SNPs is likely driven by the relatively small
547 number of SVs, there is a high degree of synteny across the genus (Sun et al. 2020) that may
548 indicate large adaptive SVs are relatively uncommon in *Bombus*. That said, prior work on

549 elevational selection signatures in honey bees has demonstrated the importance of structural
550 variation, inversions in particular, for harboring locally adaptive genetic variation (Wallberg et
551 al. 2017), so additional work on detecting the role of large scale SVs in bumble bees is
552 warranted. Despite the limited number of outlier SVs, the parallels in results between the SNP
553 and SV analysis suggest that they may be shaped by similar forces. Both analyses showed
554 isothermality and elevation as driving evolution and there was considerable overlap between the
555 genomic regions that had outliers. Some of the genes with only SV outliers, such as
556 LOC117159534 (innexin shaking-B, *ShakB* in *D. melanogaster*) also seemed to have functions
557 similar to those with SNP outliers, and such functional overlap indicates that SVs and SNPs
558 could both contribute to local adaptation. It is also of interest that SVs generally captured
559 patterns of population structure in this species (Jackson et al. 2018), which has been noted in
560 other species as well, even outperforming SNP based markers (Dorant et al. 2020, Cayuela et al.
561 2021). Our data supports this prior work as it suggests SVs may be utilized as useful genetic
562 markers generally for population genomic studies.
563
564 **Do outlier environmentally associated regions indicate selection?**
565 The cross-validated outlier regions detected in this study are clearly unusually divergent
566 across the *B. vancouverensis* range compared to most of the genome; however, assigning
567 adaptive function to these candidate loci will require additional research. Environmental
568 variation is commonly spatially autocorrelated so it is important to consider demographic effects
569 that may produce allele frequency differences among populations (Hoban et al. 2016). EAA
570 models that explicitly incorporate population structure, like LFMM2, have a reduced false
571 discovery rate compared to models that are unable to account for population structure or even

572 earlier versions of the same methods (Luo et al. 2021). The models employed in our research to
573 detect outliers and environmental associations are both relatively robust to population structure,
574 with LFMM2 incorporating the number of population clusters (k) to account for the underlying
575 genetic differentiation that could arise from population structure or isolation by distance and
576 OutFLANK also having low false positive rates under a range of demographic models (Whitlock
577 and Lotterhos 2015, Luo et al. 2021). Since the two methods have different statistical
578 assumptions, in combination, they add support for many of the detected outliers being true
579 positives. The low diversity and elevated F_{ST} in the cross-validated outlier regions are also
580 consistent with selective sweeps, but it is important to consider that outlier "islands of
581 divergence" can arise due to neutral processes (Quilodrán et al. 2020), especially in regions of
582 low recombination (Booker et al. 2020). Using the genome data alone it is difficult to determine
583 whether any of the outlier regions in high linkage regions harbor loci under selection, and
584 certainly the large number of linked outlier SNPs or SVs within certain genes (e.g.
585 LOC11515769, LOC117157568) are not all adaptive. Most SNPs in outlier regions were
586 noncoding and the co-localization with structural variants could indicate these are "structural
587 islands" that result from neutral processes and do not necessarily have adaptive value (Ravinet et
588 al. 2017), but could also be subject to hitchiking from a recent selective sweep in the region
589 (Kim & Nielsen 2004). Further, the non-coding SNPs may fall in regulatory regions and have
590 regulatory implications (Wittkopp and Kalay 2012) so this does not necessarily preclude a role
591 for selection. That said, some outlier regions (e.g., the elevation-associated *multidrug resistance*
592 *protein 4*) include non-synonymous changes, a unique association with an environmental
593 variable, and little evidence for strong linkage or elevated repeat structure. We thus find it likely
594 that outliers regions detected in our study do harbor loci that may be contributing to local

595 adaptation, but also may contain loci that are being shaped by other processes (e.g. barriers to
596 gene flow) (Ravinet et al. 2017).
597 A similar issue that results from autocorrelation of environmental variables is that
598 bioclimatic variables included in analyses may not be the direct selective pressure influencing
599 outlier SNP frequencies (Ahrens et al. 2018, Jackson et al. 2020). Using abiotic variable
600 reduction helps with correlations for modeling but assigning adaptive significance in the face of
601 such correlations remains a general challenge. The variables used in the models can provide
602 important insight into the factors shaping the biology of *B. vancouverensis* even if the direct
603 causal variable is not included. Isothermality, which had the largest number of outlier loci,
604 represents the size of the average daily temperature range relative to the annual temperature
605 range (O'Donnell and Ignizio 2012) and thus could capture the biological phenomena relating to
606 season length, day length, or elevation, in addition to effects on thermal tolerance (CT_{MIN})(Wang
607 and Dillon 2014, Diamond and Chick 2018, Jackson et al. 2020). Given the seeming importance
608 of CT_{MIN} variation across *Bombus* species ranges (Pimsler et al. 2020, Martinet et al. 2021), it is
609 possible that bumble bees in thermally variable environments that experience a wider array of
610 temperature fluctuations may require genetic changes relating to thermoregulation (Heinrich
611 2004). However, pinpointing the specific abiotic factor driving potentially adaptive shifts in
612 allele frequency at outlier loci with respect to any predictor variable will require physiology
613 experiments on populations from throughout the *B. vancouverensis* range.
614 In conclusion, we detected several regions within the genome with outlier variants that
615 have associations with environmental variables of interest in *B. vancouverensis*. Our results
616 provide detailed data on the factors shaping within species genetic diversity within this bumble
617 bee across its range and provide a useful starting point for rigorous field and lab-based

experiments to assess if these candidate genes play a role in adaptation and how they might alter fitness for niche-specific adaptation. Although lab rearing of wild bumble bees is particularly challenging, improving techniques to maintain laboratory colonies of montane species have increased the feasibility of geographically comprehensive common garden experiments (e.g., Pimsler et al. 2020) and developmental studies (Tian et al. 2019, Rahman et al. 2021b). For instance, given our detection of outlier SNPs relating to neuromuscular function, cuticle composition, and hypoxia resistance, further studies could be designed to evaluate differences in physiology between individuals from different regions of the range examining the various properties of muscle function or exoskeleton composition. Our results also provide novel insights into population divergence across complex landscapes, which could play an important role in addressing evolutionary questions and might be especially helpful to contend with more practical conservation related issues such as understanding how current and future environmental changes from the global warming and climate change may shape the future distributions of species and their underlying adaptive genetic variation.

Acknowledgements

We thank the University of Alabama College of Arts and Sciences and the National Science Foundation (DEB-1457645 and URoL 1921585 to J.D.L.) for support relating to this project.

Data Availability

Raw sequencing data is available on NCBI SRA (Bioproject PRJNA858769; accession numbers SAMN29751459 - SAMN29751581). Variant data for SNPs and SVs and scripts used

641 in data filtering and analysis are available on FigShare
642 (https://doi.org/10.6084/m9.figshare.20310522). Remaining tissues from representative samples
643 will be accessioned with the Alabama Museum of Natural History Entomology Collection.
644
645 **References:**

646 **Ahrens, C. W., P. D. Rymer, A. Stow, J. Bragg, S. Dillon, K. D. L. Umbers, and R. Y.**
647 **Dudaniec**. **2018**. The search for loci under selection: trends, biases and progress. Mol. Ecol.
648 27: 1342–1356.
649 **Andrews, S. 2010.** FastQC: A quality control tool for high throughput sequence data
650 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
651 **Booker, T. R., S. Yeaman, and M. C. Whitlock**. **2020**. Variation in recombination rate affects
652 detection of outliers in genome scans under neutrality. Mol. Ecol. 29: 4274–4279.
653 **Broad Institute.** 2019. http://broadinstitute.github.io/picard/
654 **Browning, B. L., Y. Zhou, and S. R. Browning**. **2018**. A one-penny imputed genome from
655 next-generation reference panels. Am. J. Hum. Genet. 103: 338–348.
656 **Bushnell, B.** 2020, BBMap. http://sourceforge.net/ projects/bbmap/.
657 **Cameron, S. A., H. M. Hines, and P. H. Williams**. **2007**. A comprehensive phylogeny of the
658 bumble bees (*Bombus*). Biol. J. Linn. Soc. 91: 161–188.
659 **Cameron, S. A., and B. M. Sadd**. **2020**. Global trends in bumble bee health. Annu. Rev.
660 Entomol. 65: 209–232.
661 **Capblancq, T., and B. R. Forester**. **2021**. Redundancy analysis: a swiss army knife for
662 landscape genomics. Methods Ecol. Evol. 2021: 1–12.
663 **Caye, K., B. Jumentier, J. Lepeule, and O. François**. **2019**. LFMM 2: fast and accurate

664 inference of gene-environment associations in genome-wide studies. Mol. Biol. Evol. 36:
665 852–860.
666 **Cayuela, H., Y. Dorant, C. Mérot, M. Laporte, E. Normandeau, S. Gagnon-Harvey, M.**
667 **Clément, P. Sirois, and L. Bernatchez**. **2021**. Thermal adaptation rather than demographic
668 history drives genetic structure inferred by copy number variants in a marine fish. Mol.
669 Ecol. 1–18.
670 **Cheng, S., Y. Park, J. D. Kurleto, M. Jeon, K. Zinn, J. W. Thornton, and E. Özkan**. **2019**.
671 Family of neural wiring receptors in bilaterians defined by phylogenetic, biochemical, and
672 structural evidence. Proc. Natl. Acad. Sci.116: 9837–9842.
673 **Cheviron, Z. A., and R. T. Brumfield**. **2012**. Genomic insights into adaptation to high-altitude
674 environments. Heredity. 108: 354–361.
675 **Christmas, M. J., J. C. Jones, A. Olsson, O. Wallerman, I. Bunikis, M. Kierczak, V. Peona,**
676 **K. M. Whitley, T. Larva, A. Suh, N. E. Miller-Struttmann, J. C. Geib, and M. T.**
677 **Webster**. **2021**. Genetic barriers to historical gene flow between cryptic species of alpine
678 bumblebees revealed by comparative population genomics. Mol. Biol. Evol. 38: 3126-3143
679 **Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and**
680 **D. M. Ruden**. **2012**. A program for annotating and predicting the effects of single
681 nucleotide polymorphisms, SnpEff. Fly (Austin). 6: 80–92.
682 **Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E.**
683 **Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin**. **2011**.
684 The variant call format and VCFtools. Bioinformatics. 27: 2156–2158.
685 **Diamond, S. E., and L. D. Chick**. **2018**. The Janus of macrophysiology: Stronger effects of
686 evolutionary history, but weaker effects of climate on upper thermal limits are reversed for

lower thermal limits in ants. Curr. Zool. 64: 223–230.

Dillon, M. E. 2006. Into thin air: physiology and evolution of alpine insects. Integr. Comp. Biol. 46: 49–61.

Dillon, M. E., and R. Dudley. **2014**. Surpassing Mt Everest - extreme flight performance of alpine bumblebees. Biol. Lett. 10.

Dillon, M. E., and J. D. Lozier. **2019**. Adaptation to the abiotic environment in insects: the influence of variability on ecophysiology and evolutionary genomics. Curr. Opin. Insect Sci. 36: 131–139.

Dorant, Y., H. Cayuela, K. Wellband, M. Laporte, Q. Rougemont, C. Mérot, E. Normandeau, R. Rochette, and L. Bernatchez. **2020**. Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. Mol. Ecol. 1–18.

Eckert, A. J., J. Van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. González-Martínez, and D. B. Neale. **2010**. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda L.*, Pinaceae). Genetics. 185: 969–982.

Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. 29: 51–63.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf. **2012**. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature. 491: 756–760.

Ferveur, J. F., J. Cortot, K. Rihani, M. Cobb, and C. Everaerts. **2018**. Desiccation resistance: effect of cuticular hydrocarbons and water content in *Drosophila melanogaster*

adults. PeerJ. 2018: 1–23.

Fick, S. E., and R. J. Hijmans. **2017**. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37: 4302–4315.

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. **2018**. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol. Ecol. 27: 2215–2233.

Frichot, E., and O. François. **2015**. LEA: an R package for landscape and ecological association studies. Methods Ecol. Evol. 6: 925–929.

Fuentes-Pardo, A. P., and D. E. Ruzzante. **2017**. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol. Ecol. 26: 5369–5406.

Gain, C., and O. François. **2021**. LEA 3: factor models in population genetics and ecological genomics with R. Mol. Ecol. Resour. 1–37.

Garrison, E., and G. Marth. **2012**. Haplotype-based variant detection from short-read sequencing. arXiv. 1–9.

Ghisbain, G., J. D. Lozier, S. R. Rahman, B. D. Ezray, L. Tian, J. M. Ulmer, S. D. Heraghty, J. P. Strange, P. Rasmont, and H. M. Hines. **2020**. Substantial genetic divergence and lack of recent gene flow support cryptic speciation in a colour polymorphic bumble bee (*Bombus bifarius*) species complex. Syst. Entomol. 45: 635–652.

Grote, S. 2020. GOfuncR: Gene ontology enrichment using FUNC. R package version 1.8.0. https://git.bioconductor.org/packages/GOfuncR

Hahne, F., R. Ivanek. 2016. Visualizing genomic data using gviz and bioconductor. Methods Mol Biol. 1418:335-51.

Harrison, J. F., K. J. Greenlee, and W. C. E. P. Verberk. **2018**. Functional hypoxia in insects: definition, assessment, and consequences for physiology, ecology, and evolution. Annu. Rev. Entomol. 63: 303–325.

Hartke, J., A. Waldvogel, P. P. Sprenger, T. Schmitt, F. Menzel, M. Pfenninger, and B. Feldmeyer. **2020**. Little parallelism in genomic signatures of local adaptation in two sympatric, cryptic sister species. J. Evol. Biol. 1–16.

Heinrich, B. 1976. Heat exchange in relation to blood flow between thorax and abdomen in bumblebees. J. Exp. Biol. 64: 561–585.

Heinrich, B. 2004. Bumblebee Economics. Harvard University Press, Cambridge, MA

Heinrich, B., and A. E. Kammer. **1973**. Activation of the fibrillar muscles in the bumblebee during warm up, stabilization of thoracic temperature and flight. J. Exp. Biol. 58: 677–688.

Heraghty, S. D., J. M. Sutton, M. L. Pimsler, J. L. Fierst, J. P. Strange, and J. D. Lozier. **2020**. De novo genome assemblies for three north american bumble bee species: *Bombus bifarius*, *Bombus vancouverensis*, and *Bombus vosnesenskii*. Genes|Genomes|Genetics. 10: 2585–2592.

Hijmans, R. J. 2021. Raster: geographic data analysis and modeling. R package version 3.4-13. https://CRAN.R-project.org/package=raster

Hines, H. M. 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). Syst. Biol. 57: 58–75.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. **2016**. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am. Nat. 188: 379–397.

Huang, H., and G. G. Haddad. **2007**. *Drosophila* dMRP4 regulates responsiveness to O2

deprivation and development under hypoxia. Physiol. Genomics. 29: 260–266.

Igsf, D., C. Surface, K. P. Menon, H. J. Bellen, K. C. Garcia, K. Zinn, K. P. Menon, S. Nagarkar-jaiswal, P. Lee, and M. Jeon. **2015**. Control of synaptic connectivity by a network of article control of synaptic connectivity by a network of *Drosophila* IgSF cell surface proteins. 1770–1782.

Iwamoto, H. 2009. Evidence for unique structural change of thin filaments upon calcium activation of insect flight muscle. J. Mol. Biol. 390: 99–111.

Jackson, J. M., M. L. Pimsler, K. J. Oyen, J. B. Koch-Uhuad, J. D. Herndon, J. P. Strange, M. E. Dillon, and J. D. Lozier. **2018**. Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. Mol. Ecol. 27: 2926–2942.

Jackson, J. M., M. L. Pimsler, K. J. Oyen, J. P. Strange, M. E. Dillon, and J. D. Lozier. **2020**. Local adaptation across a complex bioclimatic landscape in two montane bumble bee species. Mol. Ecol. 29: 920– 939.

de Jong, M. J., J. F. de Jong, A. R. Hoelzel, and A. Janke. **2021**. SambaR: an R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. Mol. Ecol. Resour. 21: 1369–1379.

Joost, S., A. Bonin, M. W. Bruford, L. Després, C. Conord, G. Erhardt, and P. Taberlet. **2007**. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol. Ecol. 16: 3955–3969.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers,

779 **C. D. Jiggins, and R. H. Ffrench-Constant**. **2011**. Chromosomal rearrangements maintain
780 a polymorphic supergene controlling butterfly mimicry. Nature. 477: 203–206.
781 **Keller, I., J. M. Alexander, R. Holderegger, and P. J. Edwards**. **2013**. Widespread
782 phenotypic and genetic divergence along altitudinal gradients in animals. J. Evol. Biol. 26:
783 2527–2543.
784 **Kim, Y., R. Nielson. 2004**. Linkage disequilibrium as a signature of selective sweeps. Genetics
785 167: 1513-1524
786 **Kraut, R., K. Menon, and K. Zinn**. **2001**. A gain-of-function screen for genes controlling
787 motor axon guidance and synaptogenesis in *Drosophila*. Curr. Biol. 11: 417–430.
788 **Krupp, J. J., K. Nayal, A. Wong, J. G. Millar, and J. D. Levine**. **2020**. Desiccation resistance
789 is an adaptive life-history trait dependent upon cuticular hydrocarbons, and influenced by
790 mating status and temperature in *D. melanogaster*. J. Insect Physiol. 121: 103990.
791 **De la Torre, A. R., B. Wilhite, D. B. Neale, and T. Slotte**. **2019**. Environmental genome-wide
792 association reveals climate adaptation is shaped by subtle to moderate allele frequency
793 shifts in loblolly pine. Genome Biol. Evol. 11: 2976–2989.
794 **Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan,**
795 **and V. J. Carey**. **2013**. Software for computing and annotating genomic ranges. PLoS
796 Comput. Biol. 9: 1–10.
797 **Lee, C. K. F., P. H. Williams, and R. G. Pearson**. **2019**. Climate change vulnerability higher in
798 arctic than alpine bumblebees. Front. Biogeogr. 11: 0–9.
799 **Legendre, P., J. Oksanen, and C. J. F. ter Braak**. **2011**. Testing the significance of canonical
800 axes in redundancy analysis. Methods Ecol. Evol. 2: 269–277.
801 **Li, H., and R. Durbin**. **2009**. Fast and accurate short read alignment with Burrows-Wheeler

802 transform. Bioinformatics. 25: 1754–1760.
803 **Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis,**
804 **and R. Durbin**. **2009**. The sequence alignment/map format and SAMtools. Bioinformatics.
805 25: 2078–2079.
806 **Li, H., A. Watson, A. Olechwier, M. Anaya, S. K. Sorooshyari, D. P. Harnett, H. K. (Peter)**
807 **Lee, J. Vielmetter, M. A. Fares, K. C. Garcia, E. Özkan, J. P. Labrador, and K. Zinn**.
808 **2017**. Deconstruction of the beaten path-sidestep interaction network provides insights into
809 neuromuscular system development. Elife. 6: 1–24.
810 **Lim, M. C. W., K. Bi, C. C. Witt, C. H. Graham, and L. M. Dávalos**. **2021**. Pervasive
811 genomic signatures of local adaptation to altitude across highland specialist andean
812 hummingbird Populations. J. Hered. 112: 229–240.
813 **Liu, Y., H. Zhao, Q. Luo, Y. Yang, G. Zhang, Z. Zhou, M. Naeem, and J. An**. **2020**. De novo
814 transcriptomic and metabolomic analyses reveal the ecological adaptation of high-altitude
815 *Bombus pyrosoma*. Insects. 11: 1–14.
816 **Lowry, D. B., S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, and A.**
817 **Storfer**. **2017**. Breaking RAD: an evaluation of the utility of restriction site-associated
818 DNA sequencing for genome scans of adaptation. Mol. Ecol. Resour. 17: 142–152.
819 **Lozier, J. D., Z. M. Parsons, L. Rachoki, J. M. Jackson, M. L. Pimsler, K. J. Oyen, J.**
820 **Strange, and M. E. Dillon**. **2021**. Divergence in body mass, wing loading, and population
821 structure reveals species-specific and potentially adaptive trait variation across elevations in
822 montane bumble bees. Insect Syst. Divers. 5: 1-15
823 **Lozier, J. D., J. P. Strange, and J. B. Koch**. **2013**. Landscape heterogeneity predicts gene flow
824 in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae).

Conserv. Genet. 14: 1099–1110.

Luo, L., Z. zheng Tang, S. D. Schoville, and J. Zhu. **2021**. A comprehensive analysis comparing linear and generalized linear models in detecting adaptive SNPs. Mol. Ecol. Resour. 21: 733–744.

Mahmoud, M., N. Gobet, D. I. Cruz-Dávalos, N. Mounier, C. Dessimoz, and F. J. Sedlazeck. **2019**. Structural variant calling: the long and the short of it. Genome Biol. 20: 1–14.

Manel, S., S. Joost, B. K. Epperson, R. Holderegger, A. Storfer, M. S. Rosenberg, K. T. Scribner, A. Bonin, and M. J. Fortin. **2010**. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. 19: 3760–3772.

Manenti, T., T. R. Cunha, J. G. Sørensen, and V. Loeschcke. **2018**. How much starvation, desiccation and oxygen depletion can *Drosophila melanogaster* tolerate before its upper thermal limits are affected? J. Insect Physiol. 111: 1–7.

Martinet, B., S. Dellicour, G. Ghisbain, K. Przybyla, E. Zambra, T. Lecocq, M. Boustani, R. Baghirov, D. Michez, and P. Rasmont. **2021**. Global effects of extreme temperatures on wild bumblebees. Conserv. Biol. 35: 1507–1518.

McCulloch, G. A., J. Guhlin, L. Dutoit, T. W. R. Harrop, P. K. Dearden, and J. M. Waters. **2021**. Genomic signatures of parallel alpine adaptation in recently evolved flightless insects. Mol. Ecol. 6677–6686.

Montejo-Kovacevich, G., J. E. Smith, J. I. Meier, C. N. Bacquet, E. Whiltshire-Romero, N. J. Nadeau, and C. D. Jiggins. **2019**. Altitude and life-history shape the evolution of *Heliconius* wings. Evolution (N.Y). 1–15.

Nguyen, A. D., K. DeNovellis, S. Resendez, J. D. Pustilnik, N. J. Gotelli, J. D. Parker, and S. H. Cahan. **2017**. Effects of desiccation and starvation on thermal tolerance and the heat-shock response in forest ants. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 187: 1107–1116.

O'Donnell, M.S., and D. A. Ignizio, D.A., 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691

Orr, H. A. 2005. The genetic theory of adaptation: A brief history. Nat. Rev. Genet. 6: 119–127.

Orr, M. C., A. C. Hughes, D. Chesters, J. Pickering, C. D. Zhu, and J. S. Ascher. **2020**. Global Patterns and Drivers of Bee Distribution. Curr. Biol. 31: 451-458.e4.

Overgaard, J., and H. A. MacMillan. **2017**. The Integrative Physiology of Insect Chill Tolerance. Annu. Rev. Physiol. 79: 187–208.

Oyen, K. J., S. Giri, and M. E. Dillon. **2016**. Altitudinal variation in bumble bee (*Bombus*) critical thermal limits. J. Therm. Biol. 59: 52–57.

Papadopulos, A. S. T., J. Igea, L. T. Dunning, O. G. Osborne, X. Quan, J. Pellicer, C. Turnbull, I. Hutton, W. J. Baker, R. K. Butlin, and V. Savolainen. **2019**. Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in *Howea*. Evolution. 73: 1986–1995.

Pimsler, M. L., K. J. Oyen, J. D. Herndon, J. M. Jackson, J. P. Strange, M. E. Dillon, and J. D. Lozier. **2020**. Biogeographic parallels in thermal tolerance and gene expression variation under temperature stress in a widespread bumble bee. Sci. Rep. 10: 1–11.

Quilodrán, C. S., K. Ruegg, A. T. Sendell-Price, E. C. Anderson, T. Coulson, and S. M.

Clegg. **2020**. The multiple population genetic and demographic routes to islands of genomic divergence. Methods Ecol. Evol. 11: 6–21.

Rahbek, C., M. K. Borregaard, R. K. Colwell, B. Dalsgaard, B. G. Holt, N. Morueta-Holme, D. Nogues-Bravo, R. J. Whittaker, and J. Fjeldså. **2019**. Humboldt's enigma: What causes global patterns of mountain biodiversity? Science. 365: 1108–1113.

Rausch, T., T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, and J. O. Korbel. **2012**. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 28: 333–339.

Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A. F. Noor, B. Mehlig, and A. M. Westram. **2017**. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30: 1450–1477.

Revelle, W. 2020. Psych: procedures for personality and psychological research. https://cran.r-project.org/web/packages/psych/index.html

Sadd, B. M., S. M. Barribeau, G. Bloch, D. C. de Graaf, P. Dearden, C. G. Elsik, J. Gadau, C. J. P. Grimmelikhuijzen, M. Hasselmann, J. D. Lozier, H. M. Robertson, G. Smagghe, E. Stolle, M. Van Vaerenbergh, R. M. Waterhouse, E. Bornberg-Bauer, S. Klasberg, A. K. Bennett, F. Câmara, R. Guigó, K. Hoff, M. Mariotti, M. Munoz-Torres, T. Murphy, D. Santesmasses, G. V. Amdam, M. Beckers, M. Beye, M. Biewer, M. M. G. Bitondi, M. L. Blaxter, A. F. G. Bourke, M. J. F. Brown, S. D. Buechel, R. Cameron, K. Cappelle, J. C. Carolan, O. Christiaens, K. L. Ciborowski, D. F. Clarke, T. J. Colgan, D. H. Collins, A. G. Cridge, T. Dalmay, S. Dreier, L. du Plessis, E. Duncan, S. Erler, J. Evans, T. Falcon, K. Flores, F. C. P. Freitas, T. Fuchikawa, T. Gempe, K. Hartfelder, F. Hauser, S. Helbing, F. C. Humann, F. Irvine, L. S. Jermiin,

C. E. Johnson, R. M. Johnson, A. K. Jones, T. Kadowaki, J. H. Kidner, V. Koch, A. Köhler, F. B. Kraus, H. M. G. Lattorff, M. Leask, G. A. Lockett, E. B. Mallon, D. S. M. Antonio, M. Marxer, I. Meeus, R. F. A. Moritz, A. Nair, K. Näpflin, I. Nissen, J. Niu, F. M. F. Nunes, J. G. Oakeshott, A. Osborne, M. Otte, D. G. Pinheiro, N. Rossié, O. Rueppell, C. G. Santos, R. Schmid-Hempel, B. D. Schmitt, C. Schulte, Z. L. P. Simões, M. P. M. Soares, L. Swevers, E. C. Winnebeck, F. Wolschin, N. Yu, E. M. Zdobnov, P. K. Aqrawi, K. P. Blankenburg, M. Coyle, L. Francisco, A. G. Hernandez, M. Holder, M. E. Hudson, L. R. Jackson, J. Jayaseelan, V. Joshi, C. Kovar, S. L. Lee, R. Mata, T. Mathew, I. F. Newsham, R. Ngo, G. Okwuonu, C. Pham, L. L. Pu, N. Saada, J. Santibanez, D. N. Simmons, R. Thornton, A. Venkat, K. K. O. Walden, Y. Q. Wu, G. Debyser, B. Devreese, C. Asher, J. Blommaert, A. D. Chipman, L. Chittka, B. Fouks, J. Liu, M. P. O'Neill, S. Sumner, D. Puiu, J. Qu, S. L. Salzberg, S. E. Scherer, D. M. Muzny, S. Richards, G. E. Robinson, R. A. Gibbs, P. Schmid-Hempel, and K. C. Worley. **2015**. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16: 1–31.

Savolainen, O., M. Lascoux, and J. Merilä. **2013**. Ecological genomics of local adaptation. Nat. Rev. Genet. 14: 807–820.

Schnorrer, F., C. Schönbauer, C. C. H. Langer, G. Dietzl, M. Novatchkova, K. Schernhuber, M. Fellner, A. Azaryan, M. Radolf, A. Stark, K. Keleman, and B. J. Dickson. **2010**. Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. Nature. 464: 287–291.

Seamus, J., C. Bak, M. Kuhlmann, and C. Grønkjær. **2018**. Cold exposure causes cell death by depolarization- mediated Ca 2 + overload in a chill-susceptible insect. 115: 9737–9744.

Sinclair, B. J., L. V. Ferguson, G. Salehipour-Shirazi, and H. A. Macmillan. **2013**. Cross-tolerance and cross-talk in the cold: relating low temperatures to desiccation and immune stress in insects. Integr. Comp. Biol. 53: 545–556.

Stolle, E., L. Wilfert, R. Schmid-Hempel, P. Schmid-Hempel, M. Kube, R. Reinhardt, and R. F. A. Moritz. **2011**. A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. BMC Genomics. 12.

Storey, J. D., A. J. Bass, A. Dabney and D. Robinson. 2020. qvalue: q-value estimation for false discovery rate control. R package version 2.20.0. http://github.com/jdstorey/qvalue

Storfer, A., A. Patton, and A. K. Fraik. **2018**. Navigating the interface between landscape genetics and landscape genomics. Front. Genet. 9: 1–14.

Sun, C., J. Huang, Y. Wang, X. Zhao, L. Su, G. W. C. Thomas, M. Zhao, X. Zhang, I. Jungreis, M. Kellis, S. Vicario, I. V Sharakhov, S. M. Bondarenko, M. Hasselmann, C. N. Kim, B. Paten, L. Penso-Dolfin, L. Wang, Y. Chang, Q. Gao, L. Ma, L. Ma, Z. Zhang, H. Zhang, H. Zhang, L. Ruzzante, H. M. Robertson, Y. Zhu, Y. Liu, H. Yang, L. Ding, Q. Wang, D. Ma, W. Xu, C. Liang, M. W. Itgen, L. Mee, G. Cao, Z. Zhang, B. M. Sadd, M. Hahn, S. Schaack, S. M. Barribeau, P. H. Williams, R. M. Waterhouse, and R. L. Mueller. **2020**. Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Mol. Biol. Evol. 38: 486–501.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc. **2011**. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One. 6.

Taylor, R. S., E. L. Jensen, D. W. Coltman, A. D. Foote, and S. Lamichhaney. **2021**. Seeing

the whole picture: What molecular ecology is gaining from whole genomes. Mol. Ecol. 30: 5917–5922.

Wallberg, A., M. T. Webster, M. Hasselmann, A. Wallberg, and C. Scho. **2017**. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. PLoS Genet. 1–30.

Walsh, A. T., D. A. Triant, J. J. Le Tourneau, and C. G. Elsik. **2021**. Hymenoptera genome database : new genomes and annotation datasets for improved go enrichment and orthologue analyses. Nucleic Acid Res. 50: D1032-D1039

Wang, G., and M. E. Dillon. **2014**. Recent geographic convergence in diurnal and annual temperature cycling flattens global thermal profiles. Nat. Clim. Chang. 4: 988–992.

Wellenreuther, M., C. Mérot, E. Berdan, and L. Bernatchez. **2019**. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Mol. Ecol. 28: 1203–1209.

Whitlock, M. C., and K. E. Lotterhos. **2015**. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. Am. Nat. 186: S24–S36.

Williams, P. H., J. M. Lobo, and A. S. Meseguer. **2018**. Bumblebees take the high road: climatically integrative biogeography shows that escape from Tibet, not Tibetan uplift, is associated with divergences of present-day *Mendacibombus*. Ecography (Cop.). 41: 461–477.

Wittkopp, P. J., and G. Kalay. **2012**. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. 13: 59–69.

Woodard, S. H. 2017. Bumble bee ecophysiology: integrating the changing environment and the

961 organism. Curr. Opin. Insect Sci. 22: 101–108.
962 **Yadav, S., A. Stow, and R. Y. Dudaniec**. **2020**. Microgeographic adaptation corresponds with
963 elevational distributions of congeneric montane grasshoppers. Mol. Ecol. 481–498.
964 **Zhang, C., S. S. Dong, J. Y. Xu, W. M. He, and T. L. Yang**. **2019**. PopLDdecay: a fast and
965 effective tool for linkage disequilibrium decay analysis based on variant call format files.
966 Bioinformatics. 35: 1786–1788.
967 **Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir**. **2012**. A high-
968 performance computing toolset for relatedness and principal component analysis of SNP
969 data. Bioinformatics. 28: 3326–3328.
970 **Zhong, Y., A. Shtineman-Kotler, L. Nguyen, K. G. Iliadi, G. L. Boulianne, and D. Rotin**.
971 **2011**. A splice isoform of DNedd4, DNedd4-Long, negatively regulates neuromuscular
972 synaptogenesis and viability in Drosophila. PLoS One. 6.
973
974 **Figure Captions**
975
976 **Figure 1**: Map of sampling locations (white circles) superimposed on a Maxent species
977 distribution model of the *Bombus vancouverensis* range generated following Cameron et al.
978 (2011). Areas of high suitability (darker colors) indicate the area inhabited by *Bombus*
979 *vancouverensis*. Inset image shows a picture of *Bombus vancouverensis*
980
981 **Figure 2**: Plot of -log(*q*-values) from OutFLANK across all scaffolds 100kb or lager (scaffolds
982 are of different sizes but x-axes are scaled here for graphical purposes, see Table 2 to find precise
983 SNP locations). The solid black line in each scaffold represents the *q*-value threshold of 0.01.

Point coded as "BOTH" are significant in both model outputs (LFMM2 and OutFLANK). Points coded as "LFMM" were significant in the LFMM2 model only, otherwise points above the black line were found to be significant only in OutFLANK.

Figure 3: Ordination plots of significant models ($p < 0.05$) from Redundancy analysis (RDA) showing the population structure of samples based on the different SV types; Deletions, Duplications, and Inversions. Large colored points show sample state of origin and small grey points show individual SVs.

Figure 4: Global F_{ST} and average within-population nucleotide diversity (π) averaged over 5 kb windows for outlier-dense focal scaffolds NW_02881786.1, NW_02881829.1, and NW_02881902.1 with outlier regions in the highlighted boxes

Figure 5: A) Smoothed average linkage across two largest outlier scaffolds (NW_02881829.1 and NW_02881902.1) (dark colored line) versus genome wide linkage (light colored line). B) Same as (A), but showing the average linkage per base-pair across largest outlier scaffolds (top) versus genome wide linkage (bottom). C) Smoothed average of outlier scaffolds (NW_02881829.1 and NW_02881902.1 in dark color) versus collection of 10 randomly scaffolds (Supp. Table S2) that did not contain outliers (light color). D) Diagram of positioning of *B. vancouverensis* scaffolds with outlier regions against LG11 of *Bombus terrestris*. Scaffold NW_022881902.1 was unplaced in the *B. terrestris* genome and is not shown. E) Relationship between F_{ST} and repeat content across the genome as well as across select scaffolds with high outlier density (NW_022881786.1, NW_022881829.1, and NW_022881902.1). All plots include

1007 linear regression models with *p*-values and R^2 values listed.

1008

Table 1: All cross validated SNPs from the output of LFMM2 and OutFLANK. Table shows: the NCBI Gene ID number and name for each gene (or the genes on either side of an intergenic region), the homologous gene in *D. melanogaster*, the number of cross validated SNPs found in each gene or intergenic region, the environmental variable associated with the SNP based on LFMM2, and the scaffold the SNP falls on.

Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117153469	E3 ubiquitin-protein ligase Nedd-4	Nedd4	1	BIO3	NW_022881760.1
LOC117153944-LOC117155827	zinc finger protein 100-like - uncharacterized	#N/A - #N/A	1	BIO3	NW_022881760.1
LOC117158937	heterogeneous nuclear ribonucleoprotein A3 homolog 2-like	#N/A	1	BIO3	NW_022881761.1
LOC117166763	calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1-like	Pde1c	1	BIO3	NW_022881765.1
LOC117154601	UTP--glucose-1-phosphate uridylyltransferase	UGP	2	BIO3	NW_022881773.1
LOC117154414	NADPH--cytochrome P450 reductase	Cpr	1	BIO3	NW_022881773.1
LOC117154436	uncharacterized	#N/A	2	BIO3, Elev	NW_022881773.1
LOC117154435	cilia- and flagella-associated protein 47-like	#N/A	2	BIO3, Elev	NW_022881773.1

Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117154605-LOC117154606	m7Gpppx diphosphatase - uncharacterized	CG2091 - #N/A	1	BIO3	NW_022881773.1
LOC117154606-LOC117154599	uncharacterized - macoilin	#N/A - CG30389	1	BIO3	NW_022881773.1
LOC117154426-LOC117154433	protein TAPT1 homolog - elongation of very long chain fatty acid protein 6-like	CG7218 - #N/A	1	BIO3	NW_022881773.1
LOC117154433-LOC117154432	elongation of very long chain fatty acid protein 6-like - elongation of very long chain fatty acid protein 6-like	#N/A - #N/A	2	BIO3	NW_022881773.1
LOC117155271	E3 ubiquitin-protein ligase Rnf220-like	CG4813	1	BIO3	NW_022881777.1
LOC117155885	mitochondrial sodium/calcium exchanger protein-like	CG14744	1	BIO3, BIO12	NW_022881780.1
LOC117155870-CHR_END	uncharacterized - chr end	2mit - #N/A	2	BIO3	NW_022881780.1
LOC117156425	GTPase-activating protein	RasGAP1	1	BIO3, Elev	NW_022881785.1
LOC117156434	glutamic acid-rich protein-like	Asph	2	BIO3,Elev	NW_022881785.1
LOC117156445	uncharacterized	#N/A	1	Elev	NW_022881785.1
LOC117156425-LOC117156434	GTPase-activating protein - glutamic acid-rich protein-like	RasGAP1 - Asph	1	BIO3,elev	NW_022881785.1
LOC117156686	DENN domain-containing protein 1A	CG18659	1	Elev	NW_022881786.1

Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117156494	nuclear pore complex protein Nup133	Nup133	1	Elev	NW_022881786.1
LOC117156535	multidrug resistance-associated protein 4-like	CG5789	6	Elev	NW_022881786.1
LOC117156538-LOC117156535	uncharacterized - multidrug resistance-associated protein 4-like	CG32206 - #N/A	1	elev	NW_022881786.1
LOC117157367	uncharacterized	#N/A	1	BIO3	NW_022881796.1
LOC117157569	protein sax-3-like	dpr20	36	BIO3,Elev	NW_022881829.1
LOC117157568	synaptogenesis protein syg-2-like	side-VI	66	BIO3,Elev	NW_022881829.1
LOC117157569-LOC117157568	protein sax-3-like - synaptogenesis protein syg-2-like	dpr20 - side-VI	263	BIO3,elev	NW_022881829.1
LOC117157921	UDP-glucuronosyltransferase 2B17-like	Ugt35C1	1	BIO3	NW_022881833.1
LOC117158648	tyrosine-protein kinase Drl	Drl-2	3	BIO3	NW_022881846.1
LOC117158660	protein sister of odd and bowel-like	CG4374	1	BIO3	NW_022881846.1
LOC117158593	cilia- and flagella-associated protein 20-like	Bug22	1	BIO3,Elev	NW_022881846.1
LOC117158711	UPF0489 protein C5orf22 homolog	MESR6	3	BIO3	NW_022881847.1

Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117158916	uncharacterized	#N/A	2	BIO3	NW_022881848.1
LOC117159416-LOC117159440	leucine-rich repeat-containing protein 24-like - uncharacterized	kek2 - #N/A	5	BIO3,elev	NW_022881861.1
LOC117160564	5-hydroxytryptamine receptor-like	RYa-R	1	BIO3	NW_022881886.1
LOC117160713	chaoptin-like	CG42346	1	BIO3	NW_022881888.1
LOC117160671	ATP-binding cassette sub-family G member 1	CG5853	1	BIO3	NW_022881888.1
LOC117160670-LOC117160662	ATP-binding cassette sub-family G member 1-like - lysine-specific demethylase 4c-like	CG9663 - kdm4b	2	BIO3	NW_022881888.1
LOC117160794	fatty acyl-CoA reductase 1-like	CG5065	1	BIO3	NW_022881895.1
LOC117160785-LOC117160792	uncharacterized - zinc finger protein 184-like	#N/A - #N/A	1	BIO3	NW_022881895.1
LOC117160792-LOC117160793	zinc finger protein 184-like - vicilin-like seed storage protein At2g18540	#N/A - #N/A	1	BIO3	NW_022881895.1
LOC117160794-LOC117160795	fatty acyl-CoA reductase 1-like - fatty acyl-CoA reductase 1-like	#N/A - #N/A	1	BIO3	NW_022881895.1
LOC117160795-CHR_END	fatty acyl-CoA reductase 1-like - chr end	#N/A - #N/A	2	BIO3	NW_022881895.1
LOC117161190	pro-resilin-like	Cpr50Cb	2	BIO3	NW_022881902.1

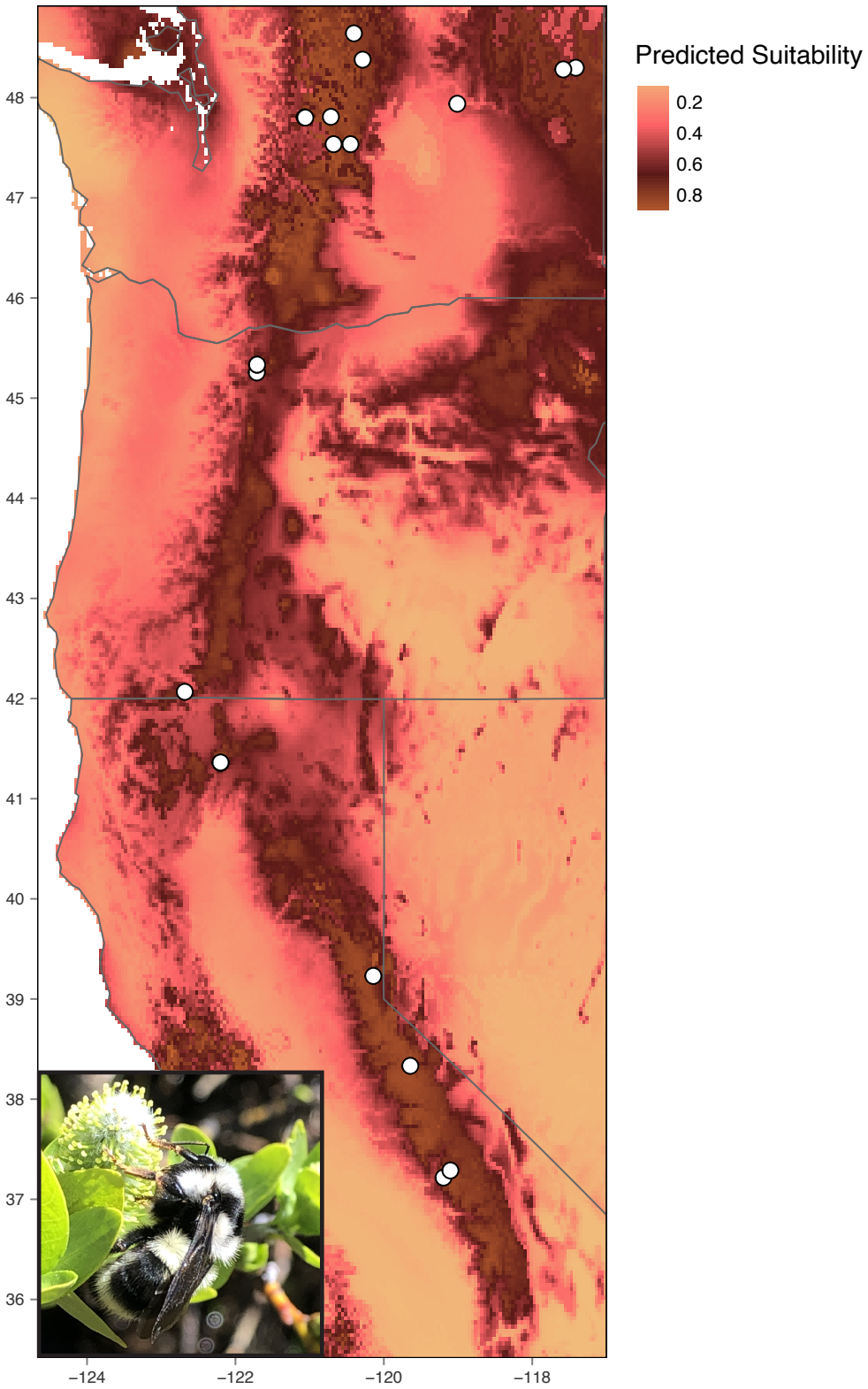
Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117161192	N66 matrix protein-like	Cpr50Cb	2	BIO3	NW_022881902.1
LOC117161197	uncharacterized	#N/A	3	BIO3	NW_022881902.1
LOC117161196	60S ribosomal protein L35	RpL35	1	BIO3	NW_022881902.1
LOC117161193	L-selectin	CG6055	1	BIO3	NW_022881902.1
LOC117161103	adenomatous polyposis coli protein-like	Apc	1	BIO3	NW_022881902.1
LOC117161064	ras-related protein Rab-11A	Rab11	1	BIO3	NW_022881902.1
LOC117161116	uncharacterized	CG13138	15	BIO3	NW_022881902.1
LOC117161115	low-density lipoprotein receptor-related protein 2	mgl	2	BIO3	NW_022881902.1
LOC117161235	transmembrane emp24 domain-containing protein bai	bai	1	BIO3	NW_022881902.1
LOC117161224	protein cordon-bleu-like	CG2841	1	BIO3	NW_022881902.1
LOC117161157	uncharacterized	beat-IIIc	5	BIO3,BIO12	NW_022881902.1
LOC117161181	phosphatidylinositol-binding clathrin assembly protein LAP	lap	2	BIO3	NW_022881902.1

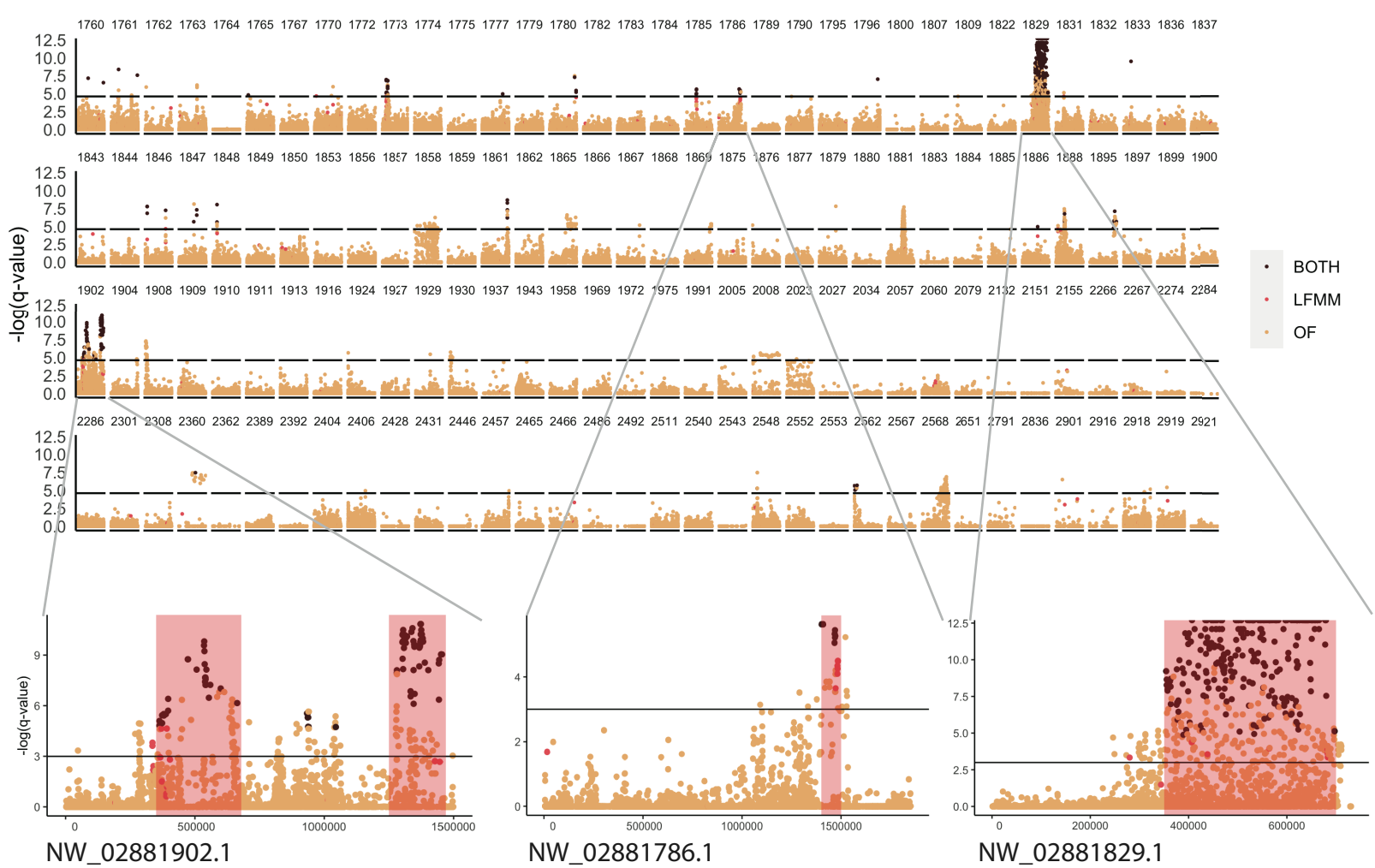
Gene ID	Gene Name	Fly Homolog	No. SNPS	Environmental Variable(s)	Scaffold
LOC117161180	xanthine dehydrogenase/oxidase-like	AOX3	36	BIO3	NW_022881902.1
LOC117161100	plasma membrane calcium-transporting ATPase 3	PMCA	36	BIO3,Elev	NW_022881902.1
LOC117161197-LOC117161189	uncharacterized - unconventional refolding RPB5 interaction-like protein	#N/A - uri	1	BIO3	NW_022881902.1
LOC117161194-LOC117161188	MIIP18 family protein galla-1 - alanine aminotransferase 1	galla-1 - CG1640	2	BIO3	NW_022881902.1
LOC117161103-LOC117161251	adenomatous polyposis colii protein-like - small nuclear ribonucleoprotein Sm D3	#N/A - SmD3	1	BIO3	NW_022881902.1
LOC117161081-LOC117161088	protein tramtrack, beta isoform-like - uncharacterized	rib - #N/A	3	BIO3	NW_022881902.1
LOC117161080-LOC117161076	synembryn-A - brain tumor protein-like	rica8a - mei-P26	2	BIO3	NW_022881902.1
LOC117162971-LOC117162978	probable serine/threonine-protein kinase MARK-A - uncharacterized	#N/A - #N/A	1	elev	NW_022882360.1
LOC117164862	mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	alpha-Man-Ia	1	BIO3	NW_022882548.1
CHR_START-LOC117165093	chr start - uncharacterized	#N/A - #N/A	1	BIO3	NW_022882562.1
LOC117165093-LOC117165099	uncharacterized - uncharacterized	#N/A - #N/A	1	BIO3	NW_022882562.1
LOC117165099-LOC117165101	uncharacterized - uncharacterized	#N/A - #N/A	2	BIO3	NW_022882562.1

Table 2: All outlier SVs identified from the Redundancy analysis (RDA). The table notes: the type of SV as either a deletion (Del) or inversion. (Inv), the scaffold the SV is located on, the start position of the SV, the NCBI gene ID and name of gene(s) spanned by the SV, and the total length of the SV in bp.

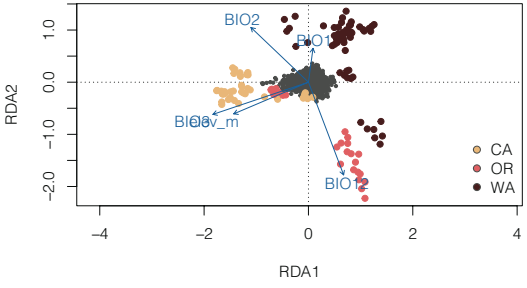
SV Type	Scaffold	Start	Location	Length
Del	NW_022881761.1	4656001	LOC117164506 (peptide-N(4)-(N-acetyl-beta-glucosaminyl) asparagine amidase)	65
Del	NW_022881772.1	3192657	LOC117154137 (uncharacterized loci)	10126
Del	NW_022881829.1	507180	intergenic	137
Del	NW_022881829.1	518415	intergenic	158
Del	NW_022881829.1	549109	intergenic	47
Del	NW_022881829.1	649945	LOC117157568 (synaptogenesis protein Syg-2-like)	65
Del	NW_022881829.1	661084	LOC117157568 (synaptogenesis protein Syg-2-like)	107
Del	NW_022881832.1	2130734	intergenic	249
Del	NW_022881861.1	2754083	intergenic	1462
Del	NW_022881862.1	2343632	LOC117159534 (innexin shaking-B)	112
Del	NW_022881865.1	581715	intergenic	52
Del	NW_022881877.1	854738	LOC117160152 (uncharacterized loci)	571
Del	NW_022881879.1	378524	LOC117160263 (RNA-binding protein Musashi homolog Rbp6)	94
Del	NW_022881881.1	715676	LOC117160393 (cadherin EGF LAG seven-pass G-type receptor 1-like)	69
Del*	NW_022881881.1	895567	multiple genes	385647
Del	NW_022881881.1	1165612	LOC117160315 (tyrosine-protein. Kinase Btk29A)	73
Del	NW_022881888.1	733080	LOC117160713 (calcium-binding mitochondrial carrier protein SCaMC-2)	281
Del	NW_022881991.1	365611	LOC117162117 (uncharacterized)	1133
Del	NW_022882286.1	2868480	LOC117162727 (cadherin-23)	867
Del	NW_022882406.1	6267958	LOC117163555 (sex determination protein fruitless)/LOC117163556 (uncharacterized)	126
Del	NW_022882540.1	1089009	multiple genes	23855
Del	NW_022882918.1	4984008	LOC117165908 (syntaxin-binding protein 5)	864
Inv	NW_022881784.1	372519	LOC117156304 (hemicentin-1-like)	136
Inv	NW_022881881.1	1233941	LOC117160315 (tyrosine-protein kinase Btk29A)	47273
Inv	NW_022882023.1	165412	intergenic	662

Footnote: * May represent an artefact in the reference genome or an issue with alignment due to large size of SV.

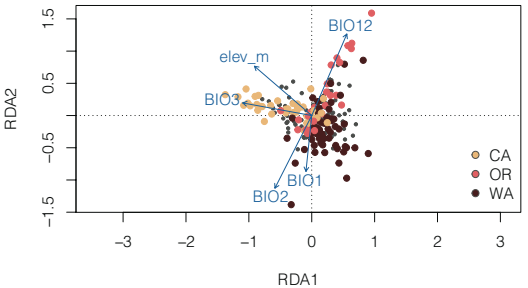

Predicted Suitability
0.2
0.4
0.6
0.8
48
47
46
45
44
43
42
41
40
39
38
37
36
−124
−122
−120
−118


-log(q-value)
BOTH
LFMM
OF
NW_02881902.1
NW_02881786.1
NW_02881829.1

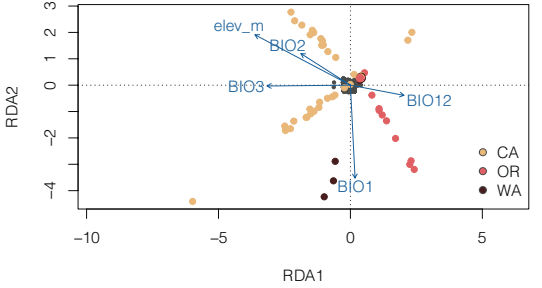
Deletions



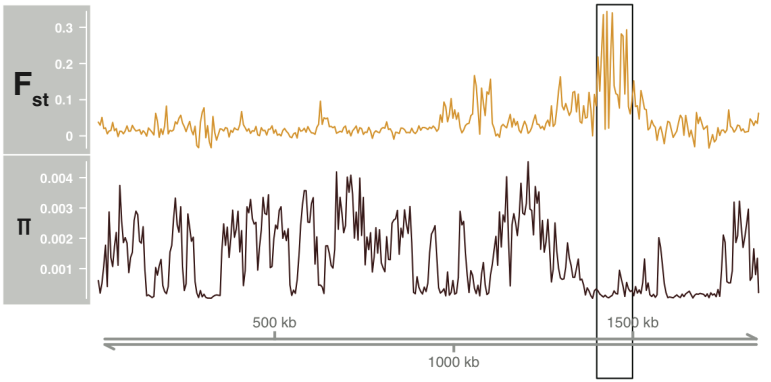
Duplications



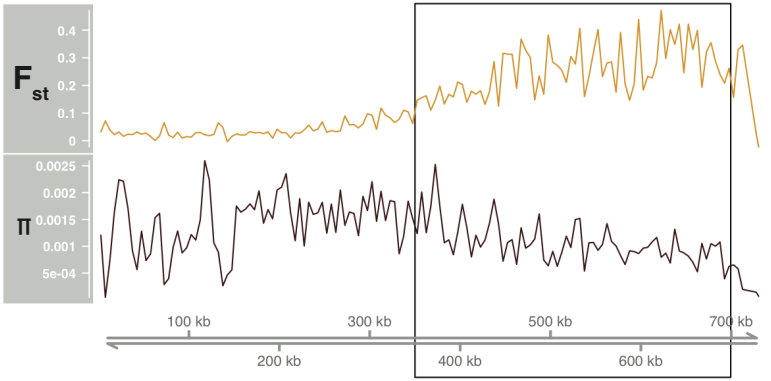
Inversions



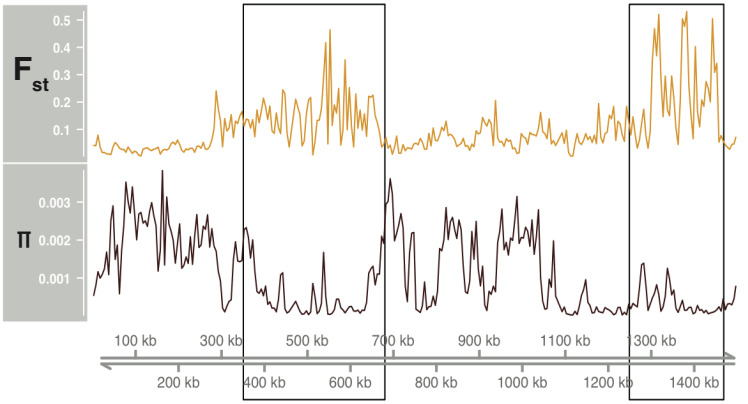
NW_022881786.1

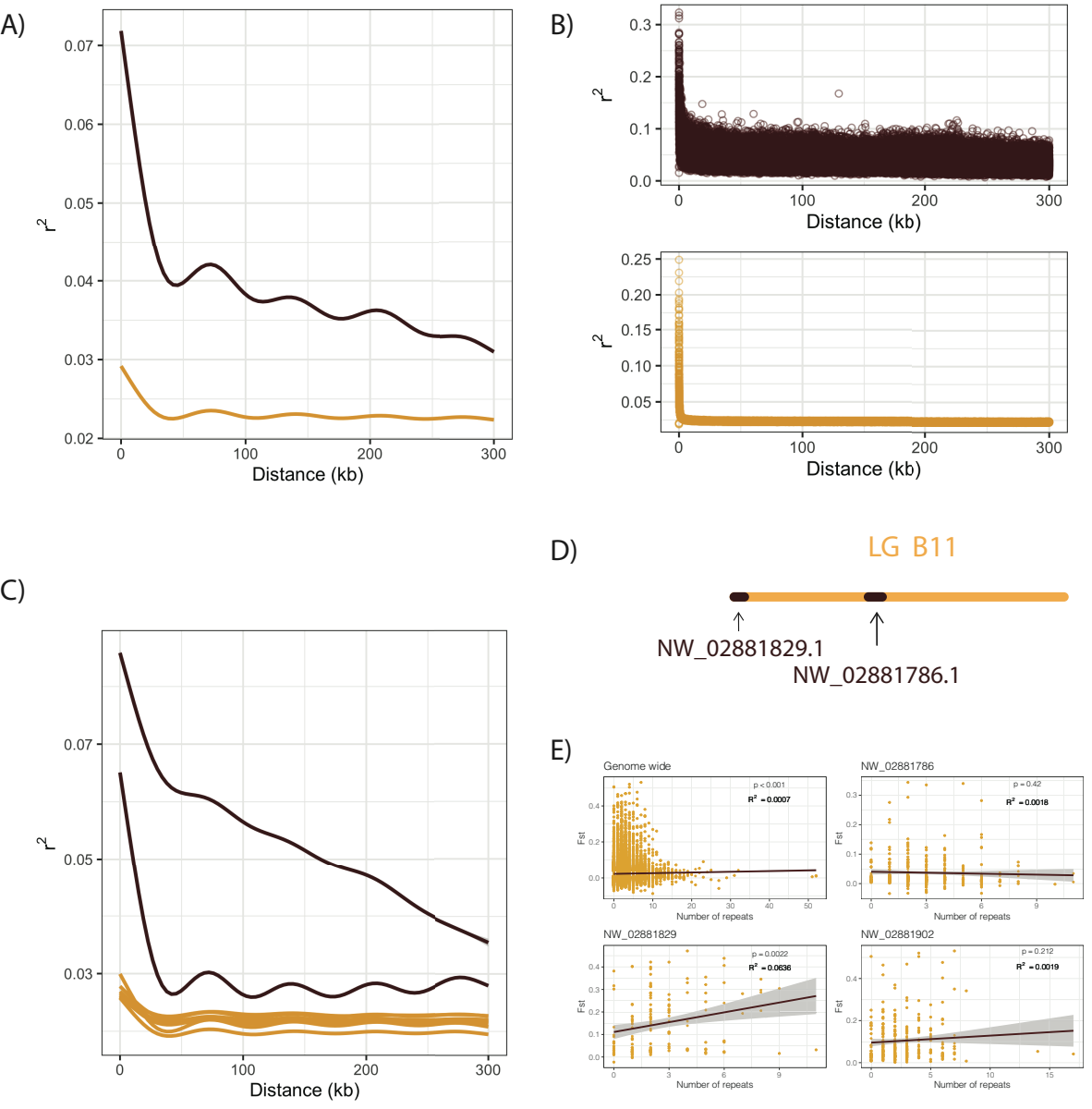


NW_022881819.1



NW_022881902.1




A)
B)
C)
D)
E)
r2
Distance (kb)
LG B11
NW_02881829.1
NW_02881786.1
Genome wide
NW_02881786
NW_02881829
NW_02881902
Fst
Number of repeats
p < 0.001
R² = 0.0007
p = 0.42
R² = 0.0018
p = 0.0022
R² = 0.0636
p = 0.212
R² = 0.0019