A high-quality carabid genome provides insights into beetle genome evolution and cold adaptation

Yi-Ming Weng[1], Charlotte Francoeur[2], Cameron Currie[2], David Kavanaugh[3], and Sean Schoville[4]

[1]University of Wisconsin Madison Graduate School
[2]University of Wisconsin-Madison
[3]California Academy of Sciences
[4]University of Wisconsin Madison

January 19, 2021

Abstract

The hyper-diverse order Coleoptera comprises a staggering ~25% of known species on Earth. Despite recent breakthroughs in next generation sequencing, there remains a limited representation of beetle diversity in assembled genomes. Most notably, the ground beetle family Carabidae, comprising more than 40,000 described species, has not been studied in a comparative genomics framework using whole genome data. Here we generate a high-quality genome assembly for Nebria riversi, to examine sources of novelty in the genome evolution of beetles, as well as genetic changes associated with specialization to high elevation alpine habitats. In particular, this genome resource provides a foundation for expanding comparative molecular research into mechanisms of insect cold adaptation. Comparison to other beetles shows a strong signature of genome compaction, with N. riversi possessing a relatively small genome (~147 Mb) compared to other beetles, with associated reductions in repeat element content and intron length. Small genome size is not, however, associated with fewer protein-coding genes, and an analysis of gene family diversity shows significant expansions of genes associated with cellular membranes and membrane transport, as well as protein phosphorylation and muscle filament structure. Finally, our genomic analyses show that these high elevation beetles have endosymbiotic Spiroplasma, with several metabolic pathways (e.g. propanoate biosynthesis) that might complement N. riversi, although its role as a beneficial symbiont or as a reproductive parasite remains equivocal.

Introduction

Beetles (Coleoptera) are the most diverse animal group on the planet, with more than 386,000 species described worldwide (Bouchard *et al.* 2017; Forbes *et al.* 2018). The extremely high diversity and abundance of beetles plays a critical role in most terrestrial ecosystems (Brooks *et al.* 2012; Erwin 1997; Schowalter 2013; Slade *et al.* 2019), with further significance to human society as they serve as food, pests, pets, biological control agents, and models in biological and engineering research (Adamski *et al.* 2019). Despite their importance, relatively few genomic studies have been conducted on beetles to date (McKenna 2018).

Coleoptera are divided into four suborders including Polyphaga, Adephaga, Myxophaga, and Archostemata, based on both morphological and molecular evidence (Beutel & Haas 2000; Zhang *et al.* 2018). Among the four suborders, Polyphaga is the most diverse group, comprising approximately 90% of coleopteran species, while most of the remaining 10% comprise Adephaga (Grimaldi & Engel 2005). Carabidae is the largest family of Adephaga, consisting of around 40,000 species, and is known for its broad use of terrestrial habitats, ecosystem services, and environmental sensitivity as bioindicators (Brooks *et al.*2012; Lorenz 2020). Except for a few species that are considered agricultural pests (e.g. *Zabrus tenebrioides*), most carabid species are predators or scavengers that serve as beneficial biological control agents and play key roles in local food

webs (Georgescu *et al.* 2017; Pizzolotto *et al.* 2018). Carabid beetles are found in a wide range of habitats, spanning all terrestrial biomes, from tropical and desert habitats, to arctic and alpine regions. As a result, they are frequently used in studies of community ecology, conservation, and biogeography (Kotze *et al.* 2011). Recent phylogenomic work has provided an understanding of the evolutionary relationships of Carabidae to other Coleoptera lineages and some patterns of gene family diversification (Gustafson *et al.* 2019; McKenna *et al.*2019; Misof *et al.* 2014; Seppey *et al.* 2019a), but has not provided a high-quality genome assembly for comparative analysis.

Carabid beetles are one of the few insect lineages that thrive in cold habitats (Mani 1968). A strong positive relationship exists between minimal environmental temperature and minimal temperature tolerance in insects (Kimura 2004), because insects, as ectotherms, do not regulate body heat through metabolism, but must instead physiologically adjust to the surrounding environment (Addo-Bediako *et al.* 2002). Cold environments consequently pose a significant challenge to insects, with few species surviving these conditions. Less than 2% of North American insect species are found in the arctic biome (Danks *et al.* 2017), and insect species richness declines dramatically in alpine and sub-nival habitats (Nagy & Grabherr 2009). Carabid beetles, however, are notable as arctic and alpine insects, comprising ˜40-50% of beetle species richness in the arctic (Chernov *et al.* 2000; Danks 1981), and setting the highest altitudinal distribution record of Coleoptera, with three taxa (*Amara altiphila* , *Bembidion maddisoni* , and *Trechus astrophilus*) reaching 5,600 m above sea level (a.s.l.) in the Himalayas (Schmidt 2009; Schmidt *et al.* 2017). The carabid beetle*Pterostichus brevicornis* additionally sets a record among beetles in demonstrating extreme cold tolerance, where the $LD_{50}$of overwintering adults occurs at -87°C (Baust 1972). The carabid genus*Nebria* (> 500 species) is a notable example of a cold-specialized beetle lineage, which exhibits its greatest diversity and abundance at high latitude and high elevation (Lorenz 2020). For example, 12 species of *Nebria* are found around Mount Rainier, Washington, with four species inhabiting the alpine and sub-nival zone (> 2000 m a.s.l., Kavanaugh 1985; Slatyer & Schoville 2016). Comparative genomic analysis of other insects has shown that adaptation to cold environments might involve constraints in genome architecture instead of gene content (Kelley *et al.* 2014; Kim*et al.* 2017), although other studies have mostly found evidence of accelerated protein evolution (Cicconardi *et al.* 2020; Parker*et al.* 2018). With the dual purpose of generating genomic resources to investigate Carabid beetle genome evolution and cold specialization, we assemble the genome of *Nebria riversi*Van Dyke (Carabidae: Nebriinae), an alpine ground beetle endemic to the Sierra Nevada in California (2900–3910 m a.s.l.). Considering its highly divergent taxonomic position from other Coleoptera genomes, and as the only cold-specialized beetle genome sequenced to date, we investigate genome architecture, gene family diversity, and differential gene expression across life-stages and sexes. Finally, as the microbial associates of high elevation insects have rarely been examined (Receveur*et al.* 2020), we scan genomic evidence for symbiotic microbes in*N. riversi* .

Methods

Sampling, Nucleic Acid Extraction and Sequencing

For the whole genome assembly, a female individual of *N. riversi*from Conness Lakes (N37.97224, W-119.30703), California, was used to conduct long-read sequencing, and a second female from the same population was used to conduct short-read sequencing. High molecular weight genomic DNA was extracted for the long-read library using the 10x Genomics protocol (DNA Extraction from Single Insects), followed by a purification and concentration step with SPRI magnetic beads. A PacBio library was prepared using the HS Large Fragment 50kb Kit by the UW-Madison Biotechnology Center, with an average library insert size of 35,640 base pairs. It was then sequenced on a PacBio Sequel II machine to approximately 24x coverage. The short-read, pair-end library (2 x 150 bp) was prepared using the Nextera DNA Flex kit and sequenced on the Illumina NovaSeq 6000 to approximately 100x coverage. Additional samples representing the egg stage, larva, adult male and adult female *N. riversi* were obtained for RNA sequencing. These samples were derived from Conness Lakes, California, with the exception that the larva was a progeny of a cross between Conness Lakes and a Central Sierra Nevada population. Total RNA was extracted using Trizol reagent (Life Technologies) and purified using the PureLink kit (Ambion). Pair-end libraries (2 x 150 bp) were then generated using the

TruSeq Stranded Total RNA Library Prep kit by the UW-Madison Biotechnology Center, and samples were sequenced on the Illumina NovaSeq 6000 to generate approximately 30 million reads per sample. All raw sequencing data has been made publicly available on NCBI (SRA Accessions: SRR12967583- SRR12967588).

Transcriptome Assembly and Transcriptome Assessment

In order to generate a comprehensive transcriptome for *N. riversi* , we combined all RNAseq data and generated a *de novo*transcriptome assembly using Trinity v2.10.0 (Haas *et al.* 2013). We used default settings, but normalized the input reads in silico based on the calculated maximum read coverage. To provide a quantitative assessment of transcriptome completeness, we first assessed the number of full-length transcripts using blastx v2.7.1 (Camacho*et al.* 2009) to query the UniProt Swiss-Prot database (UniProt Consortium 2019). We then examined alignment scores relative to a set of near-universal single-copy orthologs using the software BUSCO v2.0 (Seppey *et al.* 2019b). We selected the reference gene set for Endopterygota (OrthoDB v9), which contains 2,442 genes. To further refine the transcriptome assembly, a super-transcriptome was generated by merging the *de novo* transcriptome from Trinity and the annotated genome assembly (see below) using Necklace v1.11 (Davidson & Oshlack 2018). The goal of this step was to produce a compact, but comprehensive set of transcribed genes that reflect total evidence. The super-transcriptome assembly has been made publicly available on NCBI (TSA Accession: GIWW00000000).

Mitogenome Assembly and Annotation

The *N. riversi* mitochondrial genome was assembled using the SMART2 software package (Alqahtani & Măndoiu 2019), based on one quarter of the Illumina pair-end short read genomic data. We confirmed the gene order by mapping long-read data to this assembly in Geneious Prime 2019 (Biomatters Ltd.). The mitochondrial genome was then annotated using MITOS (Bernt *et al.* 2013) and coding regions were subsequently refined manually in Geneious with reference to published beetle mitochondrial genomes (Timmermans & Vogler 2012). We compared the resulting annotated genome to a published mitochondrial genome for *Nebria brevicollis* (KT876906; Linard*et al.* 2016) and to the gene order of other Carabidae (López-López & Vogler 2017). The mitochondrial genome assembly has been made publicly available on NCBI (Accession: MW244066).

Genome Assembly, Decontamination and Genome Assessment

To assemble the *N. riversi* genome, we first generated a long-read genome assembly using the software CANU v2.0 (Koren *et al.* 2017). Default parameters were used, with the estimated genome size set to 230 MB, except the lowCovSpan and lowCovDepth parameters were set to 0.5 and 0, respectively. To error correct the long reads, we polished the genome by first mapping the Illumina short reads to the draft genome using Minimap2 (Li 2018), then the mapped reads (sam files) and raw reads (fastq files) were used to polish the draft genome using Racon (Vaser *et al.* 2017). This draft long-read assembly was used as a benchmark to assess our final genome assembly, as well as to identify non-targeted sequences. To provide a quantitative assessment of the draft genome completeness, we used BUSCO with the Endopterygota reference gene set. We then used two methods to identify possible non-target sequences in the assembly, BLOBtools v1 (Laetsch & Blaxter 2017), which uses GC content and short-read coverage to help identify foreign DNA, and the sendsketch tool in BBMap v38.86 (Bushnell 2014), which uses a k-mer based approach to match sequences to reference databases. For sendsketch, the number of sequences was set to 100k and the sketch length was set to 200k. We identified the most abundant microbial taxa (*Spiroplasma* and*Acinetobacter*), and used their Genbank reference sequences, as well as the *N. brevicollis* mitochondrial genome, to filter the long-read data using Minimap2 (Li 2018). The short-read genomic data was also screened with these reference sequences using the bbduk tool in BBMap, and then normalized using bbnorm. With these refined read sets, we generated a final assembly using the hybrid assembler Haslr (Haghshenas*et al.* 2020). Runs of Haslr were conducted at different settings of long-read coverage (10x, 20x and 25x), and each assembly was then evaluated using BUSCO with the Endopterygota reference gene set. As a final assembly step, we employed pair-end RNAseq data to join contigs using P_RNA_scaffolder (Zhu *et al.* 2018). BUSCO was run again on this assembly, and we compared the final RNA-scaffolded haslr 20x assembly to the *Tribolium castaneum* v5.2 RefSeq assembly

(GCF_000002335.3) using QUAST v 5.1.0rc1 (Gurevich *et al.* 2013). The genome assembly has been made publicly available on NCBI (WGS Accession: JADQWA010000000).

Genome Annotation

In order to annotate both structural and functional properties of the genome, we used GenSAS v6 (Humann *et al.* 2019). We first identified and masked interspersed repeats and low complexity DNA sequences in the assembly. RepeatModeler v1.0.11 was run to identify and produce a structural annotation of repeat regions *de novo* (Smit & Hubley 2008), and subsequently RepeatMasker v4.1.0 was used to generate a modified version of the genome with these regions masked (Smit *et al.* 2015). We then employed Braker v2.1.0 (Hoff *et al.* 2019; Stanke *et al.*2008; Stanke *et al.* 2006) to automatically predict gene models for protein coding genes in the masked genome. The Braker pipeline uses the tools GeneMark-ES/ET (Lomsadze *et al.*2005; Ter-Hovhannisyan *et al.* 2008) and Augustus (Camacho *et al.* 2009), as well as evidence from RNAseq data, to predict gene models in novel eukaryotic genomes. Paired-end RNAseq reads were aligned to the genome using Hisat2 (Kim *et al.*2015), with default settings, and the alignment file was provided to Braker. Finally, PASA v2.3.3 (Haas *et al.* 2008) was used to refine the gene models, using the assembled transcriptome as input.

For the resulting consensus gene models, we assigned functional annotations using a combination of six tools. Amino acid similarity to proteins in the NCBI RefSeq invertebrate database was used for functional annotation based on searches with both Diamond v0.9.22 (Buchfink *et al.* 2015) and Blastp v2.7.1 (with the settings: matrix = BLOSUM62, expect = $1e^{-8}$, word size = 3, gap open penalty = 11, gap extend penalty = 1, maximum HSP distance =30000). The gene set was further annotated based on the presence of peptide domains, using InterProScan v5.29-68.0, Pfam v1.6 (with the settings: e-value sequence = 1 and e-value domain = 10), and SignalP v4.1 (with the settings: organism group = eukaryotes, method = best, D-cutoff for noTM networks = 0.45, D-cutoff for TM networks = 0.50, minimal predicted peptide length = 10, and truncate sequence length = 70). Finally, the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology terms were assigned to each gene using the KEGG Automatic Annotation Server (Moriya*et al.* 2007), based on bi-directional best hit searches of the nucleotide sequences using blast.

Comparison of Coleoptera Genome Structure and Synteny

To test for structural genomic differences in *N. riversi* and other beetles, we compared estimates of genome size, assembly fragmentation, intron length, and transposable element (TE) repeat content. For genome size, we compared *N. riversi* with the estimates for eight other Adephaga species reported in Pflug *et al* . (2020), coupled with 11 species of Polyphaga that had sufficient read coverage data available from NCBI's RefSeq database (O'Leary *et al.* 2016). These 11 reference taxa (characteristics of these beetle genomes can be found in **Table 1**) include:*Aethina tumida* (Atum), *Agrilus planipennis* (Apla),*Anoplophora glabripennis* (Agla), *Dendroctonus ponderosae*(Dpon), *Diabrotica virgifera* (Dvir), *Leptinotarsa decemlineata* (Ldec), *Nicrophorus vespilloides* (Nves),*Onthophagus taurus* (Otau), *Photinus pyralis* (Ppyr),*Sitophilus oryzae* (Sory), and *T. castaneum* (Tcas). For all 20 species, several k-mer based methods were used to estimate genome size, including GenomeScope, basic CovEST, and repeat CovEST (Hozza *et al.* 2015; Ranallo-Benavidez *et al.* 2020), as well as de novo assembly size as reported by Pflug*et al* . (2020) and RefSeq. The k-mer count distributions were calculated using Jellyfish v2.2.3 (Marcais & Kingsford 2011). To compare structural genomic features, we focus on *N. riversi* and the 11 annotated RefSeq genomes. Assembly statistics were obtained directly from RefSeq and the intron sequences were extracted from the latest assembled genomes and gff files (downloaded on July 27, 2020) using GFF_Ex v2.3 (Rastogi & Gupta 2014). Transposable element (TE) repeat content was either retrieved from published genome studies, or estimated from the available assembly using RepeatModeler and RepeatMasker. Intron sequences were used to calculate the intron length distribution, and then compared to those values reported in RefSeq annotation results. To test whether *N. riversi* has smaller introns, we employed Kolmogorov–Smirnov statistic to test whether the cumulative distribution of intron length of *N. riversi* was less than other Coleopteran species.

To examine the extent of synteny among beetle genomes, we used MCScanX (Wang *et al.* 2012) to measure

the number of collinear genes found along scaffolds in pairwise comparisons. We focused on comparing the *N. riversi* assembly to the six most contiguous beetle genomes available on RefSeq: *Aethina tumida* (contig N50 = 299 Kb), *Agrilus_planipennis* (scaffold N50 = 1.11 Mb), *Onthophagus taurus* (scaffold N50 = 337 Kb),*Photinus pyralis* (scaffold N50 = 47.02 Mb), *Sitophilus oryzae* (scaffold N50 = 2.86 Mb), and *Tribolium castaneum*(scaffold N50 = 4.46 Mb). We calculated the number of contiguous genes, as well as the proportion of contiguous genes, among the seven genomes. The pheatmap package in R (Kolde & Kolde 2015) was used to generate visualizations of the results.

Sex Chromosome Identification

Although karyotype analysis has been conducted in several *Nebria*species (Serrano & Galián 1998; Serrano & Yadav 1984), the sex chromosome system has not been identified. However, either XY and XO systems are present in other carabid beetles (Serrano & Yadav 1984), so we employed two computational approaches to identify the X chromosome (Palmer *et al.* 2019). First, because female individuals are expected to have two copies of the X chromosome and male individuals only have one copy, we examined sex-specific coverage to distinguish X chromosome scaffolds. We used a published dataset (Weng *et al.*2020) of genotyping by sequencing (GBS) reads from 59 *N. riversi*individuals (BioProject PRJNA645878, specimen data listed in**Table S1**), comprising 34 females and 25 males from three sites, to calculate the sequencing depth across the genome. Single-end GBS reads were mapped to the genome using Bwa-Mem v0.7.17.4 to generate the mapped bam files (Li 2013) and Samtools v1.10.03 was used to calculate the mean depth of coverage for each scaffold (Li*et al.* 2009). Welch's t-tests were conducted for each contig to test whether mean coverage of male individuals was half that of female individuals. All contigs with p-value < 0.05, and/or showing a ratio of male/female coverage between 0.45 and 0.55, were considered X chromosome candidates. Second, we further processed the GBS data to identify single nucleotide polymorphisms (SNPs) among the candidate X chromosome scaffolds. We added the read group to the bam files and sorted them using Picard v2.23.3 (Picard 2018), then called SNPs using HaplotypeCaller in GATK v.4.1.8.1 (McKenna *et al.* 2010; Poplin *et al.* 2017). We examined whether scaffolds contained SNPs that were heterozygous in males, as male individuals are expected to carry a single copy of X chromosome and thus should show single allele in those loci under an ideal situation without sequencing errors. In order to keep the high quality SNPs, we first filtered by read depth (>10x) and missing data, requiring genotype calls in at least half of female or male samples. The SNPs that passed the filter were used to calculate the mean read depths for males and females, and the scaffolds carrying SNPs with mean read depths ratio of male/female between 0.4 and 0.6 were marked as candidates. We repeated this for each population from the three sample sites (Conness Lake, Donohue Pass, and Lyell Peak; for more information about collecting information see: Weng *et al.* 2020) and all samples together, and the contigs that identified in all four data sets were considered X chromosome candidates. Finally, all X chromosome candidates were screened for the presence of heterozygote SNPs in males, and only those scaffolds that had a heterozygote rate less than the standard sequencing error rate of 0.01 were retained as X chromosome candidates.

Gene Expression Analysis

To examine differential gene expression and differential transcript usage among life stage (egg, larvae, and adult), as well as female and male adults, we mapped paired-end short-read RNAseq data to the super-transcriptome using Hisat2 to generate input count matrix for genes and exons within genes. As each RNAseq sample is unreplicated, we chose statistical tools that have performed reliably in differential expression and splicing analyses (Mehmood *et al.* 2019; Rapaport*et al.* 2013) and applied stringent significance thresholds based on multiple testing criteria. First, we used edgeR (Robinson *et al.* 2010) to conduct the differential expression analysis and assumed a recommended dispersion value for within species comparisons (0.16, based on human datasets). The count matrix was filtered by requiring a minimum count of 1 read-per-million per gene in at least three samples. Hierarchical clustering was used to measure the overall sample to sample distance based on differential expression, with read data transformed to log2 counts-per-million (logCPM), or moderated log-counts-per-million, to avoid the effect of undefined values and to shrink poorly expressed genes with low counts to zero. We then identified significantly differentially expressed genes in pairwise

contrasts of each life-stage or sex, using Fisher's Exact Test based on the negative-binomial distribution. Statistical significance in each contrast was determined based on a Benjamini-Hochberg adjusted*p-value* , or false discovery rate (FDR), of 0.01. Gene lists for the top 25 genes (lowest *p-value*) from each pairwise contrast were combined to generate a heatmap using the pheatmap package in R. For the egg vs. larvae, larvae vs. adult female, and adult female vs. male contrasts, we examined the biological function for all significantly differentially expressed genes (FDR < 0.01) using a gene set enrichment analysis (GSEA) based on gene ontology (GO) terms. The goseq package (Young *et al.* 2010) in R was used to identify GO terms that were enriched by comparing the subset of differentially expressed genes to all genes that were expressed in either the female or male samples. The Wallenius approximation was then used to calculate statistical significance based on an FDR of 0.05, while taking gene length into account using an empirically estimated probability weighting function.

We conducted another GSEA analysis based on genes exhibiting differential exon usage in the female vs. male contrast. A generalized log-linear model was fit to the exon count data, and log-fold differences among exons in each multi-exon gene were assessed using the*diffSplice* function in the limma package in R (Ritchie *et al.* 2015). Differences in expression among exons were first assessed using t-tests, and then the Simes method was used to provide a multiple testing correction for a gene-wise measure of significance. We used the Simes *p-value* with a significance threshold of 0.01 to generate a gene list for GSEA analysis, and the background list included all multi-exon genes.

Phylogeny and Gene Family Size Analysis

For the gene family expansion/contraction analysis, we first reconstructed a phylogenomic tree for *N. riversi* and the 11 beetle species with genomes in RefSeq (O'Leary *et al.*2016), as well as *Drosophila melanogaster* (Insecta: Diptera, hereafter Dmel) as an outgroup for the Coleoptera. The latest genomes were used to identify single copy orthologs using BUSCO v2.0 with the Endopterygota reference dataset. More than two thousand single copy orthologs were discovered from the 13 genomes. We then used MAFFT v7.471 (Katoh & Standley 2013) to align the amino acid (aa) sequences of 13 species for each ortholog, and found 1,017 orthologs had sequences in all 13 species. In order to reduce any effects of misalignment, we used trimAl v1.2 (Capella-Gutiérrez *et al.* 2009) to remove ambiguous bases and gaps. This trimming process reduced the total alignment size from 1,021,369 (including gaps) to 382,948 aas. We then used the R package kdetrees (Weyenberg *et al.*2014) to detect any gene tree outliers based on both topological similarity and branch length similarity to all gene trees. After removing 14 outliers, we concatenated the gene-specific alignments into a single alignment (1,013 orthologous genes, with 379,463 aas in total) and constructed a Maximum Likelihood (ML) tree using RAxML v8.2.12 (Stamatakis 2014). In RAxML, we applied the PROTGAMMAAUTO substitution model and designated *D. melanogaster*as the outgroup, and the branch support was evaluated with 100 bootstrap replicates. The final ultrametric tree was generated by calibrating the ML tree with fossil records using treePL (Smith & O'Meara 2012), with parameters obtained from the prime function. For fossil records, we used the estimated age of Tshekardocoleidae (oldest known beetle, 295 Mya) to define the minimum age of the most recent common ancestor of all Coleopteran species (Kirejtshuk 2020). The minimum estimation of crown Coleoptera (208.5 Mya) was used to define the minimum age of Polyphaga (Wolfe *et al.* 2016). The oldest Chrysomelidae fossil (*Mesolpinus* , 122 Mya) was used to define the minimum age of Chrysomeloidea (Hu *et al.* 2020; Kirejtshuk*et al.* 2015), whereas the fossil weevil (Nemonychidae, 157 Mya) was used to define the maximum age of Chrysomeloidea (Behrensmeyer & Turner 2013).

With the resulting ultrametric tree, we set out to analyze changes in the size of gene families in *N. riversi* relative to other beetles, while accounting for phylogenetic history, using CAFE v3.0 (Ganote *et al.* 2018). We first identified all one to one orthologous genes among the twelve beetle species and Dmel (as outgroup) using OMA standalone v2.4.1 (Altenhoff *et al.* 2019). The input dataset included the precomputed all-against-all genomes of Tcas and Dpon from the OMA database, and the proteome sequences of the remaining ten beetle species and Dmel from RefSeq (O'Leary*et al.* 2016). This resulted in a count matrix of the number of genes in each hierarchical orthologous group (HOG). Among the 18,438 HOGs, we dropped one HOG (HOG06033)

that had more than 200 gene copies to allow CAFE to find the optimized λ (rate of evolutionary change). We identified all rapidly expanding and contracting HOGs using a p-value threshold < 0.05. In addition, we relaxed this threshold to consider all expanding HOGs on the branch of *N. riversi* , and conducted a GSEA test using clusterProfiler (Yu *et al.* 2012) in R. We examined two levels of cutoff (0.05 and 0.1) for Benjamini-Hochberg adjusted p-values to identify enriched GO terms (Yu*et al.* 2012). The gene ontology terms associated with biological processes were further clustered based on semantic similarity using the Revigo webserver (Supek *et al.* 2011), using the*simRel* score to collapse redundant terms at 0.7 similarity and the UniProt database to determine the gene ontology term abundance, and visualized using a CirGO plot (Kuznetsova*et al.* 2019).

Metagenomic Analysis of Microbes

To better understand the microbiome associated with *N. riversi* , we first examined the taxonomic assignment for the most abundant microbial reads using Autometa (Miller *et al.* 2019). To generate an input coverage table for Autometa, we used Minimap2 to calculate coverage for the long-read data against the CANU assembly. Then, we ran the Autometa pipeline using a contig length cutoff of 500 bp. While we ran this program with all reads (beetle and non-targeted reads), Autometa separates eukaryotic reads from bacterial reads during the clustering step. We then used taxonomic identification and clustering to identify four unique bins (*Spiroplasma, Enterobacter, Acinetobacter, Rhodococcus*). Using BUSCO v4.14 and CheckM (Parks *et al.* 2015), we assessed the quality of the four bins. Specifically, we ran BUSCO in genome mode using the auto lineage selector (–auto-lineage-prok) and we used the lineage_wf CheckM command. From this analysis we determined that the contigs matching to *Spiroplasma* had high enough quality to continue with downstream functional analyses. The *Spiroplasma*sp. NR genome assembly has been made publicly available on NCBI (WGS Accession: CP065192-CP065193).

To understand the function of *Spiroplasma* in its host, we examined its phylogenetic placement relative to other *Spiroplasma*taxa, annotated its genome for functional elements, and also examined whether it was transcriptionally active at different life-stages in the RNAseq data. Previous studies have shown that *Spiroplasma* can play a role in host defense against nematodes and parasitoids through ribosome-inactivating proteins (RIPs, Ballinger & Perlman 2019), or conversely act as a reproductive parasite when the male-killing*spaid* gene is present (Harumoto & Lemaitre 2018). First, the phylogenetic relationship of *Spiroplasma* sp. NR was assessed using the 16S rRNA gene. A total of 38 *Spiroplasma* accessions were obtained from GenBank, and 1,286 bp of the 16S gene were aligned with MUSCLE v3.8.31 (Edgar 2004). An evolutionary tree was inferred using Maximum Likelihood in RAxML (Stamatakis 2014), with the GTRCAT substitution model and 1,000 bootstrap replicates to estimate support values. Next, we annotated the *Spiroplasma*genome using prokka (Seemann 2014), with settings for the Bacteria and genetic code 4. To further functionally annotate the genome, we performed several annotations with 1) the KEGG Automatic Annotation Server, 2) antiSMASH v5.0 (Blin *et al.* 2019), 3) CRISPRfinder (Grissa *et al.* 2007), and 4) DIAMOND blastx against custom databases and the UniProt SwissProt and TrEMBL databases (accessed September 8, 2020; O'Donovan *et al.* 2002).We kept only the top 5% of hits (–top 5) that had an E value below 1e-05 for each query sequence. We searched for UniProt accession numbers corresponding to 156 RIPs and the Ankyrin-rich *Spaid* protein from *S. poulsonii* (Accession A0A2R6Y5P0). Due to ambiguous results, we also used 1) blastn and blastx against the *spaid* nucleotide and peptide sequences on GenBank (Accessions MG837001.1 and AWJ64280.1), as well as 2) metabolisHMM (McDaniel *et al.* 2019) searched using the Ankyrin repeat (PF00023) and OTU-like (PF02338) Pfam hmmer models. The resulting hits were compared to NCBI's non-redundant nucleotide database using blastx. Finally, to examine if*Spiroplasma* was transcriptionally active at different life-stages and in different sexes, we used the Kaiju webserver (Menzel*et al.* 2016) to assign taxonomic classifications to microbial reads, using default parameters. We recorded the number of reads mapping to domain Bacteria, class Mollicutes, and genus *Spiroplasma* . In addition, we used kallisto v0.43.1 (Bray *et al.* 2016) to pseudo-align all non-beetle reads against the *Spiroplasma* genome from this study.

Results

Transcriptome and Mitochondrial Genome Assemblies

The assembled transcriptome was 297.65 Mb, encompassing 129,804 genes and 201,715 transcripts (**Table 2)** . Based on blast searches against full length reference proteins in the UniProt-SwissProt database, the transcriptome had 3,985 transcripts with >80% and 2,869 transcripts >90% alignment coverage on reference proteins. The BUSCO assessment of 2,442 single copy orthologs shows 97.4% completeness (complete and single copy =19.2%, complete and duplicated = 78.2%, fragmented = 1.6%, and missing = 1.0%). The high number of duplicated genes is due to the presence of several isoforms in the trinity output. When only the longest isoform per gene is used, the transcriptome size is 100.02 Mb (contig N50 = 3,991 bp, with an average and median contig length of 1,475 and 489 bp, respectively).

The mitochondrial genome of *N. riversi* is 15.8 Kb, with a full complement of 37 genes (**Figure S1** and **Table S2**). This is substantially smaller than the published mitochondrial genome of*N. brevicollis* (21.2 Kb), which has a large expansion of non-coding DNA near the replication origin. However, the mitochondrial genome of *N. riversi* is similar in size to the majority of Coleopteran mitochondrial genomes, and the gene order is largely conserved, except for a novel inversion with respect to other Carabidae that encompasses the *nad2, trnI, trnM* and *trnW* genes adjacent to the replication origin.

Genome Assembly, Annotation, Structural Comparison and Synteny

The draft long-read CANU assembly resulted in a genome size of 189 Mb, comprised of 2,098 contigs. However, a proportion of the genome was identified as non-target sequence (**Figure S** 2, **Table S3**and **Table S4**), principally as Proteobacteria and Tenericutes, and removal of these reads as well as mitochondrial reads reduced the long-read dataset by 44.6 Mb. After non-target sequence removal, a hybrid assembly was generated using both the long-read and short-read data. Different long-read coverage thresholds were explored (10x, 20x, and 25x), with genome completeness assessed using BUSCO. The 20x coverage assembly performed best according to BUSCO, with 95.0% completeness (complete and single copy = 94.8%, complete and duplicated = 0.2%, fragmented = 2.2%, and missing = 2.8%, out of 2,442 single copy orthologs), and resulted in 2,137 contigs summing to 147.3 Mb in total size (average length 69.0 Kb and max length 1.28 Mb). Alternative haslr assemblies resulted in a reduction of BUSCO completeness (93.6% for 10x and 94.8% for 25x), with increases in the missing single copy orthologs. The 20x coverage genome was then scaffolded using RNAseq pair-end read data, which resulted in 1,636 scaffolds and 147.4 Mb genome size (average scaffold length 90.1 Kb and max length 1.37 Mb). This final genome assembly had an improved BUSCO score with 95.3% completeness (complete and single copy = 95.1%, complete and duplicated = 0.2%, fragmented = 2.1%, and missing = 2.6%).

The genome was then annotated using automated prediction, database searches and RNAseq data. This resulted in 17,895 genes and 23,973 proteins (**Table 2**), as well as 150,583 repeat regions (comprising 201,059 copies in 19 families, **Table S5**). The official gene set has a 93.3% BUSCO completeness score (complete and single copy = 80.5%, complete and duplicated = 12.8%, fragmented = 2.4%, and missing = 4.3%). The genome size of *N. riversi* is smaller and less fragmented than the 11 Coleoptera species in RefSeq (**Table 1; Table S6; Figure S3**), yet the number of genes falls within the range of published gene sets. Similarly, the repeat content of *N. riversi* is falls within the observed range of other Coleoptera species, albeit at the low end (**Table 1**). On exceptional difference involves the distribution of intron length in*N. riversi* , which is truncated compared to other beetle species in having a statistically significant reduction of larger introns (> 1,000 bp) and consequently smaller total size of intronic regions (**Figure S4**). Finally, an analysis of collinear genes across *N. riversi* and six other beetle genomes (**Figure S5**) shows that estimates of the number of collinear genes and the proportion of collinear genes are impacted by genome fragmentation. Focusing only on comparisons involving *T. castaneum* (the most contiguous assembly), modest synteny is found across Coleoptera, with roughly 2000 genes (6% of the total) showing collinearity in *N. riversi* .

Identification of the X Chromosome

We assigned six scaffolds (scaffold 53, 93, 316, 1218, 1230, and 2067) to the X chromosome based on a combination of coverage and SNP genotyping (**Table S7**), but did not observe synteny when comparing

the *N. riversi* X chromosome scaffolds to the *T. castaneum* X chromosome.

Differential Gene Expression

To examine differential expression among life-stages and adult sexes of*N. riversi* , we mapped RNA sequencing reads to the super-transcriptome that merged the official gene set with the *de novo* transcriptome. This super-transcriptome reference contains 17,736 genes (**Table 2**), a slight reduction relative to the official gene set. Hierarchical clustering of overall expression differences (**Figure S6**) shows that the egg stage is very strongly differentiated, relative to comparisons among the larval and adult stages. Focusing on the top 25 differentially expressed genes among each pairwise contrast (**Figure S7**), both stage and sex-specific gene expression is evident. For example, some of the top genes expressed in the larval stage are involved in cuticle formation (cuticle, endocuticle and chitin binding proteins), while some of the top genes in the adult male are associated with sperm production. These patterns are reinforced by a gene set enrichment analysis (GSEA) of the associated gene ontology terms for differentially expressed genes (FDR < 0.01), with statistical enrichment determined below a nominal p -value of 0.05 (**Table S8**). The larvae to egg contrasts highlighted terms upregulated in larvae associated with extracellular space, monooxygenase activity, chitin binding and metabolism, fatty acid synthesis, biological adhesion, and muscle fibers. The adult female to larvae contrasts highlighted terms associated with in chitin metabolism and extracellular space, while terms upregulated only in the larvae additionally highlighted gene ontology terms involved in protein digestion and sugar metabolism. The contrast of the adult male and female *N. riversi* showed an enrichment of terms upregulated in males involving sperm production, such as the axonemal dynein complex assembly and associated terms involving the cytoskeleton and cell motility. We examined the contrast between female and male adult beetles for differential exon usage (**Table S9**), and found that among the 7,864 genes with multiple exons, 217 showed significant differences (FDR < 0.01), with the GSEA of the gene ontology terms showing significant enrichment (FDR < 0.05) of 14 GO categories, seven of which are involved in sperm structure and function.

Phylogenetic Relationships and Gene Family Evolution

We estimated a coleopteran phylogeny with high support, with *N. riversi* as the inferred outgroup (**Figure 1**). Fossil-based time calibrations suggest a divergence time of ˜300 million years. The CAFÉ analysis of gene family turnover shows at least seven large lineage-specific rapid gene expansion events in Polyphaga (Figure 1), but a very modest expansion in *N. riversi* . Most gene contractions are quite small across the Coleoptera phylogeny. An examination of the *N. riversi* orthogroup expansion events (**Table 3**) shows expansions in zinc finger proteins, trehalose transporters, protein kinases, troponins, multidrug resistance-associated proteins, and synaptic vesile glycoproteins. Contracting orthogroups in *N. riversi* (**Table S10**) include ligases, ion channel receptors, glutathione S-transferases, nuclear receptor coactivators, aminophospholipid transporters, regucalcin proteins, and dynein complex proteins.

Based on a conservative GSEA (p-value=0.001) of expanded gene families in *N. riversi* , functionally enriched terms include pathways involved in protein transport, protein phosphorylation, metabolic processes, proteolysis (ubiquitination and proteasome), organismal and brain development, response to heat and synaptic signaling (**Table S11; Figure 2**).

Microbial Associations

The four metagenomic bins identified using Autometa match to the bacterial genera *Spiroplasma, Acinetobacter, Enterobacter,*and *Rhodococcus.* A BUSCO analysis to confirm completeness indicated that the *Acinetobacter, Enterobacter,* and*Rhodococcus* genomes were not high quality (3.2%, 0%, and 0% complete, respectively) to continue downstream analysis. The draft*Spiroplasma* sp. NR genome contained one 1,414,835bp genomic contig and a 28,794bp plasmid, for a total length of 1,443,729bp, an average GC content of 26.08%, 2,085 open reading frames, and 29 tRNA encoding genes. The BUSCO analysis of 151 markers conserved among Mollicutes indicated that the *S.* sp. NR contained 148 complete and single-copy BUSCOs (98%), one fragmented BUSCO (0.7%), and two missing BUSCOs (1.3%). A CheckM analysis with 228 markers in the Mollicutes lineage found similar quality, with 97.12% of the genome being complete.

To determine the relationship of *S.* sp. NR to other*Spiroplasma* species, we generated a ML tree for the16S rRNA gene from 38 *Spiroplasma* isolates, spanning the known major clades of*Spiroplasma* (**Figure 3A**). *Spiroplasma* sp. NR and three isolates from *Harpalus pensylvanicus* (Coleoptera: Carabidae) form a monophyletic group within the lineage of*Spiroplasma* associated with insects. Annotation with KEGG indicates that *S.* sp. NR can import and phosphorylate sugars via a phosphoenolpyruvate phosphotransferase system and can use G6P and fructose-6P through a complete glycolytic pathway and near-complete pentose phosphate pathway (**Figure 3B; Figure S8**). Similar to other *Spiroplasma* species, *S.* sp. NR is predicted to produce lactate, acetate, and propanoate. *Spiroplasma* sp. NR is capable of oxidative phosphorylation as it encodes an F-type ATPase, and it contains most of the genes (SecD/F, SecY, SecG, YidC, SecA, FtsY, ffh) that encode the preprotein translocation Sec-SRP system, although its genome is missing the gene encoding the essential stabilizing SecE component. Notably, *S.* sp. NR does not have the amino acid biosynthesis pathways or genes for the citric acid cycle, which has been observed in other *Spiroplasma* species. Annotations using antiSMASH and CRISPrfinder indicated the absence of biosynthetic gene clusters and CRISPR regions. No RIPs were detected in*S.* sp. NR, but there is evidence of loci similar to the male-killing gene, *spaid* . A total of 23 ORFs exhibited significant alignment (E-value <1E-05) to *spaid* , but only one ORF remained as a significant alignment when the TrEMBL database was used. This ORF alignment has only 10% query coverage and 49.57% identity to *spaid* . Using a HMM based approach, four ORFs significantly aligned to the ankyrin-repeat HMM marker and one ORF significantly aligned to the Ovarian Tumour-like deubiquitinase domain (OTU) HMM marker, with no overlap between the ORFs. Both markers are characteristic of *spaid.* Finally, analysis of RNAseq data from *N. riversi* revealed the presence of expressed*Spiroplasma* reads, supporting transcriptional activity of*Spiroplasma* in the egg, larval, and adult life stages, for both the adult female and male (**Table S12**).

Discussion

Nebria riversi has a smaller genome than most sequenced beetles to date, which is associated with a reduction in transposable element content and a truncation in the distribution of intron length. A comparison of genome structure shows that the number of predicted genes remains within the range of other beetles and the number of collinear genes is similar (but modest) across sequenced coleopteran lineages. These results, in addition to patterns of gene family expansion, are primarily associated with molecular changes involved in cold specialization (see below). In contrast, there is more limited evidence of genome novelty in Carabidae, relative to other Coleoptera. Rapidly expanding ortholog groups associated with DNA and RNA binding suggest evolutionary divergence in gene regulation, but this is not unexpected given the long evolutionary history (˜300 Myr) separating Carabidae from other Coleoptera (Schmitz *et al.* 2016). The expansion of the troponin orthogroup might be linked to muscle fiber performance of predatory ground beetles, but also could be linked to cold adaptation (see below). Contracting gene families (**Table S7**), especially the glutathione S-transferase genes, and ion channel genes, but potentially the genes linked to monooxygenase activity and membrane transport, likely distinguish the predatory lifestyle of ground beetles from the herbivorous lifestyle of most Polyphaga (Seppey*et al.* 2019a). These gene families are known to play a role in detoxifying plant secondary chemicals, and likely diversified in Polyphaga as a result of the plant-insect coevolutionary arms race (Ehrlich & Raven 1964). Consistent with the transcriptomic analysis of Seppey *et al.* (2019a), many of the adaptive lineage-specific gene expansions in beetles (**Figure 1**) occur in Polyphaga, rather than Adephaga, and involve plant dietary innovations.

Genome Structure and Signatures of Cold Specialization in Beetles

Despite many attempts to find abiotic and biotic factors that contribute to the enormous natural variation in genome size, it remains an open question which evolutionary and ecological variables drive genome size changes (Blommaert 2020). The main challenges to resolving this question arise from the complexity of genome composition, decoupling the role of selection (if present) from many different environmental factors, and the dependence of genome size on phylogenetic relatedness (Alfsnes*et al.* 2017; Canapa *et al.* 2020; Ritchie *et al.*2017). Recent studies have shown that the genome size is positively correlated with the size of repetitive DNA in insects (Quesneville 2015), but repetitive DNA has been linked to both adaptive and non-

adaptive evolutionary processes (Canapa *et al.* 2015; Lower*et al.* 2017). To date, there has been limited effort to document evidence for a statistical association between genome size and environmental factors in insects (Alfsnes *et al.* 2017). One exception involves the Antarctic midge, *Belgica antarctica,* which is endemic to Antarctica and has (to date) the smallest genome of any insect species (99 MB), and very little TE content in the genome (Kelley*et al.* 2014). In concordance with the case of *B. antarctica* , we show that the cold-adapted alpine ground beetle *N. riversi* has the smallest assembled genome among sequenced beetles, including eight species of Adephaga and 11 species of Polyphaga (**Table S6**). This is associated with a smaller size range of introns and less TE content (**Table 1**), and may support the hypothesis that smaller genomes are associated with adaptation to extreme environments such as polar or alpine climates (Kelley *et al.* 2014). A possible adaptive driver of this pattern is that smaller genome size results in bioenergetic savings (Wagner 2005; Wright*et al.* 2014), such as a more efficient cell cycling, transcription and metabolism, allowing species to develop faster and maximize the efficiency of cellular processes. Such changes might compensate for the relatively shorter growth season in the cold environments, as well as the fact that low temperatures elongate cell cycle duration and slow cellular metabolism in ectotherms (Vinogradov 1999). However, the relationship between genome size and cold temperature is reversed in crustaceans and many vertebrates (Canapa*et al.* 2020), posing something of a conundrum. Other cold-specialized insects also defy this pattern, such as the alpine grasshopper *Gomphocerus sibiricus* , which has a spectacularly large 8.95 Gb genome (Gosalvez *et al.* 1980). Some have argued that different strategies might be employed depending on the developmental biology of organisms (Alfsnes *et al.* 2017). In insects, previous research has shown that holometabolous insects (with complete metamorphosis) have smaller genomes than hemimetabolous insects (Alfsnes *et al.* 2017), suggesting that the complex developmental changes in metamorphosis are an added factor constraining genome size. Clearly, more comparative genomics research, preferably sampling variable genome sizes among more closely related taxa, is needed to test these hypotheses.

Gene Family Novelty and Signatures of Cold Specialization in Beetles

Diverse insect lineages have evolved to cope with cold conditions, and strategies vary; insect taxa are frequently described as freeze avoidant or freeze tolerant, with further subdivision of the freeze avoidant species into chill susceptible or chill tolerant species (Bale 2002).*Nebria* (*Catonebria*) appears to be chill tolerant (Slatyer & Schoville 2016), as it maintains neuromuscular activity below freezing, has modest mortality as it approaches its supercooling point, and yet recovers within minutes from its critical thermal minimum. However, this categorization fails to capture the fact that *N. riversi* is specialized to perform at cold temperatures, as it forages nocturnally at sub-zero temperatures in high elevation riparian habitats that are fed by ice and snowmelt. *Nebria riversi* must grow, develop and reproduce under constantly cold environmental conditions, and thus its fitness depends on permanently altering molecular pathways to function efficiently at cold temperature. A number of key physiological adaptations are expected (Clark & Worland 2008; Teets & Denlinger 2013), including 1) shifts to maintain cellular ion homeostasis (ion transport mechanisms), 2) regulation of cellular metabolism (mitochondrial performance and ATP synthesis), 3) efficient removal of toxic byproducts (detoxification of reactive oxygen species), 4) homeoviscous adaptation to maintain the liquid crystalline phase of cell membranes (an increase in unsaturated fatty acid content in membrane lipids), 5) regulation of protein denaturation (heat shock response, ubiquitination and proteolysis), and 6) neuromuscular adaptation. Despite an extensive literature and considerable knowledge of the physiological strategies insects employ to tolerate cold (see reviews in: Overgaard & MacMillan 2017; Teets & Denlinger 2013), comparative genomic and transcriptomic analyses of cold specialized insects remains limited (Cicconardi *et al.* 2020; Dennis *et al.* 2015; Kelley *et al.* 2014; Kim *et al.* 2017; Schoville*et al.* 2020).

One fundamental strategy for ectotherms to maintain the physical function in cold environments is to increase metabolic rate (Addo-Bediako *et al.* 2002; Williams *et al.* 2016). This involves both energy generation and management of subsequent oxidative stress caused by the accumulation of reactive oxygen species (ROS), which cause significant damage to nucleic acids, lipids, and proteins, as well as cellular structures like the plasma membrane (Storey & Storey 2010). The detoxification of ROS is thus an important requirement in coping with cold environmental conditions (Lalouette *et al.*2011). Several studies have shown that many cellular organelles are involved in dealing with this oxidative stress, including the endoplasmic reticulum,

Golgi apparatus, and the lysosome (Butler & Bahr 2006; Jiang *et al.* 2011; Malhotra & Kaufman 2007; Pascua-Maestro*et al.* 2017). In *N. riversi* , we found enriched GO terms for expanding gene families associated with oxidative stress, including endoplasmic reticulum (GO:0005783), intracellular membrane-bounded organelle (GO:0043231), regulation of cellular metabolic process (GO:0031323), Golgi apparatus (GO:0005794), lysosomal membrane (GO:0005765), anterograde synaptic vesicle transport (GO:0048490), and carbohydrate phosphorylation (GO:0046835).

Cold acclimated insects also need to cope with damage to proteins, cells and tissues, and often compensate by modification of canonical temperature stress pathways (Lyzenga & Stone 2012). In *N. riversi* , we found numerous enriched GO terms that are linked to degradation of damaged proteins and DNA repair, including response to heat (*i.e* . heat shock proteins involved in protein folding; GO:0009408), protein polyubiquitination (GO:0000209), ubiquitin protein ligase binding (GO:0031625), proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161), and protein K48-linked ubiquitination (GO:0070936). This finding is concordant with other studies that focus on other cold adapted insect species (Dennis *et al.* 2015; Kim *et al.* 2017; Schoville *et al.* 2020). The most severe damage to cells results from intracellular freezing, which results in unrepairable damage to cellular membranes in freeze-intolerant species (Bale 1987). One of the common strategies insects use to overcome freezing temperature is to increase the concentration of solutes in hemolymph and thereby lower the ice nucleation point (Storey & Storey 2012; Teets *et al.* 2011; Teets*et al.* 2019). In *N. riversi* , we found that genes associated with trehalose transport have rapidly expanded, and note that this sugar is widely implicated in cold tolerance among insects as it readily increases blood sugar content (Khani *et al.* 2007; Sinclair *et al.* 2003). Similarly, the enriched GO term phosphatidylinositol binding (GO:0035091) is involved in ice nucleation (Neven *et al.* 1989), which limits supercooling and prevents intracellular damage from ice formation. This pathway often includes proteins (i.e. protein ice nucleators, PINS, and antifreeze proteins, AFPs) that are known to cooperate with trehalose in freeze avoidance in beetles (Duman 2001).

Finally, for an insect species like *N. riversi* to be active in a constantly subzero environment, an important challenge is to maintain neuromusculatory activity for foraging, mating, and predator avoidance behaviors. Indeed, several studies have shown that species distributed in cooler areas have lower critical thermal minimum (CT_{min}) compared to closely related species living in warmer environments (Bishop *et al.* 2017; Slatyer *et al.*2016). In *N. riversi* , we found rapid expansion of troponin I genes compared to other coleopteran species, which is a cardiac and skeletal muscle protein that serves an inhibitory function to actomyosin ATPase and muscular contraction, but also prevents muscle damage during exercise (Farah *et al.* 1994; Fishilevich *et al.* 2019). Troponin I is typically single copy in most insects, and increasing the number of protein isoforms is expected to increase muscle fiber type specificity and/or developmental regulation of muscle development. Amino acid changes in troponin I (and other members of the troponin complex) have been linked to improved muscle performance under cold conditions (Shaffer & Gillis 2010).

Among the expected changes, two categories were not enriched in gene family expansion events. Cellular ion homeostasis, although it is also crucial for ensuring the function of neuronal signal transduction and muscle contraction (Lubawy *et al.* 2020; MacMillan *et al.*2014; Pivovarov *et al.* 2019), was not associated with gene family expansions in *N. riversi* . Similarly, changes to the structure of phospholipid bilayer, which is a common strategy to avoid damage to the cytoplasmic membrane and internal membranes (Enriquez & Colinet 2019; Holmstrup *et al.* 2002), was not found among the gene expansion events. There are three possible reasons we do not find evidence for these physiological changes. First, both mechanisms are often seen in overwintering insects or following laboratory acclimation treatments (Danks 2005; Enriquez & Colinet 2019; Parker *et al.* 2018), so may instead be associated with plastic gene regulatory responses to cold instead of turnover in gene families. Second, additional evidence of cold adaptation might be discovered in *N. riversi* with improved taxon sampling. As *N. riversi* sits on a long branch in our Coleoptera tree, the power to distinguish smaller gene expansion events is limited, because gene expansion tests must adjust for the rate of evolutionary change. Third, it is notable that insect species that lack signatures of these two physiological responses include the two Antarctic midges (Kelley *et al.* 2014; Kim *et al.* 2017), and like *N. riversi* , they live in constantly cold habitats. Another mechanism of cold specialization, which we did not evaluate

here, likely arises from selection on protein-coding genes. Structural changes to protein coding genes could act to permanently alter cellular properties involving homeostasis and membrane viscosity in these species, and has been seen in several cold-specialized insects (Dennis*et al.* 2015; Parker *et al.* 2018; Schoville *et al.*2020). Since *N. riversi* sits on a long branch in our Coleoptera tree, tests for positive selection, such as the branch-site test, will lack power due to saturation of synonymous sites and the occurrence of multi-nucleotide substitutions (Gharib & Robinson-Rechavi 2013; Venkat*et al.* 2018). Therefore, sampling the genomes of species of Carabidae or Adephaga, especially groups with contrasting cold and warm specialized species, are needed to increase our ability to detect signatures of temperature adaptation.

Microbial Associates

Bacterial symbionts have been implicated in a range of important roles in insect hosts. In Carabidae, *Spiroplasma* are known from at least two subfamilies (Lundgren *et al.* 2007; McManus *et al.* 2018). However, little research to date has focused on the microbiome of alpine insects (but see: Receveur *et al.* 2020), let alone documented the role of endosymbionts. *Spiroplasma* is a common symbiont in insects, with a recent estimation of infection in 8% of arthropod species, although it is rarely detected in beetles (Medina*et al.* 2019). *Spiroplasma* spp. has been observed to have a range of effects in their hosts, including male-killing and protection from parasites, such as wasps and nematodes (Ammar & Hogenhout 2006; Ballinger & Perlman 2019; Harumoto & Lemaitre 2018; Haselkorn & Jaenike 2015; Xie *et al.* 2010). In this study, a complete*Spiroplasma* genome was recovered from PacBio sequencing of*N. riversi* . This finding coupled with the presence of*Spiroplasma* reads in the RNAseq datasets of *N. riversi*egg, larvae, and adult specimens, imply the prevalence of*Spiroplasma* in *N. riversi* . The detection of*Spiroplasma* in egg sample also suggests the endosymbiosis and vertical transmission of this bacterium in *N. riversi* . However, additional work should characterize the distribution and prevalence of*Spiroplamsa* in *N. riversi* . Furthermore, our phylogenetic analyses places *S.* sp. NR in Carabid-specific clade of*Spiroplasma* isolates. These results suggest that*Spiroplasma* may be widespread in Carabidae. Future work that characterizes the prevalence of *S.* sp. NR in *N. riversi* , and related *Nebria* , would help elucidate its role as an endosymbiont.

There are several possible ecological roles that the *Spiroplasma*might play in its host. First, analysis of this genome revealed the presence of ankyrin repeats and the OTU domain, which could indicate male-killing and a role as a reproductive parasite (Harumoto & Lemaitre 2018). However, homology to the *spaid* gene locus was partial, and ankyrin repeats are a common protein-protein interaction motif occurring in proteins with diverse function. No single ORF contained both ankyrin repeats and the OTU domain; however, recent work has demonstrated that *spaid* is a rapidly evolving locus and*spaid*- like genes containing only the OTU domain were identified in *Spiroplasma* and *Wolbachia* symbionts (Gerth *et al.* 2020). Our ecological survey data supports a female-skewed sex ratio in *N. riversi* (**Figure S9**), as might be expected from a long-term infection with a male-killing parasite. Second, because we have observed nematode infections (*Gordius* sp., Weng and Schoville, personal observation) in *N. riversi* ,*Spiroplasma* might have a host defensive role. This role was suggested for another riparian carabid species with *Spiroplasma* ,*Brachinus elongatulus,* which is frequently observed to have nematode infections (McManus *et al.* 2018). However, we did not detect the presence of RIPs in *S.* sp. NR, which would provide stronger evidence of host defense against nematodes. Third,*Spiroplasma* might play a role in the production of useful metabolites, which is suggested by the presence of propanoate biosynthesis genes. One possible role of propanoate (also known as propanoic or propionic acid) is serving as a constituent of the reproductive pheromones or defensive chemistry of *N. riversi*(Attygalle *et al.* 1992; Francke & Dettner 2005). Propanoate has been found in the defensive gland of at least one carabid beetle (Attygalle *et al.* 1992), and is considered to have insecticidal properties (Krzyżowski *et al.* 2020). While it has a pungent smell as a fatty acid volatile, it is also thought to facilitate biosynthesis of strong-smelling benzoquinones (Rocha *et al.* 2013). Notably,*N. riversi* seems to complement this pathway (**Figure S8**) by encoding the genes to produce Propanoyl-CoA, which *Spiroplasma*is predicted to transform into propanoate through propanoyl phosphate.*Nebria riversi* has an alternative path to producing propanoate, from propanoyl-CoA to propionyl adenylate, which implies there may be different derivatives or functions of propanoate synthesis by *S.*sp. NR and *N. riversi* . Fourth, and finally, evidence has shown that the presence of *Spiroplasma* could help insect hosts to survive over winter (Ebbert & Nault 1994). Future work,

including an assessment of the frequency and distribution of *Spiroplasma* among*N. riversi* populations, will be required to distinguish among these alternative endosymbiotic roles.

Acknowledgements

We thank Anna Graneiro, Lee Brady, and the UW Biotech Center for help with genomic library preparations and sequencing. We also thank Matthew Bolek and Ming-Chung Chiu for identifying the parasitic nematode of*N. riversi* . This study is partially supported by Ministry of Education of the Republic of China (Taiwan) by offering a scholarship to Y.M.W. Support was also provided by the Valentine Eastern Sierra Reserve (UC Santa Barbara Natural Reserve System) through a VESR Graduate Student Funding to Y.M.W. and a UW Research Competition Award to S.D.S.

Data Accessibility and Benefit Sharing Statement

All raw sequencing data, assemblies and the official gene set have been made publicly available on NCBI though BioProject PRJNA671697. Accession numbers are noted in the Methods section. Additional functional annotation data, input data files, and scripts are available on Dryad (https://doi.org/10.5061/dryad.8cz8w9gp5). All research described in the publication complies with U.S. national laws, and is consistent with the Convention on Biological Diversity and Nagoya Protocol agreements.

Author Contributions

This study was designed by Y.M.W. and S.D.S. The fieldwork was done by Y.M.W. and S.D.S. The data collection and analyses were conducted by Y.M.W., S.D.S., and C. B. F. All authors contributed to the writing of the manuscript, with the first draft written by Y.M.W., S.D.S., and C.B.F.

References

Adamski Z, Bufo SA, Chowański S*, et al.* (2019) Beetles as model organisms in physiological, biomedical and environmental studies–a review. *Frontiers in physiology* **10** , 319.

Addo-Bediako A, Chown S, Gaston K (2002) Metabolic cold adaptation in insects: a large-scale perspective. *Functional Ecology***16** , 332-338.

Alfsnes K, Leinaas HP, Hessen DO (2017) Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecology and evolution* **7** , 5939-5947.

Alqahtani F, Măndoiu I (2019) SMART2: Multi-library Statistical Mitogenome Assembly with Repeats, 184-198.

Altenhoff AM, Levy J, Zarowiecki M*, et al.* (2019) OMA standalone: orthology inference among public and custom genomes and transcriptomes.*Genome Research* **29** , 1152-1163.

Ammar E-D, Hogenhout SA (2006) Mollicutes associated with arthropods and plants. In: *Insect Symbiosis, Volume 2* (eds. Bourtzis K, Miller TA), pp. 119-140. CRC press, Boca Raton.

Attygalle AB, Meinwald J, Eisner T (1992) Defensive secretion of a carabid beetle,*Helluomorphoides clairvillei* . *Journal of Chemical Ecology* **18** , 489-498.

Bale J (1987) Insect cold hardiness: freezing and supercooling—an ecophysiological perspective. *Journal of Insect Physiology***33** , 899-908.

Bale J (2002) Insects and low temperatures: from molecular biology to distributions and abundance. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357** , 849-862.

Ballinger MJ, Perlman SJ (2019) The defensive *Spiroplasma* .*Current Opinion in Insect Science* **32** , 36-41.

Baust JG (1972) Insect Freezing Protection in *Pterostichus brevicornis* (Carabidae). *Nature New Biology* **236** , 219-221.

Behrensmeyer A, Turner A (2013) Taxonomic occurrences of Suidae recorded in the Paleobiology Database. Fossilworks.

Bernt M, Donath A, Jühling F*, et al.* (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69** , 313-319.

Beutel RG, Haas F (2000) Phylogenetic relationships of the suborders of Coleoptera (Insecta). *Cladistics* **16** , 103-141.

Bishop TR, Robertson MP, Van Rensburg BJ, Parr CL (2017) Coping with the cold: minimum temperatures and thermal tolerances dominate the ecology of mountain ants. *Ecological Entomology* **42** , 105-114.

Blin K, Shaw S, Steinke K*, et al.* (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **47** , W81-W87.

Blommaert J (2020) Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B* **287** , 20201441.

Bouchard P, Smith AB, Douglas H*, et al.* (2017) Biodiversity of Coleoptera. *Insect Biodiversity: Science and Society. John Wiley & Sons Ltd* , 337-417.

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology***34** , 525-527.

Brooks DR, Bater JE, Clark SJ*, et al.* (2012) Large carabid beetle declines in a United Kingdom monitoring network increases evidence for a widespread loss in insect biodiversity. *Journal of Applied Ecology* **49** , 1009-1019.

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12** , 59-60.

Bushnell B (2014) BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Butler D, Bahr BA (2006) Oxidative stress and lysosomes: CNS-related consequences and implications for lysosomal enhancement strategies and induction of autophagy. *Antioxid Redox Signal* **8** , 185-196.

Camacho C, Coulouris G, Avagyan V*, et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10** , 421.

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E (2015) Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research* **147** , 217-239.

Canapa A, Biscotti M, Barucca M*, et al.* (2020) Shedding light upon the complex net of genome size, genome composition and environment in chordates. *The European Zoological Journal* **87** , 192-202.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.*Bioinformatics* **25** , 1972-1973.

Chernov YI, Makarov K, Eremin P (2000) Family of ground beetles (Coleoptera, Carabidae) in the Arctic fauna: Communication 1.*Entomological Review* **80** , 1009-1021.

Cicconardi F, Krapf P, D'Annessa I*, et al.* (2020) Genomic signature of shifts in selection in a subalpine ant and its physiological adaptations. *Molecular Biology and Evolution***37** , 2211-2227.

Clark MS, Worland MR (2008) How insects survive the cold: molecular mechanisms—a review. *Journal of Comparative Physiology B***178** , 917-933.

Danks HV (1981) Arctic arthropods. *A review of systematics and ecology with particular reference to the north American fauna.***608** .

Danks HV (2005) Key themes in the study of seasonal adaptations in insects I. Patterns of cold hardiness. *Applied Entomology and Zoology* **40** , 199-211.

Danks HV, Smith AB, Foottit R, Adler P (2017) Insect biodiversity in the Nearctic region. *Insect biodiversity science and society***1** , 47-63.

Davidson NM, Oshlack A (2018) Necklace: combining reference and assembled transcriptomes for more comprehensive RNA-Seq analysis.*Gigascience* **7** , giy045.

Dennis AB, Dunning LT, Sinclair BJ, Buckley TR (2015) Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects.*Scientific reports* **5** , 13965-13965.

Duman JG (2001) Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annual Review of Physiology* **63** , 327-357.

Ebbert MA, Nault LR (1994) Improved overwintering ability in Dalbulus maidis (Homoptera: Cicadellidae) vectors infected with Spiroplasma kunkelii (Mycoplasmatales: Spiroplasmataceae). *Environmental Entomology* **23** , 634-644.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32** , 1792-1797.

Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* **18** , 586-608.

Enriquez T, Colinet H (2019) Cold acclimation triggers lipidomic and metabolic adjustments in the spotted wing drosophila. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **316** , R751-R763.

Erwin TL (1997) Biodiversity at its utmost: tropical forest beetles.*Biodiversity II: understanding and protecting our biological resources* , 27-40.

Farah CS, Miyamoto CA, Ramos C*, et al.* (1994) Structural and regulatory functions of the NH2-and COOH-terminal regions of skeletal muscle troponin I. *Journal of Biological Chemistry* **269** , 5230-5240.

Fishilevich E, Bowling AJ, Frey ML*, et al.* (2019) RNAi targeting of rootworm Troponin I transcripts confers root protection in maize.*Insect Biochemistry and Molecular Biology* **104** , 20-29.

Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA (2018) Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology* **18** , 21.

Francke W, Dettner K (2005) Chemical signalling in beetles. In:*The chemistry of pheromones and other semiochemicals II* , pp. 85-166. Springer.

Ganote C, Mendes F, Henschel R, Hahn M, Fulton B (2018) Introducing CAFE: Computational Analysis of (gene) Family Evolution.*Bioinformatics* **22** , 1269–1271.

Georgescu E, Rîșnoveanu L, Toader M*, et al.* (2017) Actual problems concerning protection of the wheat crops against cereal ground beetle (*Zabrus tenebrioides* Goeze) attack in south-east of the Romania. *Scientific Papers-Series A, Agronomy* **60** , 256-263.

Gerth M, Montoya HM, Ramirez P*, et al.* (2020) Rapid molecular evolution of Spiroplasma symbionts of Drosophila. *BioRxiv* .

Gharib WH, Robinson-Rechavi M (2013) The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution* **30** , 1675-1686.

Gosalvez J, López-Fernandez C, Esponda P (1980) Variability of the DNA content in five Orthopteran species. *Caryologia* **33** , 275-281.

Grimaldi D, Engel MS (2005) *Evolution of the Insects* Cambridge University Press.

Grissa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats.*Nucleic Acids Research* **35** , W52-W57.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics***29** , 1072-1075.

Gustafson GT, Baca SM, Alexander AM, Short AE (2019) Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Systematic Entomology* .

Haas BJ, Papanicolaou A, Yassour M*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols***8** , 1494.

Haas BJ, Salzberg SL, Zhu W*, et al.* (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9** , R7.

Haghshenas E, Asghari H, Stoye J, Chauve C, Hach F (2020) HASLR: Fast Hybrid Assembly of Long Reads. *iScience* **23** , 101389.

Harumoto T, Lemaitre B (2018) Male-killing toxin in a bacterial symbiont of *Drosophila* . *Nature* **557** , 252-255.

Haselkorn TS, Jaenike J (2015) Macroevolutionary persistence of heritable endosymbionts: acquisition, retention and expression of adaptive phenotypes in *Spiroplasma* . *Molecular Ecology***24** , 3752-3765.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. *Methods in Molecular Biology***1962** , 65-95.

Holmstrup M, Hedlund K, Boriss H (2002) Drought acclimation and lipid composition in *Folsomia candida* : implications for cold shock, heat shock and acute desiccation stress. *journal of insect physiology* **48** , 961-970.

Hozza M, Vinař T, Brejová B (2015) How big is that genome? Estimating genome size and coverage from k-mer abundance spectra, 199-209.

Hu L, Zhao T, Pan Y (2020) Spinicaudatans from the Yixian Formation (Lower Cretaceous) and the Daohugou Beds (Jurassic) of Western Liaoning, China. *Cretaceous Research* **105** , 104073.

Humann JL, Lee T, Ficklin S, Main D (2019) Structural and functional annotation of eukaryotic genomes with GenSAS. In: *Gene Prediction* , pp. 29-51. Springer.

Jiang Z, Hu Z, Zeng L*, et al.* (2011) The role of the Golgi apparatus in oxidative stress: is this organelle less significant than mitochondria? *Free Radical Biology and Medicine* **50** , 907-917.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30** , 772-780.

Kavanaugh DH (1985) On wing atrophy in carabid beetles (Coleoptera: Carabidae), with special reference to Nearctic *Nebria* . In:*Series Entomologica (Dordrecht)* (ed. Ball GE), pp. 408-431. Dr W. Junk Publishers, Dordrecht.

Kelley JL, Peyton JT, Fiston-Lavier A-S*, et al.* (2014) Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nature Communications* **5** , 1-8.

Khani A, Moharramipour S, Barzegar M (2007) Cold tolerance and trehalose accumulation in overwintering larvae of the codling moth, Cydia pomonella (Lepidoptera: Tortricidae). *European Journal of Entomology* **104** , 385.

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12** , 357.

Kim S, Oh M, Jung W*, et al.* (2017) Genome sequencing of the winged midge, *Parochlus steinenii* , from the Antarctic Peninsula.*Gigascience* **6** , 1-8.

Kimura MT (2004) Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia***140** , 442-449.

Kirejtshuk A, Moseyko A, Ren D (2015) Mesozoic leaf beetles of the tribe Mesolpinini trib. nov.(Coleoptera, Chrysomelidae) from the Lower Cretaceous. *Paleontological Journal* **49** , 1424-1435.

Kirejtshuk AG (2020) Taxonomic review of fossil Coleopterous families (Insecta, Coleoptera). Suborder Archostemata: Superfamilies Coleopseoidea and Cupedoidea. *Geosciences* **10** , 73.

Kolde R, Kolde MR (2015) Package 'pheatmap'. *R Package***1** , 790.

Koren S, Walenz BP, Berlin K*, et al.* (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27** , 722-736.

Kotze DJ, Brandmayr P, Casale A*, et al.* (2011) Forty years of carabid beetle research in Europe–from taxonomy, biology, ecology and population studies to bioindication, habitat assessment and conservation. *ZooKeys* , 55.

Krzyżowski M, Francikowski J, Baran B, Babczyńska A (2020) The short-chain fatty acids as potential protective agents against Callosobruchus maculatus infestation. *Journal of Stored Products Research* **86** , 101570.

Kuznetsova I, Lugmayr A, Siira SJ, Rackham O, Filipovska A (2019) CirGO: an alternative circular way of visualising gene ontology terms.*BMC Bioinformatics* **20** , 1-7.

Laetsch DR, Blaxter ML (2017) BlobTools: Interrogation of genome assemblies. *F1000Research* **6** , 1287.

Lalouette L, Williams CM, Hervant F, Sinclair BJ, Renault D (2011) Metabolic rate and oxidative stress in insects exposed to low temperature thermal fluctuations. *Comparative Biochemistry and Physiology - Part A Molecular & Integrative Physiology* **158** , 229-234.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* .

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences.*Bioinformatics* **34** , 3094-3100.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25** , 2078-2079.

Linard B, Arribas P, Andújar C, Crampton-Platt A, Vogler AP (2016) Lessons from genome skimming of arthropod-preserving ethanol.*Molecular Ecology Resources* **16** , 1365-1377.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm.*Nucleic Acids Research* **33** , 6494-6506.

Lopez-Lopez A, Vogler AP (2017) The mitogenome phylogeny of Adephaga (Coleoptera). *Molecular Phylogenetics and Evolution* **114** , 166-174.

Lorenz W (2020) CarabCat: Global database of ground beetles. In:*Species 2000 & ITIS Catalogue of Life, 2020-12-01* eds. Roskov Y, Ower G, Orrell T*, et al.*). Species 2000: Naturalis, Leiden, the Netherlands.

Lower SS, Johnston JS, Stanger-Hall KF*, et al.* (2017) Genome size in North American fireflies: substantial variation likely driven by neutral processes. *Genome Biology and Evolution* **9** , 1499-1512.

Lubawy J, Urbański A, Colinet H, Pflüger HJ, Marciniak P (2020) Role of the Insect Neuroendocrine System in the Response to Cold Stress.*Frontiers in Physiology* **11** , 376.

Lundgren JG, Lehman RM, Chee-Sanford J (2007) Bacterial communities within digestive tracts of ground beetles (Coleoptera: Carabidae).*Annals of the Entomological Society of America* **100** , 275-282.

Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. *Journal of Experimental Botany* **63** , 599-616.

MacMillan HA, Findsen A, Pedersen TH, Overgaard J (2014) Cold-induced depolarization of insect muscle: differing roles of extracellular K+ during acute and chronic chilling. *Journal of Experimental Biology* **217** , 2930-2938.

Malhotra JD, Kaufman RJ (2007) Endoplasmic reticulum stress and oxidative stress: a vicious cycle or a double-edged sword?*Antioxidants & redox signaling* **9** , 2277-2294.

Mani MS (1968) *Ecology and Biogeography of High Altitude Insects*Dr. W. Junk N.V. Publishers, The Hague.

Marcais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics***27** , 764-770.

McDaniel EA, Anantharaman K, McMahon K (2019) metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. *bioRxiv* .

McKenna A, Hanna M, Banks E*, et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20** , 1297-1303.

McKenna DD (2018) Beetle genomes in the 21st century: prospects, progress and priorities. *Curr Opin Insect Sci* **25** , 76-82.

McKenna DD, Shin S, Ahrens D*, et al.* (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences* **116** , 24729-24737.

McManus R, Ravenscraft A, Moore W (2018) Bacterial associates of a gregarious riparian beetle with explosive defensive chemistry.*Frontiers in microbiology* **9** , 2361.

Medina P, Russell SL, Corbett-Detig R (2019) Deep data mining reveals variable abundance and distribution of microbial reproductive manipulators within and among diverse host species. *bioRxiv* , 679837.

Mehmood A, Laiho A, Venäläinen MS*, et al.* (2019) Systematic evaluation of differential splicing tools for RNA-seq studies.*Brief Bioinform* .

Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature communications***7** , 1-9.

Miller IJ, Rees ER, Ross J*, et al.* (2019) Autometa: automated extraction of microbial genomes from individual shotgun metagenomes.*Nucleic Acids Research* **47** , e57-e57.

Misof B, Liu S, Meusemann K*, et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346** , 763-767.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server.*Nucleic Acids Res* **35** , W182-185.

Nagy L, Grabherr G (2009) *The biology of alpine habitats* Oxford University Press.

Neven LG, Duman JG, Low MG, Sehl LC, Castellino FJ (1989) Purification and characterization of an insect hemolymph lipoprotein ice nucleator: evidence for the importance of phosphatidylinositol and apolipoprotein in the ice nucleator activity. *Journal of Comparative Physiology B* **159** , 71-82.

O'Donovan C, Martin MJ, Gattiker A*, et al.* (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Briefings in bioinformatics* **3** , 275-284.

O'Leary NA, Wright MW, Brister JR*, et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44** , D733-745.

Overgaard J, MacMillan HA (2017) The integrative physiology of insect chill tolerance. *Annual Review of Physiology* **79** , 187-208.

Palmer DH, Rogers TF, Dean R, Wright AE (2019) How to identify sex chromosomes and their turnover. *Molecular Ecology* **28** , 4709-4724.

Parker DJ, Wiberg RAW, Trivedi U*, et al.* (2018) Inter and intraspecific genomic divergence in *Drosophila montana* shows evidence for cold adaptation. *Genome Biology and Evolution***10** , 2086-2101.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research***25** , 1043-1055.

Pascua-Maestro R, Diez-Hermano S, Lillo C, Ganfornina MD, Sanchez D (2017) Protecting cells by protecting their vulnerable lysosomes: Identification of a new mechanism for preserving lysosomal functional integrity upon oxidative stress. *PLoS Genet* **13** , e1006603.

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR (2020) Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). *G3 (Bethesda)***10** , 3047-3060.

Picard (2018) Picard toolkit. Broad Institute, GitHub repository, http://broadinstitute.github.io/picard/.

Pivovarov AS, Calahorro F, Walker RJ (2019) Na+/K+-pump and neurotransmitter membrane receptors. *Invertebrate Neuroscience***19** , 1.

Pizzolotto R, Mazzei A, Bonacci T*, et al.* (2018) Ground beetles in Mediterranean olive agroecosystems: Their significance and functional role as bioindicators (Coleoptera, Carabidae). *PLoS ONE***13** , e0194551.

Poplin R, Ruano-Rubio V, DePristo MA*, et al.* (2017) Scaling accurate genetic variant discovery to tens of thousands of samples.*bioRxiv* , 201178.

Quesneville H (2015) Impact of transposable elements on insect genomes and biology. *Current Opinion in Insect Science* **7** , 1-7.

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.*Nature Communications* **11** , 1432.

Rapaport F, Khanin R, Liang Y*, et al.* (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology* **14** , 1-13.

Rastogi A, Gupta D (2014) GFF-Ex: a genome feature extraction package.*BMC Research Notes* **7** , 315.

Receveur JP, Fenoglio S, Benbow ME (2020) Insect-associated bacterial communities in an alpine stream. *Hydrobiologia* **847** , 331-344.

Ritchie H, Jamieson AJ, Piertney SB (2017) Genome size variation in deep-sea amphipods. *Royal Society Open Science* **4** , 170862.

Ritchie ME, Phipson B, Wu D*, et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43** , e47.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.*Bioinformatics* **26** , 139-140.

Rocha DF, Wouters FC, Zampieri DS*, et al.* (2013) Harvestman phenols and benzoquinones: characterisation and biosynthetic pathway.*Molecules* **18** , 11429-11451.

Schmidt J (2009) Taxonomic and biogeographical review of the genus *Trechus* Clairville, 1806, from the Tibetan Himalaya and the southern central Tibetan Plateau (Coleoptera: Carabidae: Trechini). *Zootaxa* **2178** , 1-72.

Schmidt J, Böhner J, Brandl R, Opgenoorth L (2017) Mass elevation and lee effects markedly lift the elevational distribution of ground beetles in the Himalaya-Tibet orogen. *PLoS ONE* **12** , e0172939.

Schmitz JF, Zimmer F, Bornberg-Bauer E (2016) Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Research***44** , 6287-6297.

Schoville SD, Simon S, Bai M, *et al.* (2020) Comparative transcriptomics of ice-crawlers demonstrates cold specialization constrains niche evolution in a relict lineage. *Evolutionary Applications* .

Schowalter TD (2013) *Insects and sustainability of ecosystem services* CRC Press.

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30** , 2068-2069.

Seppey M, Ioannidis P, Emerson BC, *et al.* (2019a) Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. *Genome biology* **20** , 98.

Seppey M, Manni M, Zdobnov EM (2019b) BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology***1962** , 227-245.

Serrano J, Galián J (1998) A review of karyotypic evolution and phylogeny of carabid beetles (Coleoptera). *Phylogeny and classification of Caraboidea (Coleoptera: Adephaga). Museo Regionale di Scienze Naturali, Torino* , 191-228.

Serrano J, Yadav J (1984) Chromosome numbers and sex-determining mechanisms in Adephagan Coleoptera. *The Coleopterists' Bulletin* , 335-357.

Shaffer JF, Gillis TE (2010) Evolution of the regulatory control of vertebrate striated muscle: the roles of troponin I and myosin binding protein-C. *Physiological Genomics* **42** , 406-419.

Sinclair BJ, Vernon P, Klok CJ, Chown SL (2003) Insects at low temperatures: an ecological perspective. *Trends in Ecology & Evolution* **18** , 257-262.

Slade EM, Bagchi R, Keller N, Philipson CD (2019) When do more species maximize more ecosystem services? *Trends in Plant Science***24** , 790-793.

Slatyer RA, Nash MA, Hoffmann AA (2016) Scale-dependent thermal tolerance variation in Australian mountain grasshoppers. *Ecography* **39** , 572-582.

Slatyer RA, Schoville SD (2016) Physiological limits along an elevational gradient in a radiation of montane ground beetles. *PLoS ONE* **11** , e0151959.

Smit A, Hubley R, Green P (2015) RepeatMasker Open-4.0 [http://www. repeatmasker. org] Accessed September, 2020. Accessed.

Smit AF, Hubley R (2008) RepeatModeler Open-1.0. *Available fom http://www. repeatmasker. org* .

Smith SA, O'Meara BC (2012) treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics***28** , 2689-2690.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics***30** , 1312-1313.

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24** , 637-644.

Stanke M, Schoffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7** , 62.

Storey KB, Storey JM (2010) Oxygen: Stress and adaptation in cold-hardy insects.

Storey KB, Storey JM (2012) Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology* **90** , 456-475.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE***6** , e21800.

Teets NM, Denlinger DL (2013) Physiological mechanisms of seasonal and rapid cold-hardening in insects. *Physiological Entomology***38** , 105-116.

Teets NM, Kawarasaki Y, Lee RE, Denlinger DL (2011) Survival and energetic costs of repeated cold exposure in the Antarctic midge,*Belgica antarctica* : a comparison between frozen and supercooled larvae. *Journal of Experimental Biology* **214** , 806-814.

Teets NM, Kawarasaki Y, Potts LJ*, et al.* (2019) Rapid cold hardening protects against sublethal freezing injury in an Antarctic insect. *Journal of Experimental Biology* **222** .

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18** , 1979-1990.

Timmermans MJ, Vogler AP (2012) Phylogenetically informative rearrangements in mitochondrial genomes of Coleoptera, and monophyly of aquatic elateriform beetles (Dryopoidea). *Molecular Phylogenetics and Evolution* **63** , 299-304.

UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge.*Nucleic Acids Research* **47** , D506-D515.

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research***27** , 737-746.

Venkat A, Hahn MW, Thornton JW (2018) Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature ecology & evolution* **2** , 1280-1288.

Vinogradov AE (1999) Genome in toto. *Genome* **42** , 361-362.

Wagner A (2005) Energy constraints on the evolution of gene expression.*Molecular Biology and Evolution* **22** , 1365-1374.

Wang Y, Tang H, DeBarry JD*, et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.*Nucleic Acids Research* **40** , e49-e49.

Weng YM, Kavanaugh DH, Schoville SD (2020) Drainage basins serve as multiple glacial refugia for alpine habitats in the Sierra Nevada Mountains, California. *Molecular Ecology* .

Weng YM, Francoeur CB, Currie C, Kavanaugh DH, Schoville SD (2020) A high-quality carabid genome provides insights into beetle genome evolution and cold adaptation, Dryad, Dataset, https://doi.org/10.5061/dryad.8cz8w9gp5

Weyenberg G, Huggins PM, Schardl CL, Howe DK, Yoshida R (2014) kdetrees: Non-parametric estimation of phylogenetic tree distributions.*Bioinformatics* **30** , 2280-2287.

Williams CM, Szejner-Sigal A, Morgan TJ*, et al.* (2016) Adaptation to Low Temperature Exposure Increases Metabolic Rates Independently of Growth Rates. *Integrative and Comparative Biology* **56** , 62-72.

Wolfe JM, Daley AC, Legg DA, Edgecombe GD (2016) Fossil calibrations for the arthropod Tree of Life. *Earth-Science Reviews* **160** , 43-110.

Wright NA, Gregory TR, Witt CC (2014) Metabolic 'engines' of flight drive genome size reduction in birds. *Proceedings of the Royal Society B: Biological Sciences* **281** , 20132780.

Xie J, Vilchez I, Mateos M (2010) *Spiroplasma* cacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp*Leptopilina heterotoma* . *PLoS ONE* **5** , e12149.

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11** , R14.

Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16** , 284-287.

Zhang SQ, Che LH, Li Y*, et al.* (2018) Evolutionary history of Coleoptera revealed by extensive sampling of genes and species.*Nature Communications* **9** , 205.

Zhu BH, Xiao J, Xue W*, et al.* (2018) P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads.*BMC Genomics* **19** , 175.

Figure Legend

Figure 1. Time-calibrated maximum likelihood phylogeny of Coleoptera, estimated with *Drosophila melanogaster* (Dmel) as outgroup. The branches are labeled with the number of rapid gene family expansions (yellow boxes) and contractions (blue boxes), or labeled zero if no change was estimated. Topological relationships were estimated in RAxML using 379,463 amino acid residues from 1,013 orthologous genes. All nodes have 100% bootstrap support, and nodes that were used to time-calibrate the tree are labeled with stars.

Figure 2 . CirGO plot showing semantic clustering of GO terms for all expanding gene families in the *Nebria riversi* branch.

Figure 3. A) 16S rRNA phylogeny of *Spiroplasma* isolates, showing the placement of *S* . sp. NR in the Carabid clade. B) Schematic showing metabolic functions of *S* . sp. NR.

Table 1 . Genome statistics for *Nebria riversi* (bold type) and eleven Coleoptera species in RefSeq.

	Nebria riversi **(Nriv)**	*Anoplophora glabripennis* (Agla)	*Agrilus planipen*
Contig Count	2136	26749	23186
Contig N50 (Kb)	197	80	28
Contig L50	207	2230	3132
Scaffold Count	1636	9867	3613
Scaffold N50 (Kb)	302	678	1113
Scaffold L50	140	269	91
Complete BUSCOs	2322	2352	2210
Total genes	17,895	16,200	15,336
Genome size (Mb)	147.35	706.97	353.07
Estimated total length of introns (Mb)	47.49	234.24	209.81
NCBI total length of introns (Mb)	*NA*	316.15	249.4082
Estimated largest intron	160753	526744	539725
NCBI largest intron	*NA*	526741	539722
Intron content (%)	32.23	33.13	59.42
NCBI intron content (%)	*NA*	44.72	70.64
TE content (%)	18.43	41.45	24.55

Table 2 . List of predicted features from the *Nebria riversi* transcriptome, genome assembly and merged super-transcriptome.

	Number	Tools

	Number	Tools
Transcriptome		
Genes	129,804	Trinity
Transcripts	201,715	Trinity
Genome		
Repeat regions	150,583	RepeatModeler
Genes	17,895	Braker & PASA
mRNA	24,102	Braker & PASA
Proteins	23,973	Braker & PASA
Super-transcriptome		
Genes	17,736	Necklace

Table 3 . Rapidly expanding hierarchical orthogroups (HOGs), with corresponding genes and gene ontology (GO) terms, for the*Nebria riversi* branch.

HOG	HOG membership size change	Gene products	GO ID	GO Fu
HOG08696	+8	leucine-rich repeat-containing protein	-	-
HOG08786	+6	facilitated trehalose transporter	GO:0008643	carbohyc
			GO:0016020	membra
			GO:0016021	integral
			GO:0022857	transme
			GO:0055085	transme
HOG09114	+16	zinc finger protein	GO:0003676	nucleic a
			GO:0003677	DNA bi
HOG09541	+16	zinc finger protein	GO:0003676	nucleic a
			GO:0003677	DNA bi
HOG10308	+7	cyclin-dependent kinase-like	GO:0000166	nucleotic
			GO:0004672	protein
			GO:0004674	protein s
			GO:0005524	ATP bi
			GO:0006468	protein
HOG11304	+8	troponin I	GO:0005865	striated
			GO:0032991	protein-c
HOG11869	+12	synaptic vesicle glycoprotein 2B-like or 2C-like	GO:0016020	membra
			GO:0016021	integral
			GO:0022857	transme
			GO:0055085	transme
HOG14219	+8	multidrug resistance-associated protein	GO:0005524	ATP bi
			GO:0005773	vacuole
			GO:0005774	vacuolar
			GO:0005886	plasma
			GO:0008559	ATPase-
			GO:0016021	integral
			GO:0016887	ATPase
			GO:0042908	xenobiot
HOG16874	+12	zinc finger RNA-binding protein	GO:0003725	double-s
			GO:0003727	single-st

HOG	HOG membership size change	Gene products	GO ID	GO Fu
			GO:0005634	nucleus
			GO:0008270	zinc ion

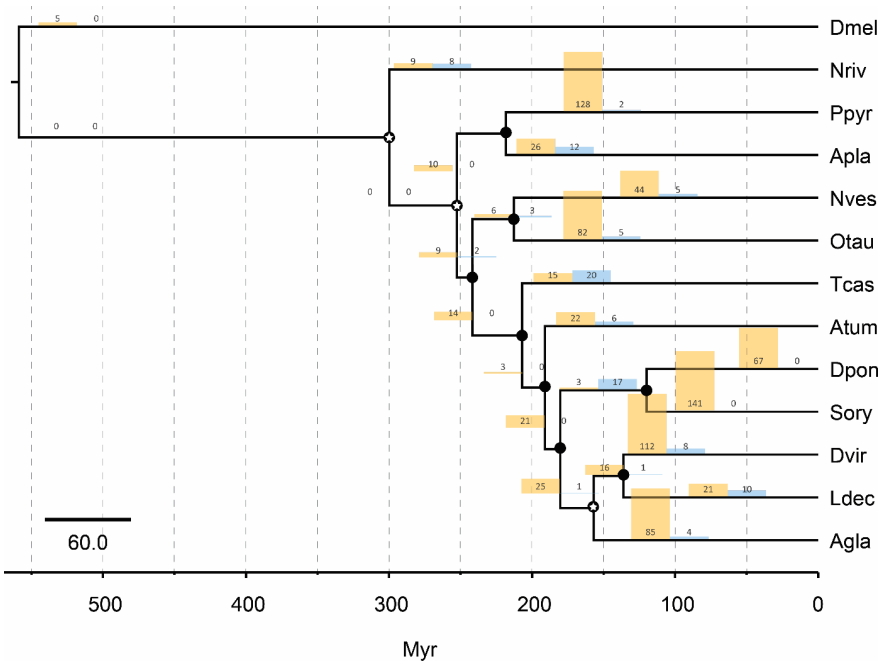

Figure 1. Time-calibrated maximum likelihood phylogeny of Coleoptera, estimated with *Drosophila melanogaster* (Dmel) as outgroup. The branches are labeled with the number of rapid gene family expansions (yellow boxes) and contractions (blue boxes), or labeled zero if no change was estimated. Topological relationships were estimated in RAxML using 379,463 amino acid residues from 1,013 orthologous genes. All nodes have 100% bootstrap support, and nodes that were used to time-calibrate the tree are labeled with stars.

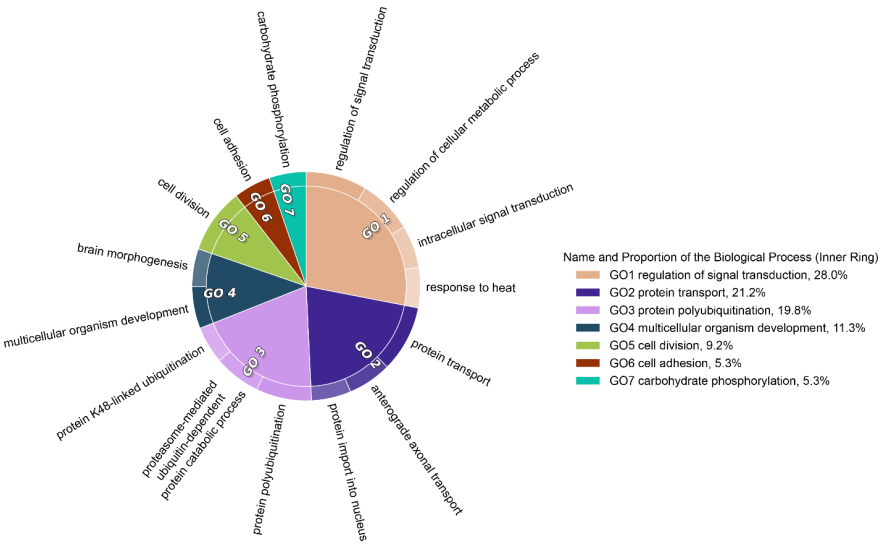

Figure 2. CirGO plot showing semantic clustering of GO terms for all expanding gene families in the *Nebria riversi* branch.

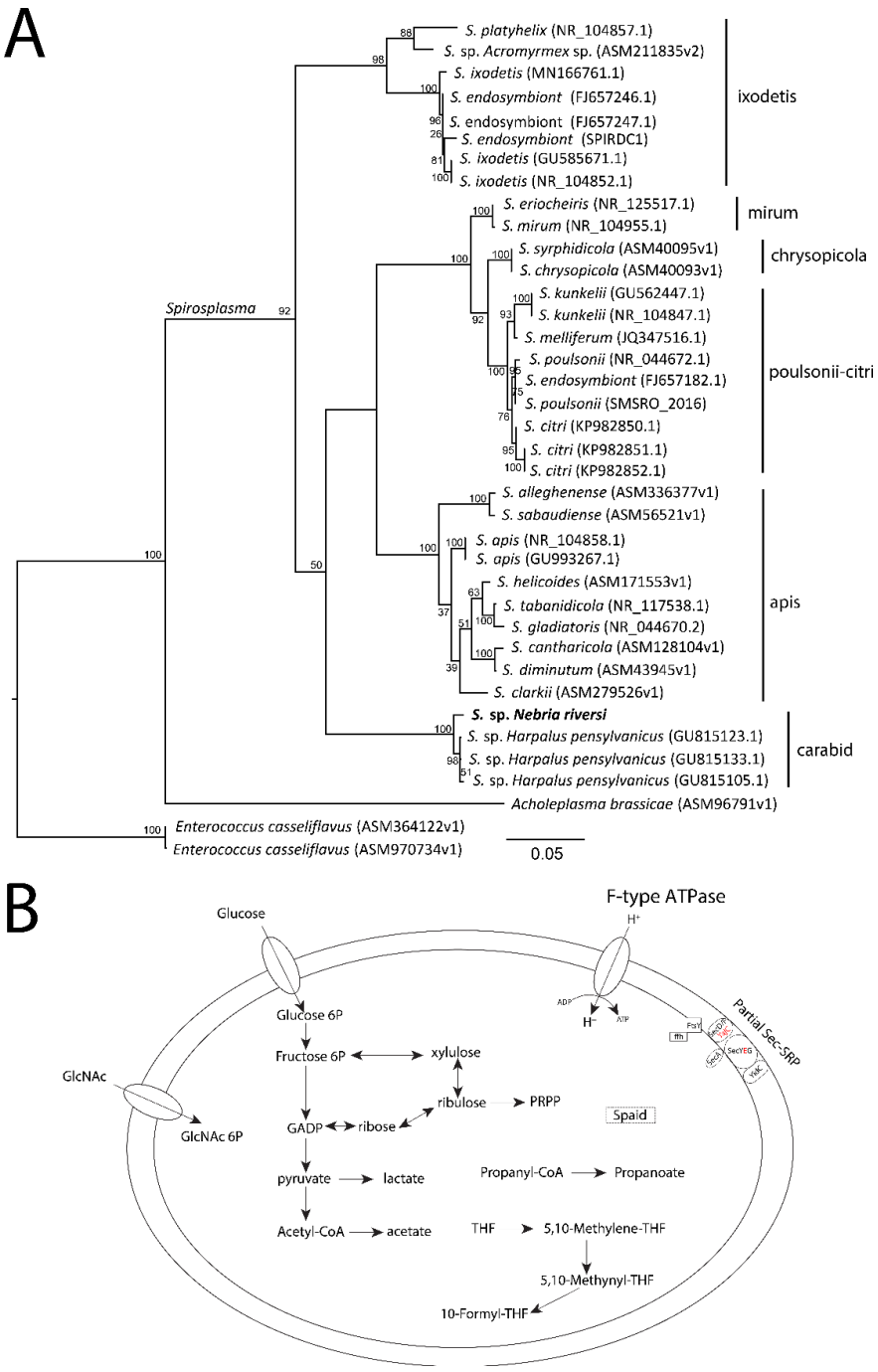

Figure 3. A) 16S rRNA phylogeny of *Spiroplasma* isolates, showing the placement of S . sp. NR in the Carabid clade. B) Schematic showing metabolic functions of S . sp. NR.