UC Berkeley
UC Berkeley Previously Published Works

Title

A chromosome-scale genome assembly and evaluation of mtDNA variation in the willow leaf beetle Chrysomela aeneicollis.

Permalink

Journal

Authors

Bracewell, Ryan
Stillman, Jonathon
Dahlhoff, Elizabeth
et al.

Publication Date

DOI

A chromosome-scale genome assembly and evaluation of mtDNA variation in the willow leaf beetle *Chrysomela aeneicollis*

Ryan R. Bracewell,[1,*] Jonathon H. Stillman,[1,2] Elizabeth P. Dahlhoff,[3] Elliott Smeds,[4] Kamalakar Chatla,[1] Doris Bachtrog,[1] Caroline Williams,[1] Nathan E. Rank[4]

[1]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Department of Biology, San Francisco State University, San Francisco, CA 94132, USA
[3]Department of Biology, Santa Clara University, Santa Clara, CA 95053, USA
[4]Department of Biology, Sonoma State University, Rohnert Park, CA 94928, USA

*Corresponding author: Department of Biology, Indiana University, 1001 E 3rd St, Bloomington, IN 47405, USA. Email: rbracewe@iu.edu

Abstract

The leaf beetle *Chrysomela aeneicollis* has a broad geographic range across Western North America but is restricted to cool habitats at high elevations along the west coast. Central California populations occur only at high altitudes (2,700–3,500 m) where they are limited by reduced oxygen supply and recent drought conditions that are associated with climate change. Here, we report a chromosome-scale genome assembly alongside a complete mitochondrial genome and characterize differences among mitochondrial genomes along a latitudinal gradient over which beetles show substantial population structure and adaptation to fluctuating temperatures. Our scaffolded genome assembly consists of 21 linkage groups; one of which we identified as the X chromosome based on female/male whole genome sequencing coverage and orthology with *Tribolium castaneum*. We identified repetitive sequences in the genome and found them to be broadly distributed across all linkage groups. Using a reference transcriptome, we annotated a total of 12,586 protein-coding genes. We also describe differences in putative secondary structures of mitochondrial RNA molecules, which may generate functional differences important in adaptation to harsh abiotic conditions. We document substitutions at mitochondrial tRNA molecules and substitutions and insertions in the 16S rRNA region that could affect intermolecular interactions with products from the nuclear genome. This first chromosome-level reference genome will enable genomic research in this important model organism for understanding the biological impacts of climate change on montane insects.

Keywords: genome assembly, mitochondria, Hi-C

Introduction

Breakthroughs in sequencing technology are now allowing us to produce high-quality chromosome-scale genome assemblies for organisms across the tree of life. Beetles (order Coleoptera) constitute an exceptionally large group with more described species than any other insect order, yet their genomic resources lag behind other arthropods (Hotaling *et al.* 2021). This is surprising given that Coleoptera contains numerous well-studied species of significant economic and ecological importance. By broadening genomic resources in Coleoptera, we will begin to understand what has made this group one of the most successful animal radiations on earth and will deepen our knowledge of how they integrate into natural and agricultural systems.

The leaf beetle *Chrysomela aeneicollis* (Fig. 1) is found in montane and coastal habitats across Western North America (Brown 1956; Dellicour *et al.* 2014). It occurs at high elevation (2,700–3,500 m) in the Sierra Nevada range in California, where it feeds on willows in bogs and along streams (Smiley and Rank 1986). Decades of ecological and evolutionary studies have focused on populations found at high elevation in the southern Sierra Nevada in

California (Smiley *et al.* 1985; Rank 1992; Dahlhoff and Rank 2000; Rank *et al.* 2007). These studies have revealed that Sierra Nevada populations of this insect are strongly dependent on annual fluctuations in temperature and snowpack (Dahlhoff *et al.* 2019; Roberts 2022) and show evidence of local adaptation to thermal microclimate and environmental oxygen (Rank 1992; McMillan *et al.* 2005; Millstein 2006; Rank *et al.* 2020). Allele frequencies of key metabolic enzyme loci fluctuate, in a putatively adaptive way, with changes in climate (Rank and Dahlhoff 2002; Dahlhoff *et al.* 2008). Furthermore, recent evidence suggests that relationships between the mitochondrial genotype, performance, and reproductive success in Sierra Nevada populations depend on the genotype at the nuclear enzyme locus *phosphoglucose isomerase* (Dahlhoff *et al.* 2019; Rank *et al.* 2020), providing a case study for the importance of mitonuclear epistasis in natural populations under field conditions (Camus 2020). The broader context for historical phylogeographic diversification of *C. aeneicollis* populations in Western North America has been described for both mitochondrial and nuclear loci (Dellicour *et al.* 2014). Together with extensive ecological data on host plant relationships and food web dynamics (Smiley *et al.* 1985; Rank 1994; Otto *et al.* 2008), these

studies provide a unique understanding of the natural history, physiology, and evolutionary ecology of an insect species in its natural context.

Here, we report both the nuclear and mitochondrial genome of *C. aeneicollis*. This is the first genome assembly for this species. We also provide careful annotations of transcribed protein-coding and RNA gene products of mitochondria. In addition, we assess differences in mitochondria of 12 individuals sampled from three high-altitude populations in the Eastern Sierra Nevada (Rock Creek, Bishop Creek, and Big Pine Creek). Finally, we identify potential secondary structure differences in mitochondrial rRNA and tRNA molecules in these populations and discuss the role these variants may play in mechanisms of local adaptation to environmental temperature and oxygen levels.

## Materials and methods

Initial assembly with 10× genomics

A single adult female beetle from Mosquito Flat in Rock Creek (37° 26′ 29.6″ N, 118° 44′ 48.3″ W; 3,067 m ASL) was collected in September 2017, kept alive in the laboratory until Feb 2018, and then frozen at −80°C until DNA extraction. High-molecular weight genomic DNA was extracted by staff of the University of California Davis DNA Technologies and Expression Analysis Core Laboratory (Davis, CA, USA). Blue Pippin size selection was used to retain fragments > 40 kb, and the resulting DNA was used to construct sequencing linked read libraries using 10× Genomics Technology. PippinHT size selection of the final libraries (350–650 bp) was used to maximize read 2 quality when sequenced using an Illumina HiSeqX10 at Novogene (Sacramento, CA USA). The resulting sequence data (150 bp PE) was assembled into a draft genome using Supernova v2.0.0 (Weisenfeld *et al.* 2017). Assemblies were made using 56× coverage (224M reads, following "optimal" amount of data recommendations) and 99× coverage (395M reads, all the sequence data). For each assembly, Supernova output was generated in three styles (Pseudohap, Megabubble, and Raw), though only the Pseudohap output, which generates a single record per scaffold, was used for subsequent analysis. We used BUSCO v.2.0 (Simão *et al.* 2015) on CyVerse with three references (Metazoa, Arthropoda, and Insecta_odb9) to determine which assembly was more complete.

Hi-C scaffolding

To further scaffold the Supernova-based assembly, we used chromatin conformation capture to generate Hi-C data (Lieberman-Aiden *et al.* 2009). Hi-C libraries were generated as outlined in (Bracewell *et al.* 2019) using a Dnase digestion method (Ramani *et al.* 2016) and a single female adult as input. The resulting DNA library was prepared using an Illumina TruSeq Nano library prep kit and was sequenced on a HiSeq 4000 with 100 bp PE reads. We used the Juicer pipeline (Durand *et al.* 2016b) to map Hi-C reads and generate contact maps based on 3D interactions. We then used the 3D-DNA pipeline (Dudchenko *et al.* 2017) to orient and place contigs. 3D-DNA output files were visualized and checked for accuracy using Juicebox (Durand *et al.* 2016a) with verification and modifications to scaffolding done using built-in tools. The final assembly was scaffolded together with 500 Ns between each contig.

Identifying the putative X chromosome scaffold

We used whole genome sequencing data of sexed individuals (described below in the mtDNA section) to identify the putative X chromosome scaffold. In male heterogametic taxa (most
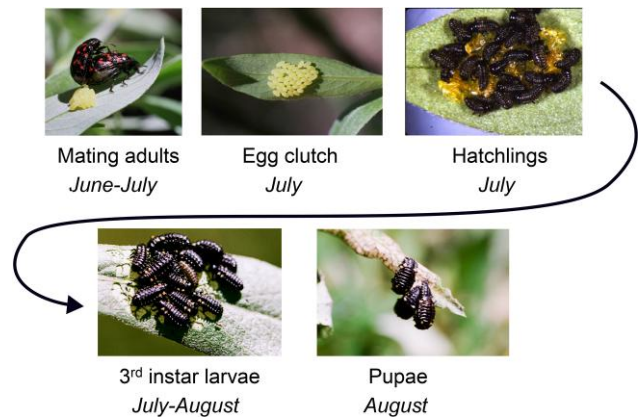
Fig. 1. The willow leaf beetle *C. aeneicollis* (Schaeffer), a natural model system for evolutionary, physiological, and ecological genomics. Life cycle stages of this univoltine insect, which completes larval development during midsummer.

Coleoptera), females are expected to have roughly double the sequencing coverage of a male over the X chromosome. Therefore, we mapped Illumina 150 bp PE reads to our draft genome using BWA MEM (Li and Durbin 2009) and compared normalized female/male coverage over all linkage groups using bedtools2 (Quinlan and Hall 2010).

Repetitive element identification and genome masking

We used EDTA (Ou *et al.* 2019) and RepeatModeler2 (Flynn *et al.* 2020), to identify repetitive sequences in the genome assembly. With these repeat libraries, we then masked the genome with RepeatMasker version 4.0.7 using the -no_is and -nolow flags. To characterize the proportion of sequence that was identified as repetitive and the genomic distribution of specific repetitive elements, we used bedtools2 (Quinlan and Hall 2010). We identified repeat-enriched regions of the genome as those areas where at least two consecutive genomic intervals (i.e. 100 kb region given 50 kb windows) were estimated to be at least double the genome-wide average.

Transcriptome data

To generate a transcriptome to aid in genome annotation (below), we used available RNA-seq data from whole bodies of third instar larvae from an experiment investigating insect performance and gene expression (Elmore 2021). We chose data from three larvae, one each from a different experimental treatment (A73, A220, and A223), in an attempt to build a diverse transcript library.

Genome annotation

To annotate the assembly, we used the repeat-masked genome and the MAKER2 annotation pipeline (Holt and Yandell 2011) to identify gene models. The ab initio gene predictors SNAP (Korf 2004) and AUGUSTUS (Stanke and Waack 2003) were used to guide the annotation with the protein set from *Tribolium castaneum*. We also cleaned RNA-seq raw reads using Seqyclean (https://github.com/ibest/seqyclean) and then used HISAT2 (Kim *et al.* 2019) and the -dta flag to align the reads to the draft genome. We used StringTie2 (Kovaka *et al.* 2019) to build a *C. aeneicollis* transcriptome for use as a guide in annotation. Input transcripts used for annotation were created from the transcriptome file using gffread. To compare our final MAKER2 annotation to the model
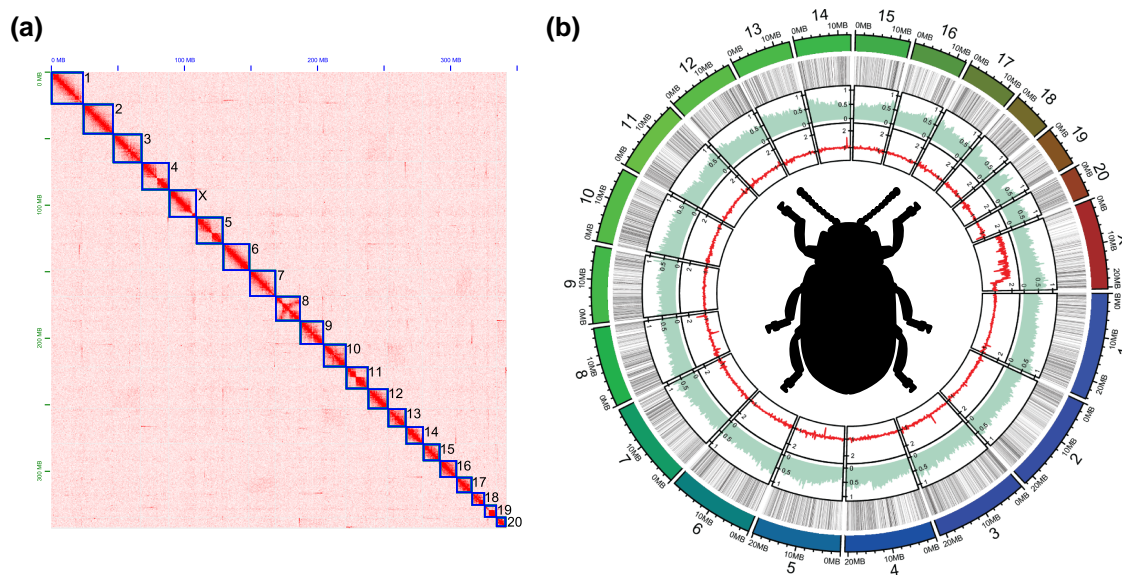
Fig. 2. Nuclear genome structure. a) Hi-C contact heatmap showing chromosome level scaffolding of the 10× assembly. Twenty-one total linkage groups were found (shown in blue boxes). The observed bowtie patterns in the contact heatmap for some scaffolds (e.g. 4, 5, and 11) are consistent with metacentric chromosomes. b) Circular plot showing the draft genome assembly and from the outermost track moving inwards a) the lengths of the 21 linkage groups, b) a heatmap showing low (white) to high (black) density of scaffolding Ns, c) the percentage of sequence identified as repetitive (50 kb windows), and d) the normalized female/male whole genome sequencing coverage (red).

beetle *T. castaneum* annotation, we used orthoDB (Kriventseva *et al.* 2018) and translated protein sequences from the two species (version Tcas_5.2) and compared them locally.

Final genome assembly cleaning

As a final cleaning step, we re-examined small sequences that were not incorporated into Hi-C scaffolds. These sequences were screened to determine if they were nontarget sequences (i.e. mtDNA, *Wolbachia*, or *Spiroplasma*) or contained annotated protein-coding genes. We also removed unscaffolded contigs that were not found to have any protein-coding genes.

mtDNA assembly and analysis

We used the mitochondrial genome program NOVOPLASTY to assemble mitochondrial genomes of 12 *C. aeneicollis* individuals, four each from Rock Creek, Bishop Creek, and Big Pine Creek. These individuals had been sampled as first instar larvae and reared to third instar under common garden conditions at 3,000 m ASL during summer 2015 (Smeds 2022). Before extraction, larvae were dissected, flash frozen in liquid nitrogen, and stored at −8°C until DNA extractions were performed in the laboratory of G. Lanzaro (University of California-Davis). Individual larval abdomens were placed in 2.0 mL microcentrifuge tubes along with 100 mL of a MagMAX PK enzyme mix (2 mL of 100 mg/mL proteinase K and 98 mL of PK buffer, both from Thermo Fisher) and a 3-mm-diameter steel bead, and then homogenized at 30 Hz for 30 s in a Qiagen Tissuelyser. Samples were then centrifuged at 10,000 g for 1 min in a benchtop microcentrifuge followed by 2 h incubation at 56°C in a shaking incubator. After incubation, samples were centrifuged and transferred to a 96-well plate with 100 µL of MagMAX DNA lysis buffer, and incubated at room temperature for 10 min. A solution containing 215 µL of the MagAttract mix (100 µL Buffer AL, 100 mL isopropanol, and 15 µL MagAttract Suspension) was added to each well, and the plate was loaded into a Qiagen BioSprint 96 extraction machine. DNA was extracted using BioSprint DNA Blood Kit reagents and BS96 DNA Tissue protocol. We also assembled the mitochondrial

genome for a closely related species using one *Chrysomela lapponica* individual collected in Fjellas, Norway (69° 18′ 15″N; 19° 31′ 03″E) in June 2012 and stored at −80°C until DNA extraction in 2019 using a Qiagen DNAEasy animal tissue kit.

We constructed genomic libraries for *C. aeneicollis* and *C. lapponica* using a Nextera DNA Flex kit with Nextera DNA CD indexes (Illumina). Each library was constructed with 500 ng of starting material. We quantified input DNA preps and finished libraries on a Qubit 3.0 Fluorometer using dsDNA high-sensitivity reagents and assessed the library quality using an Agilent 2100 Bioanalyzer with a dsDNA high-sensitivity chip. 150 bp PE sequencing was performed by Novogene (Davis, California) on an Illumina Hiseq X Ten, with twelve *C. aeneicollis* libraries sequenced on one lane and one *C. lapponica* library on a second lane.

We used Novoplasty 4.0 to assemble mitochondrial genomes. Before running Novoplasty, we used the perl script "filter_reads.pl" to extract reads with putative mitochondrial sequence. When we ran Novoplasty, we used *cytochrome oxidase I* sequence from a prior draft genome, which was also compared to Sanger sequences obtained from prior studies (Rank *et al.* 2020) as seed sequence and used default settings in the Novoplasty config.txt file for initial assembly. To annotate genomes of *C. aeneicollis* and *C. lapponica*, we used the program MITOS (http://mitos.bioinf.uni-leipzig.de/index.py). After initial automated annotation, we manually compared gene lengths and positions among both species and to a complete mitochondrial genome of the related species *Gonioctena intermedia* (Coleoptera: Chrysomelidae; accession number KX922881.1). Based on recommendations made by Cameron 2014, we extended protein-coding genes to their corresponding stop codons, allowing for more overlapping genes than the automated output provided. Once we completed these annotations, we validated nucleotide and amino acid sequences for 13 protein-coding genes and ribosomal RNA regions using BLAST and comparisons to relatives of *C. aeneicollis*. Annotations for tRNA coding genes were validated using the web-based tRNAscan-SE program (Chan and Lowe 2019). This application was also used to compare predicted secondary structure

Table 1. *C. aeneicollis* mtDNA and nuclear genome assembly statistics.

Linkage group	Length (bp)	Protein-coding genes	GC (%)	N (%)	Repetitive (%)
Scaf_01	24,271,041	678	35.0%	15.9%	42.8%
Scaf_02	22,353,350	632	34.9%	15.2%	41.7%
Scaf_03	21,438,272	654	34.7%	17.6%	43.1%
Scaf_04	20,712,704	613	35.2%	16.9%	44.7%
Scaf_05	20,149,410	580	35.3%	16.0%	43.6%
Scaf_06	19,848,268	632	35.4%	16.8%	43.4%
Scaf_07	19,643,986	510	34.7%	16.3%	42.5%
Scaf_08	18,176,930	552	35.0%	17.3%	41.4%
Scaf_09	17,506,527	606	34.9%	17.5%	41.1%
Scaf_10	17,186,144	427	35.0%	19.2%	42.8%
Scaf_11	16,482,901	535	35.2%	18.1%	44.2%
Scaf_12	15,131,411	462	35.4%	17.6%	45.5%
Scaf_13	13,599,833	355	34.8%	15.2%	48.2%
Scaf_14	12,890,779	421	35.2%	17.6%	44.1%
Scaf_15	12,471,685	414	35.4%	16.2%	40.1%
Scaf_16	12,249,352	388	35.9%	13.6%	43.8%
Scaf_17	11,871,507	384	35.7%	17.5%	44.9%
Scaf_18	9,236,171	347	34.7%	18.9%	42.1%
Scaf_19	9,108,583	268	35.9%	14.7%	46.0%
Scaf_20	7,025,367	202	36.2%	20.9%	49.2%
Scaf_X	20,422,211	663	35.2%	12.2%	40.7%
Unincorporated	12,174,703	2,263	35.8%	17.0%	42.1%
mtDNA	18,007	13	23.9%	0%	*

*not estimated.

configurations for variants at the tRNA-Leu2 and tRNA-Lys regions. Predicted secondary structures for rRNA variants were analyzed and visualized using the web-based Vienna RNA Websuite (Lorenz *et al.* 2011). Once the genome was annotated, we removed nucleotides corresponding to the AT-rich region from genomes before final alignment and comparison of genetic variants. The alignments of the transcribed region and annotations for the 12 sequenced individuals were made using the genomics software platform Geneious (version 9.1.8), which we also used to generate the VCF file.

To understand the evolution of mtDNA in *C. aeneicollis*, we inferred relationships by using maximum likelihood and a general time reversible model with 1,000 bootstrap replicates in MEGA-X (Kumar *et al.* 2018). Branch lengths were measured in the number of substitutions per site. This analysis involved 13 nucleotide sequences with a total of 14,622 positions in the final dataset. The alignment is in Supplementary materials. All plots of relevant nuclear and genomic features were plotted using the *circlize* (Gu *et al.* 2014) and *karyploteR* (Gel and Serra 2017) packages in R.

Results

Our initial assembly using 56× (Assembly 1) and 99× (Assembly 2) coverage data produced 31,942 and 37,551 scaffolds, respectively. The total length of Assembly 1 was 475,329,771 bp and the length of Assembly 2 was 536,018,598 bp. The scaffold N50 for Assembly 1 was 1,120,830 and was 1,107,799 for Assembly 2. To determine which genome assembly to move forward with for the Hi-C scaffolding, we compared BUSCO results between the two and found them to be similar, with the number of complete BUSCOs higher, and with less missing, in Assembly 1 [Assembly 1: C:95.0% (S:91.6%, D:3.4%), F:3.6%, M:1.4%, n:1066. Assembly 2: C:94.7% (S:92.0%, D:2.7%), F:3.4%, M:1.9%, n:1066]. Given these results and assembly software recommendations, we moved forward with Assembly 1.

Our Hi-C data allowed us to scaffold Assembly 1, and we confidently identified 21 total linkage groups ranging in length from 24,271,041 bp to 7,025,367 bp (Fig. 2a and Table 1). Whole genome

sequencing coverage differences between a female and male allowed us to identify the X chromosome (Fig. 2b). The karyotype is unknown for *C. aeneicollis*, and therefore, it is currently unclear if these linkage groups represent the entire chromosomes or are individual arms of metacentric chromosomes. Some scaffolds likely represent complete metacentric chromosomes given clear associations on the diagonal of what would appear to be two chromosomal arms (Fig. 2a; LG4, LG5, and LG11). This Hi-C contact pattern is similar to contact maps of metacentric chromosomes in Drosophila (Bracewell *et al.* 2019; Bracewell *et al.* 2020). In addition to 21 linkage groups, a number of contigs and scaffolds were unable to be placed into linkage groups but were retained for additional downstream filtering.

Repetitive sequences (transposable elements and tandemly repeated satellite sequence) were found to make up 43.5% of the genome assembly (Fig. 2b). We did not observe any enrichment in repetitive elements across LGs when all repeat families were grouped together (Fig. 2b). However, we found that specific repeat families were more frequently encountered than others, and LINEs constitute 7.1% of masked bases in the genome, LTR elements make up 5.4%, and DNA elements 3.8%. We found 26.2% of masked bases were unclassified and await future work. We found an increase in LTRs on the ends of most scaffolds suggesting a nonrandom distribution of these repeats that hints to their enrichment near centromeric or telomeric regions (Supplementary Fig. 1).

The cleaning of raw RNAseq reads from three individuals left us with a total of 21M, 25M, and 21M pairs of reads, and our combined transcriptome identified a total of 38,792 transcripts that were then used as evidence during annotation. Maker2 identified a total of 12,586 gene models in *C. aeneicollis*, with the number of protein-coding genes on the 21 linkage groups ranging from 202 to 678 (Table 1). Using OrthoDB, we were able to identify a total of 8,187 orthologous genes with *T. castaneum* (Herndon *et al.* 2020) thus providing some functional characterization of the *C. aeneicollis* annotation for future exploration. As a final cleaning step, we removed all contigs and scaffolds where the top two BLAST hits were either *Wolbachia* or *Spiroplasma* (known endosymbionts),
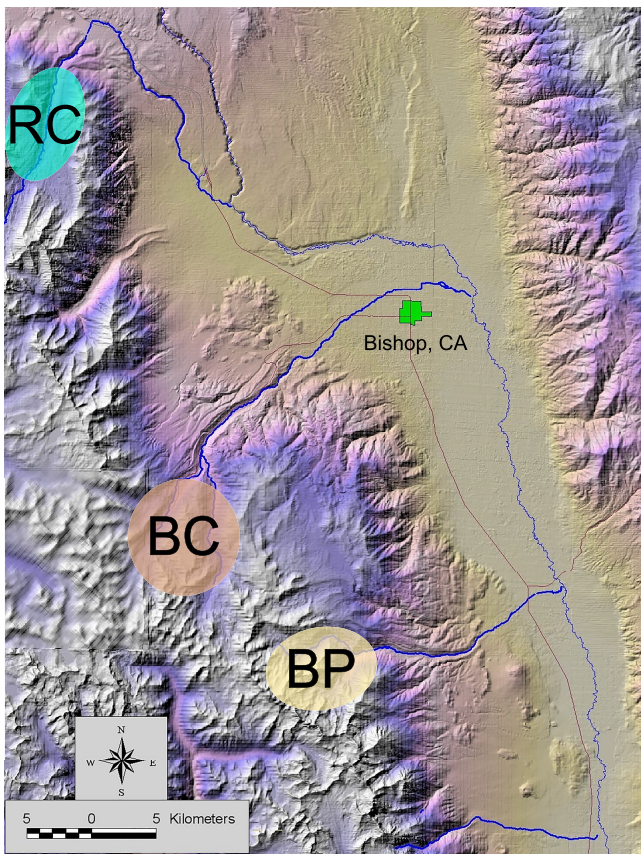
Fig. 3. Collection locations of 12 *Chrysomela aeneicollis* used in whole genome resequencing for mtDNA characterization. Individuals were initially collected from three distinct drainages identified as Rock Creek (RC), Bishop Creek (BC) and Big Pine (BP).

the contig/scaffold did not have at least one annotated protein-coding gene from above, or the contig/scaffold was a duplicated sequence already found in a linkage group or another scaffold/contig. After these filters, we retained 2,015 unplaced scaffolds (contigs) that contained 2,263 genes (Table 1). The above filters removed a total of 121.4 Mb of sequence. We visualized gene density across all LGs and any potential negative correlation with repeat density (Supplementary Fig. 1). For some linkage groups, we observed a general trend where one end of an LG was gene poor and showed increased repetitiveness, which is indicative of a telocentric or subtelocentric chromosome (e.g. LG15 and LGX). However, most LGs showed no obvious spatial patterns of gene enrichment/depletion, suggesting genes are fairly evenly distributed across most LGs.

A query of genes involved in the Electron Transport Chain pathway of cellular respiration based on KEGG identifiers revealed that 59% of genes found in the published *T. castaneum* genome were found in our *C. aeneicollis* assembly (41 of 69 genes, Supplementary Table 2). The assembly also includes the full coding sequence for the nuclear metabolic enzyme *phosphoglucose isomerase* (PGI), which is located on LG3. Preliminary results indicate that the PGI electromorphs that have been investigated extensively in prior studies (Dahlhoff and Rank 2000; Rank *et al.* 2007) are determined by a nonsynonymous substitution located in the first exon of the PGI coding sequence (Smeds 2022).

The reference sequence for the *C. aeneicollis* mitochondrion was generated in Novoplasty as a circularized genome of 18,007 bp. In the reference individual, the control region had a length of

3,390 bp, with 14,617 bp remaining for the transcribed region. For eight *C. aeneicollis* individuals and the *C. lapponica* individual, Novoplasty also assembled a circular genome. For four individuals, Novoplasty was only able to assemble linear contigs with multiple contigs in the control region. Sizes of *C. aeneicollis* mt genomes ranged from 17,158 to 18,794 bp (mean = 17,882.2, S.D. = 521.8, $n = 8$); the mt genome of the *C. lapponica* individual was somewhat smaller (16,750 bp). The transcribed region differed among individuals by only two base pairs, and therefore, the observed variation in mitochondrial genome size is largely due to differences in control region size rather than size of the transcribed region. One individual from Bishop Creek had an extra A nucleotide in the tRNA^{Gln} molecule, and six individuals (four from Big Pine Creek and two from Bishop Creek) possessed a 2-bp insertion in the large ribosomal subunit. The final annotations for the reference sequence revealed seventeen cases of overlapping gene products within the transcribed region. Seven cases involved genes that overlapped with more than one other gene, eight between protein-coding genes and tRNA gene products, four among protein-coding genes alone, and four tRNA gene regions. One was found between the small ribosomal subunit and a tRNA molecule.

The mitochondrial genomes of 12 sampled *C. aeneicollis* individuals (Fig. 3) showed substantial variation in the transcribed region, with nine distinct haplotypes recognized and four individuals (three from Big Pine Creek and one from Bishop Creek) sharing an identical DNA sequence (Fig. 4a AND Supplementary Table 1 for pairwise percent sequence divergence). Two distinct groups of haplotypes can be easily recognized; half of Bishop Creek individuals clustered together with individuals from the northern drainage Rock Creek and half clustered with the southern drainage Big Pine Creek (Fig. 4b). Fifty-three nucleotide differences occurred between RC64 and four individuals, all from Big Pine Creek, and the minimum number of differences between Rock Creek and Big Pine Creek individuals was forty-nine. This results in a substantial level of differentiation among northern and southern populations. The number of substitutions separating *C. lapponica* from *C. aeneicollis* was considerably greater (ranging from 253 to 260), indicating that the North American *C. aeneicollis* is monophyletic compared to the European *C. lapponica*.

Closer examination of genetic differences among individuals revealed that four nonsynonymous (amino acid changing) substitutions occurred in protein-coding genes, and all of these were unique to single individuals. Other differences were found either at synonymous sites within protein-coding genes or in RNA–encoding regions (Fig. 4a). Two substitutions occurred at the tRNA^{Leu} and tRNA^{Lys} regions that bracket the protein-coding *cytochrome oxidase II* gene. These substitutions cooccur with three synonymous substitutions within *cytochrome oxidase II*, which were used as RFLP markers of different populations in prior studies (Dahlhoff *et al.* 2019; Rank *et al.* 2020). Two more substitutions occurred in the large ribosomal subunit (16S) region, and beetles with southern haplotypes (4 BP and 2 BC) possessed a two-base pair insertion in this region. No substitutions occurred in regions of the mitochondrial genome where overlapping genes are found.

Analysis of the putative secondary structure of tRNA and rRNA molecules produced by these substitutions revealed that nucleotide differences between northern and southern haplotypes are located in regions that could affect functional properties of the molecules. Within the 65 bp tRNA^{Leu} region, northern beetles possess a "C" nucleotide at position 34, while southern beetles possess an "A" nucleotide at the same position. This substitution is located in the loop

(a)

(c)

(b)
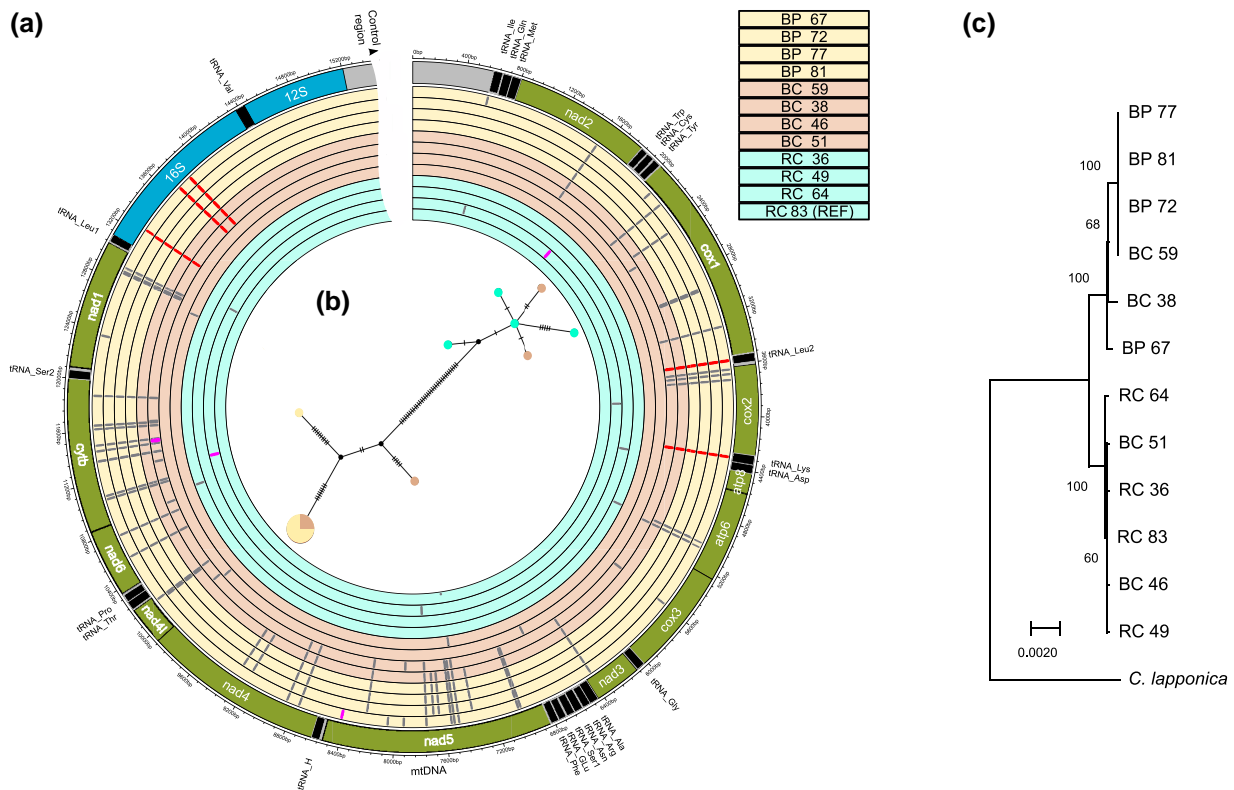
Fig. 4. Mitochondrial genome structure. a) Circular representation of the assembled mtDNA genome from *C. aeneicollis* (outer track) with the locations of tRNAs (black), protein-coding genes (green), and rRNA (blue). The control region is not shown in its entirety. The inner twelve tracks show variants found in 4 individuals each from Big Pine (BP), Bishop Creek (BC), and Rock Creek (RC). An individual from Rock Creek was used as the reference (RC_83) and is the innermost track. Variants found in each mtDNA genome are shown as tick marks in their respective location. Grey marks show variants in tRNAs or protein-coding genes that were synonymous or did not influence the tRNA structure. Magenta marks are locations of nonsynonymous variants. Red tick marks highlight variants that were found in tRNAs or rRNA and appear to alter the secondary structure of the molecule. b) Haplotype network showing the relationships and number of nucleotide differences (black ticks) between mitotypes. c) Maximum likelihood phylogeny showing evolutionary relationships between mtDNA from 12 individuals with *C. lapponica* as an outgroup. Bootstrap values ≥ 60 are shown.

portion of the anticodon, two base pairs away from the "TAA" anticodon (Supplementary Fig. 2). Within the 69-base pair tRNA^Lys region, the substitution from "T" in the northern populations to "C" in the southern ones occurs at position 37, which occurs in the final base in the stem portion of the anticodon and results in noncanonical base pairing, which would affect stability of a critical region within the molecule (Supplementary Fig. 3). Finally, within the 16S ribosomal RNA region, two substitutions and one insertion distinguish northern and southern beetles, and these sequence differences result in differences in predicted secondary structure and free energy values (Supplementary Fig. 3).

Discussion

Using a combination of 10× genomics sequencing, Hi-C scaffolding, transcriptomics, and population-level whole genome resequencing, we were able to assemble highly-quality nuclear and mitochondrial genomes for the willow leaf beetle, *C. aeneicollis*. The total length of our nuclear assembly (356 Mb) is in the range of other chromosome-length beetle genome assemblies and provides an important resource for this underrepresented order of insects. Our annotation of 12,586 genes is also consistent with other high-quality beetle genome assemblies (Fallon *et al.* 2018; Van Dam *et al.* 2021; Keeling *et al.* 2022), and by establishing orthology with the model species, *T. castaneum* (Herndon *et al.* 2020), we have putative gene functions for over 60% of the genes in our annotation. The putative number of linkage groups proposed here (21)

is roughly consistent with cytological studies of other *Chrysomela* (Petitpierre 2011). Within Chrysomelini, the ancestral haploid chromosome number appears to be 17. Our genome assembly suggests a genome size that is also consistent with earlier cytological investigations (Petitpierre 2011).

Mitochondrial differences along the latitudinal gradient show substantial differences between northern and southern populations and are consistent with nuclear genetic data suggesting that Bishop Creek populations reflect introgression between north and south (Rank *et al.* 2020). The number of substitutions separating Rock Creek and Big Pine mitotypes suggests that they have been distinct for a considerable length of time. Levels of differentiation are greater at mitochondrial than at nuclear loci in *C. aeneicollis* (Dellicour *et al.* 2014), which is consistent with predictions that mitochondria should evolve faster than the nuclear genome due to higher mutation rates (Burton and Barreto 2012), reduced effective population sizes (Neiman and Taylor 2009), and potential functional differences among mitotypes (Hill 2015) (although see Cooper *et al.* 2015; Edwards *et al.* 2021).

In *C. aeneicollis*, mitochondrial variation appears to be partly a result of natural selection and adaptation to local environmental conditions (Dahlhoff *et al.* 2019; Rank *et al.* 2020). Specifically, beetles are found at higher elevations in Big Pine Creek in the south than in Rock Creek, and larvae with the Big Pine Creek mitotype developed twice as fast as those with the northern mitochondrial haplotype when reared at 3,800 m ASL at the Barcroft Station of the White Mountain Research Center (Dahlhoff *et al.* 2019).

Conversely, when larvae were reared at 1,240 m at the Owens Valley Station, running speed of beetles with northern mitotypes recovered from a nonlethal heat exposure more effectively than those with the southern mitotype (Rank *et al.* 2020). In addition, beetles with matching mitochondrial and nuclear genetic backgrounds (southern/southern or northern/northern) show higher reproductive success and faster larval development in the field (Rank *et al.* 2020).

Our results show that most differences between northern and southern mitotypes occur at synonymous sites within protein-coding genes, and all nonsynonymous substitutions were unique to a single beetle. In contrast, differences between northern and southern mitotypes at two tRNA loci and the large ribosomal subunit are predicted to generate functional differences in secondary structure. These differences could result in incompatibilities between nuclear-encoded genes that function inside mitochondria to perform protein synthesis and their mitochondrial RNA counterparts (Hoekstra *et al.* 2013). Our results provide a potential mechanism for the mitonuclear epistasis described in prior studies (Rank *et al.* 2020) as three critical metabolic processes require integration between mitochondrial and nuclear gene products (Hill 2015, 2016; Hill *et al.* 2019; Levin *et al.* 2014). These processes include interactions among proteins involved in the electron transport chain, mitochondrial replication inside cells, and protein synthesis inside the mitochondrion (Levin *et al.* 2014). This third mechanism has been proposed to operate in lab-generated lines of Drosophila that were created to generate mitochondrial mismatch (Hoekstra *et al.* 2013).

In summary, the genomic information presented in this paper adds important resources that expand our understanding of beetle genomics and will help those studying the genetics and genomics of host plant use in phytophagous insects. Coupled with our mtDNA assembly, these resources will be of great use to those investigating the evolutionary physiology of ectotherm metabolism in natural environments characterized by multiple axes of abiotic stress.

Data Availability

All data used in the genome assembly and annotation are described and available under bioproject PRJNA693623. The genome assembly accession at NCBI is GCA_027562985.1 and the mtDNA accession is OPT787486. All associated files are also available here: https://figshare.com/projects/A_chromosome_scale_genom e_assembly_and_evaluation_of_mtDNA_variation_in_the_willow _leaf_beetle_Chrysomela_aeneicollis/152286.

Supplemental material available at G3 online.

Acknowledgements

Funding

Conflicts of interest statement

The author(s) declare no conflict of interest.

Literature cited

Bracewell R, Chatla K, Nalley MJ, Bachtrog D. Dynamic turnover of centromeres drives karyotype evolution in Drosophila. Elife. 2019;8:e49002. doi:10.7554/eLife.49002.

Bracewell R, Tran A, Chatla K, Bachtrog D. Chromosome-level assembly of *Drosophila bifasciata* reveals important karyotypic transition of the X chromosome. G3 (Bethesda). 2020;10(3):891–897. doi:10.1534/g3.119.400922.

Brown WJ. The new world species of Chrysomela l. (Coleoptera: Chrysomelidae). Mem Entomol Soc Canada. 1956;88(S3):5–54. doi:10.4039/entm8803fv.

Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol Ecol. 2012;21(20): 4942–4957. doi:10.1111/mec.12006.

Cameron SL. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Syst Entomol. 2014;39(3):400–411. doi:10.1111/syen.12071.

Camus MF. Digest: mitonuclear interactions modulate life history phenotypes in the wild. Evolution. 2020;74(8):1877–1878. doi:10. 1111/evo.13988.

Chan PP, Lowe TM. Trnascan-se: searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019;1962:1–14. doi:10.1007/978-1-4939-9173-0_1.

Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL. Similar efficacies of selection shape mitochondrial and nuclear genes in both *Drosophila melanogaster* and *Homo sapiens*. G3 (Bethesda). 2015;5-(10):2165–2176. doi:10.1534/g3.114.016493.

Dahlhoff EP, Dahlhoff VC, Grainger CA, Zavala NA, Otepola-Bello D, Sargent BA, Roberts KT, Heidl SJ, Smiley JT, Rank NE. Getting chased up the mountain: high elevation may limit performance and fitness characters in a montane insect. Funct Ecol. 2019; 33(5):809–818. doi:10.1111/1365-2435.13286.

Dahlhoff EP, Fearnley SL, Bruce DA, Gibbs AG, Stoneking R, McMillan DM, Deiner K, Smiley JT, Rank NE. Effects of temperature on physiology and reproductive success of a montane leaf beetle: implications for persistence of native populations enduring climate change. Physiol Biochem Zool. 2008;81(6):718–732. doi:10. 1086/590165.

Dahlhoff EP, Rank NE. Functional and physiological consequences of genetic variation at phosphoglucose isomerase: heat shock protein expression is related to enzyme genotype in a montane beetle. Proc Natl Acad Sci U S A. 2000;97(18):10056–10061. doi:10. 1073/pnas.160277697.

Dellicour S, Fearnley S, Lombal A, Heidl S, Dahlhoff EP, Rank NE, Mardulyn P. Inferring the past and present connectivity across the range of a North American leaf beetle: combining ecological niche modeling and a geographically explicit model of coalescence. Evolution. 2014;68(8):2371–2385. doi:10.1111/evo.12426.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356(6333):92–95. doi:10.1126/science.aal3327.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016a;3(1):99–101. doi:10.1016/j.cels.2015.07.012.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b;3(1):95–98. doi:10.1016/j.cels.2016.07.002.

Edwards DM, Røyrvik EC, Chustecki JM, Giannakis K, Glastad RC, Radzvilavicius AL, Johnston IG. Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. PLoS Biol. 2021;19(4):e3001153. doi:10.1371/journal.pbio.3001153.

Elmore. Effects of temperature and oxygen stress on gene expression and performance in a montane leaf beetle. Department of Biology, Sonoma State University; 2021.

Fallon TR, Lower SE, Chang C-H, Bessho-Uehara M, Martin GJ, Bewick AJ, Behringer M, Debat HJ, Wong I, Day JC, *et al.* Firefly genomes illuminate parallel origins of bioluminescence in beetles. Elife. 2018;7:e36495. doi:10.7554/eLife.36495.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117-(17):9451–9457. doi:10.1073/pnas.1921046117.

Gel B, Serra E. Karyoploter: an R/bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 2017;33(19):3088–3090. doi:10.1093/bioinformatics/btx346.

Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in r. Bioinformatics. 2014;30(19):2811–2812. doi:10.1093/bioinformatics/btu393.

Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, Waterhouse RM, Liang C, Damm C, Siemanowski J, *et al.* Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. BMC Genomics. 2020;21(1):47. doi:10.1186/s12864-019-6394-6.

Hill GE. Mitonuclear ecology. Mol Biol Evol. 2015;32(8):1917–1927. doi:10.1093/molbev/msv104.

Hill GE. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecol Evol. 2016;6(16):5831–5842. doi:10.1002/ece3.2338.

Hill GE, Havird JC, Sloan DB, Burton RS, Greening C, Dowling DK. Assessing the fitness consequences of mitonuclear interactions in natural populations. Biol Rev Camb Philos Soc. 2019;94(3):1089–1104. doi:10.1111/brv.12493.

Hoekstra LA, Siddiq MA, Montooth KL. Pleiotropic effects of a mitochondrial-nuclear incompatibility depend upon the accelerating effect of temperature in Drosophila. Genetics. 2013;195(3):1129–1139. doi:10.1534/genetics.113.154914.

Holt C, Yandell M. Maker2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12(1):491. doi:10.1186/1471-2105-12-491.

Hotaling S, Kelley JL, Frandsen PB. Toward a genome sequence for every animal: where are we now? Proc Natl Acad Sci U S A. 2021;118(52):e2109019118. doi:10.1073/pnas.2109019118.

Keeling CI, Campbell EO, Batista PD, Shegelski VA, Trevoy SAL, Huber DPW, Janes JK, Sperling FAH. Chromosome-level genome assembly reveals genomic architecture of northern range expansion in the mountain pine beetle, *Dendroctonus ponderosae* hopkins (Coleoptera: Curculionidae). Mol Ecol Resour. 2022;22(3):1149–1167. doi:10.1111/1755-0998.13528.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. Nat Biotechnol. 2019;37(8):907–915. doi:10.1038/s41587-019-0201-4.

Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5(1):59. doi:10.1186/1471-2105-5-59.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with stringtie2. Genome Biol. 2019;20(1):278. doi:10.1186/s13059-019-1910-1.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. Orthodb v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2018;47(D1):D807–D811. doi:10.1093/nar/gky1053.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega x: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–1549. doi:10.1093/molbev/msy096.

Levin L, Blumberg A, Barshad G, Mishmar D. Mito-nuclear co-evolution: the positive and negative sides of functional ancient mutations. Front Genet. 2014;5:448. doi:10.3389/fgene.2014.00448.

Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–293. doi:10.1126/science.1181369.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. Viennarna package 2.0. Algorithms Mol Biol. 2011;6(1):26. doi:10.1186/1748-7188-6-26.

McMillan DM, Fearnley SL, Rank NE, Dahlhoff EP. Natural temperature variation affects larval survival, development and hsp70 expression in a leaf beetle. Funct Ecol. 2005;19(5):844–852. doi:10.1111/j.1365-2435.2005.01031.x.

Millstein RL. Natural selection as a population-level causal process. Br J Philos Sci. 2006;57(4):627–653. doi:10.1093/bjps/axl025.

Neiman M, Taylor DR. The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci. 2009;276(1660):1201–1209. doi:10.1098/rspb.2008.1758.

Otto SB, Berlow EL, Rank NE, Smiley J, Brose U. Predator diversity and identity drive interaction strength and trophic cascades in a food web. Ecology. 2008;89(1):134–144. doi:10.1890/07-0066.1.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20(1):275. doi:10.1186/s13059-019-1905-y.

Petitpierre E. Cytogenetics, cytotaxonomy and chromosomal evolution of Chrysomelinae revisited (Coleoptera, Chrysomelidae). Zookeys. 2011;157:67–79. doi:10.3897/zookeys.157.1339.

Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

Ramani V, Cusanovich DA, Hause RJ, Ma W, Qiu R, Deng X, Blau CA, Disteche CM, Noble WS, Shendure J, *et al.* Mapping 3d genome architecture through in situ dnase Hi-C. Nat Protoc. 2016;11(11):2104–2121. doi:10.1038/nprot.2016.126.

Rank NE. A hierarchical analysis of genetic differentiation in a montane leaf beetle *Chrysomela aeneicollis* (Coleoptera: Chrysomelidae). Evolution. 1992;46(4):1097–1111. doi:10.2307/2409759.

Rank NE. Host-plant effects on larval survival of a salicin-using leaf beetle *Chrysomela aeneicollis* Schaeffer (Coleoptera: Chrysomelidae). Oecologia. 1994;97(3):342–353. doi:10.1007/BF00317324.

Rank NE, Bruce DA, McMillan DM, Barclay C, Dahlhoff EP. Phosphoglucose isomerase genotype affects running speed and

heat shock protein expression after exposure to extreme temperatures in a montane willow beetle. J Exp Biol. 2007;210(5): 750–764. doi:10.1242/jeb.02695.

Rank NE, Dahlhoff EP. Allele frequency shifts in response to climate change and physiological consequences of allozyme variation in a montane insect. Evolution. 2002;56(11):2278–2289. doi:10.1111/ j.0014-3820.2002.tb00151.x.

Rank NE, Mardulyn P, Heidl SJ, Roberts KT, Zavala NA, Smiley JT, Dahlhoff EP. Mitonuclear mismatch alters performance and reproductive success in naturally introgressed populations of a montane leaf beetle. Evolution. 2020;74(8):1724–1740. doi:10.1111/evo.13962.

Roberts KT. Transcriptomic evidence suggests that the montane leaf beetle prioritizes digestion and reproductive in a sex-specific manner. Department of Integrative Biology, UC Berkeley; 2022.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19): 3210–3212. doi:10.1093/bioinformatics/btv351.

Smeds E. mtDNA variation in a montane beetle. Department of Biology, Sonoma State University; 2022.

Smiley JT, Horn JM, Rank NE. Ecological effects of salicin at three trophic levels: new problems from old adaptations. Science. 1985;229(4714):649–651. doi:10.1126/science.229.4714.649.

Smiley JT, Rank NE. Predator protection versus rapid growth in a montane leaf beetle. Oecologia. 1986;70(1):106–112. doi:10.1007/ BF00377117.

Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19 Suppl 2: ii215–ii225. doi:10.1093/bioinformatics/btg1080.

Van Dam MH, Cabras AA, Henderson JB, Rominger AJ, Pérez Estrada C, Omer AD, Dudchenko O, Lieberman Aiden E, Lam AW. The easter egg weevil (Pachyrhynchus) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution. PLoS Genet. 2021;17(8):e1009745. doi:10.1371/journal.pgen.1009745.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27(5): 757–767. doi:10.1101/gr.214874.116.

Editor: E. Betran