

# Chromosome-Scale Genomes of the Flightless Caterpillar Hunter Beetles *Calosoma tepidum* and *Calosoma wilkesii* From British Columbia (Coleoptera: Carabidae)

Jérémy Gauthier <sup>1,2,\*</sup>, Mickael Blanc <sup>1</sup>, Emmanuel F.A. Toussaint <sup>1</sup>

<sup>1</sup>Natural History Museum of Geneva, Geneva, Switzerland

<sup>2</sup>Naturéum—State Museum of Natural Sciences, Lausanne, Switzerland

\*Corresponding author: E-mail: jeremy.gauthier@vd.ch.

Accepted: November 08, 2024

## Abstract

The giant ground beetle genus *Calosoma* (Coleoptera, Carabidae) comprises ca. 120 species distributed worldwide. About half of the species in this genus are flightless due to a process of wing reduction likely resulting from the colonization of remote habitats such as oceanic islands, highlands, and deserts. This clade is emerging as a new model to study the genomic basis of wing evolution in insects. In this framework, we present the *de novo* assemblies and annotations of two *Calosoma* species genomes from British Columbia, *Calosoma tepidum* and *Calosoma wilkesii*. Combining PacBio HiFi and Hi-C sequencing, we produce high-quality reference genomes for these two species. Our annotation using long-read RNAseq and existing Coleoptera protein evidence identified a total of 21,976 genes for *C. tepidum* and 26,814 genes for *C. wilkesii*. Using synteny analyses, we provide an in-depth comparison of genomic architectures in these two species. We infer an overall pattern of chromosome-scale conservation between the two species, with only minor rearrangements within chromosomes. These new reference genomes represent a major step forward in the study of this group, providing high-quality references that open the door to different approaches such as comparative genomics or population scale resequencing to study the implications of flight evolution.

**Key words:** Adephaga genomes, Carabinae, PacBio HiFi sequencing, South Okanagan Grasslands Protected Area, Hi-C.

## Significance

The caterpillar hunter beetle genus *Calosoma* is emerging as a new model to study the evolution of wing morphology. Several lineages in this genus are capable of winged flight, a rare condition in the subfamily Carabinae, when the rest is either brachypterous, *i.e.* with reduced wings, or apterous, *i.e.* without wings. Unfortunately, few genomic resources exist to date that would allow a better understanding of flight evolution in *Calosoma*. In this study, we present the chromosomal-scale genomes of two *Calosoma* species with different wing morphologies that will hopefully be useful to future investigations into the genomic basis of flight evolution in this group.

## Introduction

Beetles (Insecta, Coleoptera) are the most diverse lineage of animals on Earth, with >380,000 species described to date and many more to discover and name (Stork 2018). Surprisingly, this order also presents one of the lowest

number of available high-quality animal genomes available to date (McKenna 2018). It is therefore key to develop such genomic resources across the Coleoptera evolutionary tree as these are a window into a better understanding of this astonishing insect radiation.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

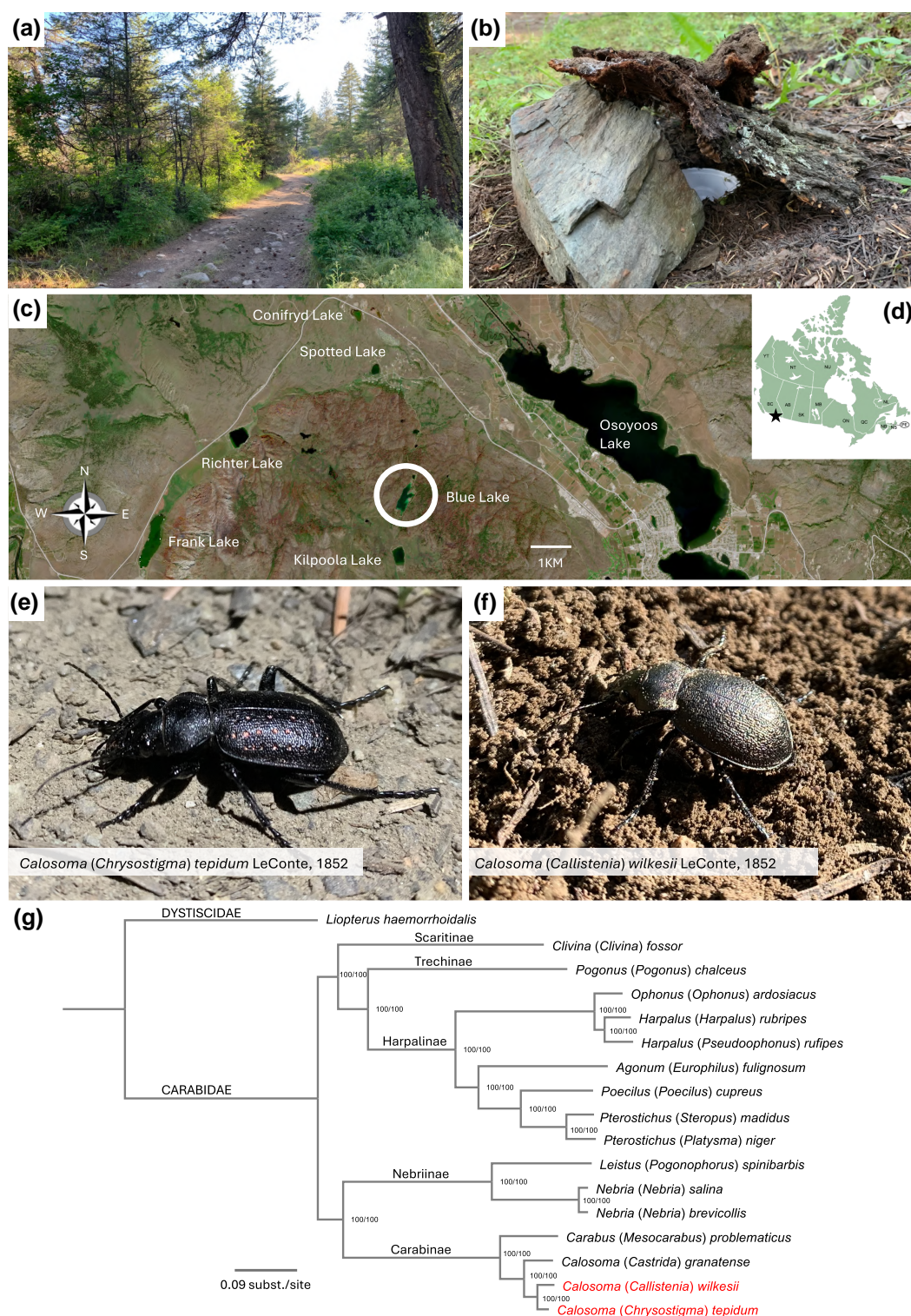
The order Coleoptera comprises four suborders, the Adephaga, Archostemata, Myxophaga, and Polyphaga. As of September 2024, most chromosome-scale beetle genomes belong to the suborder Polyphaga (149 out of 165 available, *i.e.* >90%), while none have been generated for the species-poor suborders Archostemata and Myxophaga (Challis et al. 2023; Genomes on a Tree database last accessed 16.09.2024). Few chromosome-scale genomes of Adephaga have been generated to date with respect to the remarkable diversity of species and ecologies in this clade (Beutel et al. 2020; Baca et al. 2021). Ground beetles in the family Carabidae represent the most species-rich family of the Adephaga suborder, with *ca.* 40,000 described species (Carabcat database 2021). Despite a wide interest in their ecology and evolution, phylogenetic relationships in Carabidae are poorly understood (Maddison et al. 1999, 2009) and the lack of high-quality genomes precludes a better understanding of their evolution in general. In this study, we focus on the subfamily Carabinae that comprises several genera of large predatory beetles most often brachypterous and therefore flightless. This subfamily harbors very few genomes, the two available as of July 2024 being that of *Calosoma* (*Castrida*) *granatense* Géhin, 1885 (Vangestel et al. 2024) and *Carabus* (*Mesocarabus*) *problematicus* Herbst, 1786 (GCA\_963422195.1). Caterpillar hunter beetles of the genus *Calosoma* Weber, 1801 have progressively emerged as an evolutionary model because they showcase several morphological transitions from fully developed functional wings to brachyptery and aptery (Su et al. 2005; Toussaint and Gillett 2018; Sota et al. 2020; Toussaint et al. 2021; Sota et al. 2022). These unique morphological transitions are often associated with the colonization of relatively remote or extreme environments such as oceanic islands, highlands, and deserts. Based on the evolutionary trees reconstructed so far for the genus (Toussaint and Gillett 2018; Sota et al. 2020; Toussaint et al. 2021; Sota et al. 2022), it appears plausible that complex functional wings were reevolved from a brachypterous ancestor in different lineages as observed for instance in stick insects (Whiting et al. 2003; Bank and Bradler 2022). Reconstructing the evolutionary sequence of flight loss and gain in *Calosoma* is key to understanding the genomic basis of wing evolution globally. So far, only the wing polymorphic species *Calosoma granatense* from the Galápagos archipelago has been sequenced (Vangestel et al. 2024). In this study, we report the sequencing of two species presenting different stages of flightlessness, *Calosoma* (*Chrysostigma*) *tepidum* LeConte, 1852 that is macropterous but likely unfit to fly due to thoracic muscle reduction and *Calosoma* (*Callistenia*) *wilkesii* LeConte, 1852 that is brachypterous (Bruschi 2013; pers. obs.). These new genomic resources will likely prove key to investigating wing genomics in the future (Fig. 1e,f).

## Results and Discussion

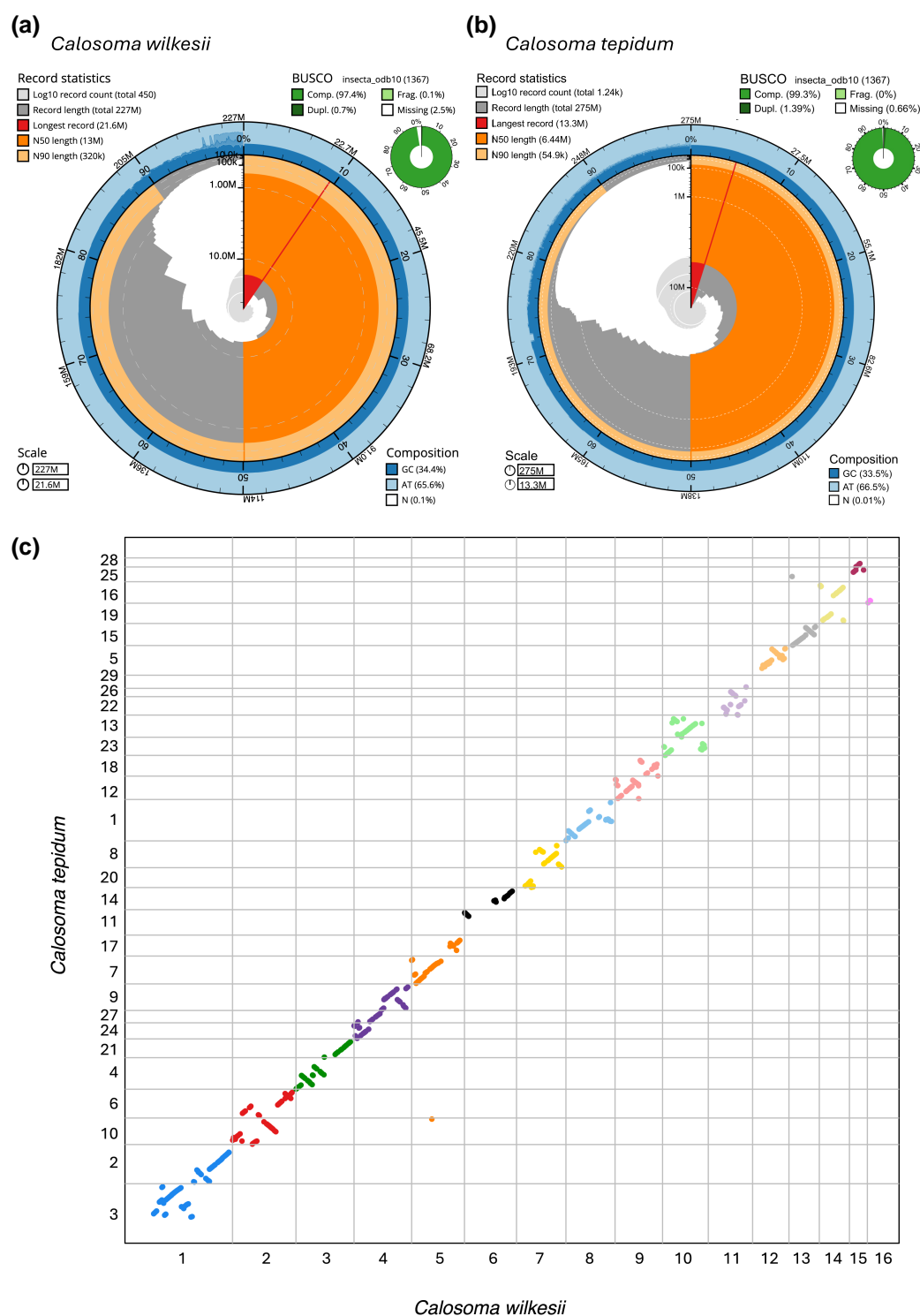
### Sequencing and Assembly Statistics

Using PacBio HiFi sequencing, we obtained 447,595 reads (3.6 Gb) with an average size of 9.0 kb and a N50 of 8.1 kb for *Calosoma wilkesii*; and 572,182 reads (6.7 Gb) with an average size of 11.7 kb and a N50 of 13.1 kb for *Calosoma tepidum*. Genome-size estimations predicted using k-mer count were ~242 Mb for *Calosoma wilkesii*, with a coverage of 7× and ~207 Mb for *Calosoma tepidum*, with a coverage of 13× (supplementary figs. S1 and S3, table S1, Supplementary Material online). The primary assemblies were composed of 1,077 contigs with a N50 of 519 kb for *Calosoma wilkesii* and 1,302 contigs with a N50 of 2.0 Mb for *Calosoma tepidum*. Using Hi-C data, contigs were organized to shape final assemblies. The genome of *Calosoma wilkesii* is composed of 450 scaffolds with a total length of 227 Mb, quality statistics of N50 = 13.0 Mb, and 97.4% completeness based on the universal, single-copy ortholog gene set for Insecta (Fig. 2a). Note that 16 scaffolds are longer than 3 Mb, and create a gap in the size distribution with smaller scaffolds. These 16 scaffolds represent 84.8% of the total genome length and certainly correspond to the chromosomes (Fig. 2c). Chromosome formula seems consistent between *Calosoma* species (five species studied by Serrano and Yadav 1984), and is composed of 13 autosomal chromosomes and two sexual chromosomes (X and Y) with variations in Y chromosome size. Among these 16 scaffolds, only one chromosome appears to have been split in two scaffolds. For reasons of convenience and in view of the quality of the *Calosoma wilkesii* assembly, the scaffolds from this species will be referred to as pseudo-chromosomes. The genome of *Calosoma tepidum* is composed of 1,243 scaffolds with a total length of 273 Mb, with quality statistics of N50 = 6.5 Mb, and 99.3% BUSCO completeness (Fig. 2b). The contingency of this assembly is marginally inferior to that of *Calosoma wilkesii*. Twenty-nine scaffolds are larger than 1 Mb, and they represent 73.2% of the genome (Fig. 2c). This implies that some chromosomes remain fragmented despite the use of Hi-C data to reconstruct chromosome architecture. The quality of the assemblies we obtained is comparable to the genomes currently available for the Carabidae group, such as the genome of *Calosoma granatense*, composed of 6,044 scaffolds with a total size of 168 Mb, a N50 of 5.6 Mb, and a BUSCO completeness of 98.3%. The use of this latter genome as a reference for population genomic studies across the Galápagos archipelago reveals the usefulness of such reference genomes (Vangestel et al. 2024). The other available genome is that of *Carabus problematicus*, which belongs to the sister genus to *Calosoma*. This genome is composed of 222 scaffolds





**Fig. 1.** a) Habitat in which both species were found, forested paths close to the Blue Lake in South Okanagan Grasslands Protected Area (credit: Emmanuel Toussaint), b) pitfall trap baited with red wine vinegar used to collect the two species (credit: Emmanuel Toussaint). c) Satellite map of the research area (open source Copernicus Sentinel-2 data). d) Map of Canada highlighting the location of the research area in British Columbia (open source Wikimedia Commons), e) photograph of *Calosoma (Chrysostigma) tepidum* LeConte, 1852, at the Blue Lake locality (credit: Emmanuel Toussaint), f) photograph of *Calosoma (Callistenia) wilkesii* LeConte, 1852, at the Blue Lake locality (credit: Emmanuel Toussaint), g) maximum likelihood phylogenomic tree of Carabidae inferred using a matrix of BUSCO genes extracted from available genomic-scale resources. Branch support metrics across the topology correspond to SH-aLRT and ultrafast bootstrap (UFBoot) as calculated in IQ-TREE. Newly generated genomes are labeled in red.



**Fig. 2.** Snail plot including genome statistics for the final assemblies *C. wilkesii* a) and *C. tepidum* b). c) Synteny plot. The positions of BUSCO genes mapping uniquely to both genomes are shown in the order of the *C. wilkesii* pseudo-chromosomes. The colors reflect the different *C. wilkesii* pseudo-chromosomes. A fully conserved chromosome would be reflected as a single diagonal line. Gray lines indicate scaffold ends.

with a total size of 254 Mb, a N50 of 17.5 Mb, and a BUSCO completeness of 99.2%.

Using a selection of universal loci from the genomic data (BUSCO genes), we inferred a phylogenomic tree including



all available high-quality long-read genomes of Carabidae as well as of a single Dytiscidae. This newly generated genomic data refines our understanding of Carabidae phylogenetic relationships. Despite a reduced taxon sampling, the results of our phylogenomic inference highlight some interesting placements that have remained controversial until recently (Maddison et al. 1999, 2009; Vasilikopoulos et al. 2021). Within Carabidae, we recover two main clades, the first one comprises the subfamilies Harpalinae, Scaritinae, and Trechinae, while the second one comprises the subfamilies Carabinae and Nebriinae. This latter relationship represents an interesting result with robust branch support that was only recently evidenced by target capture phylogenomics (Vasilikopoulos et al. 2021). Previous studies based on Sanger sequencing of few loci had alternatively recovered Carabinae as an early-diverging branch in Carabidae with no closely related lineages (Maddison et al. 1999, 2009) or with possible ties with the subfamily Elaphrinae (Ribera et al. 2005). Additional sampling at a genomic scale is needed to resolve with more accuracy the placement of Carabinae within Carabidae. Within Carabinae, the genus *Calosoma* is inferred as sister to *Carabus problematicus*, thereby forming a monophyletic tribe Carabini, a result in line with both earlier molecular studies (Osawa et al. 2004) and recent phylogenomic ones (Toussaint et al. 2021; Vasilikopoulos et al. 2021; Sota et al. 2022). Finally, the two *Calosoma* species sequenced in this study are recovered as sister taxa and are nested in a monophyletic genus *Calosoma* along with the species *C. granatense*. The phylogeny of *Calosoma* has recently received increased attention due to the putative existence of functional wing re-evolution in this group (Toussaint and Gillett 2018; Toussaint et al. 2021). Although several subgenera within *Calosoma* have been inferred to be either para- or polyphyletic, the two species *Calosoma (Chrysostigma) tepidum* and *C. (Callistenia) wilkesii* were recovered in Clade II in Toussaint and Gillett (2018), while *C. (Castrida) granatense* was recovered in the phylogenetically distant Clade CVI. The results of this study are therefore in line with the ones of previous studies on this genus as well as with the seminal work of René Jeannel who already noted the morphological similarity of the subgenera *Calosoma (Chrysostigma)* Kirby, 1837, *C. (Callistenia)* Lapouge, 1929, and *C. (Callisthenes)* Fischer von Waldheim, 1822 (Jeannel 1940), these three lineages forming a monophyletic group in recent molecular studies (Toussaint and Gillett 2018; Toussaint et al. 2021). Future more densely sampled phylogenetic trees of *Calosoma* caterpillar hunter beetles will be helpful to better understand their evolutionary history especially with respect to wing morphology.

Analysis of the synteny between the two species is informative with respect to the genomic architecture (Fig. 2). Using the pseudo-chromosomes of *Calosoma*

*wilkesii* as a reference, we observed that many *C. tepidum* scaffolds can be associated with pseudo-chromosomes. Most of the time, a *Calosoma wilkesii* pseudo-chromosome corresponds to two *C. tepidum* scaffolds, for instance *C. wilkesii* pseudo-chromosome 1 corresponds to *C. tepidum* scaffold\_2 and scaffold\_3. In some cases, the assembly is a direct match, as in the case of pseudo-chromosome 8 and scaffold\_1. Numerous breaks were identified but only in the pseudo-chromosomes identified and almost never between pseudo-chromosomes, revealing recombination events within chromosomes but not between chromosomes. This is consistent with the divergence time between the two species that belong to two different subgenera within the comparatively recent genus *Calosoma* estimated to be ca. 25 to 30 million years old (Sota et al. 2022). It should be noted that scaffold\_16 from *Calosoma wilkesii*, which appears to be the supernumerary scaffold in terms of the karyotype, appears to be syntenic with scaffold\_16 from *C. tepidum*, which itself is largely syntenic with pseudo-chromosome 14 from *C. tepidum*. Additional data, not based on synteny with the risks of recombination, could allow this scaffold to be reintegrated into the overall assembly and thus achieve a chromosome-scale assembly consistent with the karyotype.

## Genome Annotation

The structural annotation performed using BRAKER3 and combining hints from Coleoptera proteins and long-read RNAseq resulted in the annotation of 26,814 genes with 88.5% BUSCO completeness and 21,976 genes with 91.0% BUSCO completeness for *Calosoma wilkesii* and *C. tepidum*, respectively. These statistics are lower than those obtained on genomes, but consistent with the literature although comparison is difficult, as no annotation exists for the Adephaga genomes available. However, among the 50 existing annotations of coleopteran genomes, the average number of genes identified is 20,005 genes (sd = 7,872) (data extracted from the Arthropoda Assembly Assessment Catalogue (Feron and Waterhouse 2022)). This makes the annotations proposed in this study the first annotations available for Carabidae species. The identification of orthologous families between the two species has resulted in the assignment of 40,976 genes (84.1% of total) to 12,884 orthogroups.

## Materials and Methods

### Sample Collection

Samples of both *Calosoma tepidum* and *C. wilkesii* were collected in forested paths near the Blue Lake (841 m, 49°02'23.4"N 119°33'26.1"W) in South Okanagan Grasslands protected area, British Columbia, Canada between the 5<sup>th</sup> and 7<sup>th</sup> of June 2023 under a BC Park Use

Permit No. 111631 to EFAT and MB (Fig. 1c). Pitfall traps baited with red wine vinegar and active hunting at night with headlamps were used (Fig. 1a and b). The pitfall traps were nonlethal, using a central emerged piece of wood or rock to ensure that specimens would not drown overnight. The traps were checked every day to ensure that non-targeted species would not be collaterally killed. All live specimens collected were kept in containers with a humid cotton ball and in a cooler before molecular lab procedures.

### DNA Extraction Library Construction and Sequencing

High-molecular weight (HMW) DNA was extracted from the thorax, which contains the most muscle and therefore potentially the most DNA, avoiding potential contamination from the gut. Body parts of the beetles were reduced to powder in liquid nitrogen using Bessman Tissue Pulverizer and kept at  $-80^{\circ}\text{C}$ . DNA was extracted using the MagAttract HMW DNA Kit from Qiagen including a proteinase K digestion and was eluted in 100  $\mu\text{L}$  Elution Buffer (Qiagen). DNA was quantified by fluorescence using the Quant-iT DNA Assay Kit, High Sensitivity from Thermo Fisher Scientific. Purity was checked on DS-11 spectrophotometer from DeNovix. Profile and size were assessed with the HS DNA kit on Fragment Analyzer (Agilent Technologies). A quantity of 500 to 1,000 ng gDNA was fragmented with the Megaruptor 3 from Diagenode to obtain 12 to 15 kb average size DNA fragments. After purification of fragmented gDNA, each sample was used to prepare one SMARTbell library with the SMARTbell prep kit 3.0 according to the whole genome protocol from PacBio. The polymerase SMRT Library complex was prepared using the Sequel II Binding Kit 3.2, and sequencing was performed on the Sequel II System on SMRTcell at a loading concentration of 80 pM and with a movie run-time of 30 h. The BAM files of PacBio (HiFi reads) underwent processing using SMRT Link (v11.1.0.166339), following the Demultiplex Barcodes utility, leading to barcode-sorted BAM files.

The Proximo Hi-C library was constructed following the instructions in the Proximo Hi-C kit manual (Phase Genomics, Seattle, WA, USA). Sequencing of the Hi-C library was performed on an Illumina NovaSeq 6000 platform with a  $2 \times 150$  bp paired-end reads.

RNA was extracted from the head which contains the most specific transcripts to cover as much as possible the full transcript register. RNA was extracted using RNeasy Fibrous Tissue Mini Kit from Qiagen including a DNase treatment, eluted in 50  $\mu\text{L}$  RNase-free water and then submitted to initial quality control. RNA was quantified by fluorescence using the Quant-iT RNA Assay Kit, High Sensitivity from Thermo Fisher Scientific. Purity was checked on DS-11 spectrophotometer from DeNovix. Quality was assessed with the HS RNA kit on Fragment Analyzer (Agilent Technologies). Before processing, RNA samples were purified using the

RNA Clean and Concentrator-5 kit (R1013) from Zymo Research. First-strand cDNA synthesis was performed using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification kit (New England Biolabs, E6421) from 300 ng of total RNA input according to PacBio's instructions. Template Switching reaction was achieved using the Iso-Seq express template switching oligo (TSO) from PacBio. cDNA was then purified before amplification. A total of 15 PCR cycles of amplification was performed with indexed primers designed by PacBio for each RNA sample using the NEBNext Single Cell cDNA PCR Master Mix. After purification, amplified indexed cDNA products were pooled together to prepare one SMRTbell library with the SMRTbell prep kit 3.0 according to the Iso-Seq protocol from PacBio. The polymerase SMRT Library complex was prepared using the Sequel II Binding Kit 3.1, and sequencing was performed on the Sequel II System on one SMRTcell at a loading concentration of 80 pM and with a movie run-time of 24 h.

The BAM files of PacBio (HiFi reads) underwent processing using SMRT Link (v11.1.0.166339), following the Demultiplex Barcodes utility, to obtain barcode-sorted files where the generic SMRTbell adapter sequences are trimmed. These were then processed using the IsoSeq SMRT tools analysis pipeline. The Iso-Seq Analysis involved primer removal and demultiplexing using "lima" with parameters "--isoseq" (to get rid of Iso-Seq primer sequences). To generate full-length non-chimeric (FLNC) reads, we removed chimeric reads and trimmed poly-A tails using the "isoseq refine" with parameter "--require-polya".

### Genome Assembly and Evaluation

Initial assembly was performed using Hifiasm (Cheng et al. 2021), and duplications were removed using purge\_dups (Guan et al. 2020). Hi-C data was mapped on the initial assembly using bwa-mem2 (Vasimuddin et al. 2019) followed by cleaning steps to sort and remove PCR duplicates using samtools 1.19 (Li et al. 2009) and following the HiC\_pipeline (<https://github.com/esrice/hic-pipeline>). The assembly was then scaffolded integrating Hi-C data using YaHS (Zhou et al. 2023) with default parameters. Manual curation and visualization were performed using PretextView (<https://github.com/sanger-tol/PretextView>) (supplementary fig. S2, Supplementary Material online). The mitochondrial genomes were assembled using MitoHiFi (Uliano-Silva et al. 2023), integrating MitoFinder (Allio et al. 2020). PacBio reads were mapped on the final assembly using minimap2 2.22.6 (Li 2021) with map-hifi option, and coverage was calculated using samtools depth (Li et al. 2009). Putative contaminations were investigated by cutting the genome in 1 Mb fragments and blasting against the NR database using diamond 2.19 (Buchfink et al. 2021). Completeness was evaluated using BUSCO (Seppey et al. 2019), and the insecta\_odb10 composed of 1,367 genes. To evaluate contamination, scaffolds

were cut into 1 Mb fragments and blasted against the UniProt database (UniProt Consortium 2023) using Diamond blastx (Buchfink et al. 2021). Putative contamination scaffolds were excluded by combining the results of BLAST, GC content, and coverage. All this data has been integrated into the BlobToolKit environment (Challis et al. 2020) for visualization.

### Transcriptomic Data and Gene Structure Prediction

Protein coding genes were predicted using BRAKER3 (Gabriel et al. 2024) following dedicated long read protocol and integrating evidence from RNAseq data and protein data. First, long-read RNAseq, i.e. PacBio CSS, were clustered using IsoSeq workflow (PacBio) and mapped on the genome using minimap2 and splice:hq option (Li 2021). Gene predictions were generated using GeneMarkS-T (Tang et al. 2015). Second, reference proteins were extracted from OrthoDBv11 (Kuznetsov et al. 2023) to keep Insecta data and used in BRAKER2 pipeline to design hints combining diamond (Buchfink et al. 2021), spaln2 (Iwata and Gotoh 2012), GeneMark\_ES (Lomsadze et al. 2005), ProHint (Brůna et al. 2020), and Augustus (Hoff and Stanke 2019). Finally, the two sets of predictions were combined using TSEBRA (Gabriel et al. 2021). Proteins were extracted from the annotation, and their completeness was evaluated using BUSCO under protein mode (Seppey et al. 2019). The proteins predicted by these annotations for the two species were compared using OrthoFinder v2.5.5 (Emms and Kelly 2015) in order to identify the genes shared and specific to each species.

### Phylogenetic Inference and Synteny

BUSCO single-copy orthologs were extracted from 15 closely related species and outgroups, aligned using MAFFT (Katoh and Standley 2013) and concatenated using AMAS (Borowiec 2016). The species phylogeny was performed on this alignment composed of 1,367 orthologs using IQ-TREE 2.0.5 (Minh et al. 2020) using the edge-linked partition model (Chernomor et al. 2016). Best partitioning schemes were estimated using PartitionFinder 2.1.1 (Lanfear et al. 2017), and best-fit models of nucleotide substitution were identified using ModelFinder (Kalyaanamoorthy et al. 2017). To estimate branch support, we calculated 1,000 ultrafast bootstraps along with 1,000 SH-aLRT tests in IQ-TREE (Guindon et al. 2010; Hoang et al. 2018). Synteny between the two genomes was investigated based on the positions of the BUSCO single-copy orthologs and using a custom-made R script (Gauthier et al. 2023).

### Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

### Acknowledgments

We warmly thank the Syilx People of the Okanagan Nation for allowing us to conduct research in protected areas of high cultural importance. We thank BC Parks (British Columbia Ministry of Environment and Climate Change Strategy) for granting us permits to collect specimens in the South Okanagan Grasslands and White Lake Grasslands Protected Areas in concert with the First Nation representatives. We especially want to thank Wendy Pope and Mark Weston for fruitful discussions and help in permit applications. We kindly thank Harlan M. Gough and Stan Gough for assistance in fieldwork and logistics. We would also like to thank the Fasteris team (Geneva, Switzerland) for the sequencing and their support. We would also like to thank the editors, John Wang and Laura A. Katz, and the reviewers for their constructive feedback.

### Funding

This project was funded by a grant from the Swiss National Science Foundation (SNSF) 310030\_200491 to E.F.A.T.

### Data Availability

Genome assemblies and annotations have been made available on the NCBI under JBIFNN000000000 accession number for *C. wilkesii* and JBIFNO000000000 for *C. tepidum*. Raw genomic, long-read PacBio and Hi-C, and transcriptomic, long-read RNAseq, data can be found under NCBI BioProjects PRJNA1160031. The assembly and annotation pipelines including custom scripts have been made available in the GitHub repository [https://github.com/JeremyL-Gauthier/Calosoma\\_genomes](https://github.com/JeremyL-Gauthier/Calosoma_genomes). The genome accessions are included below and are present in the “genome info” file of this submission at <https://submit.ncbi.nlm.nih.gov/subs/genome/>. Genome accession: SAMN43589864, JBIFNN000000000; SAMN43589865, JBIFNO000000000.

### Literature Cited

- Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz R, Delsuc F. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour.* 2020;20(4):892–905. <https://doi.org/10.1111/1755-0998.13160>.
- Baca SM, Gustafson GT, Alexander AM, Gough HM, Toussaint EFA. Integrative phylogenomics reveals a Permian origin of Adephaga beetles. *Syst Entomol.* 2021;46(4):968–990. <https://doi.org/10.1111/syen.12506>.
- Bank S, Bradler S. A second view on the evolution of flight in stick and leaf insects (Phasmatodea). *BMC Ecol Evol.* 2022;22(1):62. <https://doi.org/10.1186/s12862-022-02018-5>.
- Beutel RG, Ribera I, Fikacek M, Vasilikopoulos A, Misof B, Balke M. The morphological evolution of the Adephaga (Coleoptera). *Syst Entomol.* 2020;45(2):378–395. <https://doi.org/10.1111/syen.12403>.



- Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 2016;4:e1660. <https://doi.org/10.7717/peerj.1660>.
- Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform*. 2020;2(2):lqaa026. <https://doi.org/10.1093/nargab/lqaa026>.
- Bruschi S. Calosoma of the world. Bologna.: Natura Edizioni Scientifiche; 2013.
- Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18(4):366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Carabcat database. 2021. [Accessed July 29, 2024] <https://www.catalogueoflife.org/data/dataset/1146>.
- Challis R, Kumar S, Sotero-Caio C, Brown M, Blaxter M. Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Res*. 2023;8:24. <https://doi.org/10.12688/wellcomeopenres.18658.1>.
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda)*. 2020;10(4):1361–1374. <https://doi.org/10.1534/g3.119.400908>.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18(2):170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 2016;65(6):997–1008. <https://doi.org/10.1093/sysbio/syw037>.
- Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16(1):157. <https://doi.org/10.1186/s13059-015-0721-2>.
- Feron R, Waterhouse RM. Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *Gigascience*. 2022;11:giac006. <https://doi.org/10.1093/gigascience/giac006>.
- Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res*. 2024;34(5):769–777. <https://doi.org/10.1101/gr.278090.123>.
- Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. 2021;22(1):566. <https://doi.org/10.1186/s12859-021-04482-0>.
- Gauthier J, Meier J, Legeai F, McClure M, Whibley A, Bretaudeau A, Boulain H, Parrinello H, Mugford ST, Durbin R, et al. First chromosome scale genomes of ithomiine butterflies (Nymphalidae: Ithomiini): comparative models for mimicry genetic studies. *Mol Ecol Resour*. 2023;23(4):872–885. <https://doi.org/10.1111/1755-0998.13749>.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36(9):2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35(2):518–522. <https://doi.org/10.1093/molbev/msx281>.
- Hoff KJ, Stanke M. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics*. 2019;65(1):e57. <https://doi.org/10.1002/cpbi.57>.
- Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res*. 2012;40(20):e161. <https://doi.org/10.1093/nar/gks708>.
- Jeannel R. Les Calosomes (Coleoptera, Carabidae). Paris (France): Éditions du Muséum; 1940.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–589. <https://doi.org/10.1038/nmeth.4285>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva E, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res*. 2023;51(D1):D445–D451. <https://doi.org/10.1093/nar/gkac998>.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 2017;34:772–773. <https://doi.org/10.1093/molbev/msw260>.
- Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572–4574. <https://doi.org/10.1093/bioinformatics/btab705>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 2005;33(20):6494–6506. <https://doi.org/10.1093/nar/gki937>.
- Maddison DR, And MDB, Ober KA. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). *Syst Entomol*. 1999;24(2):103–138. <https://doi.org/10.1046/j.1365-3113.1999.00088.x>.
- Maddison DR, Moore W, Baker MD, Ellis TM, Over KA, Cannone JJ, Gutell RR. Monophyly of terrestrial adephagan beetles as indicated by three nuclear genes (Coleoptera: Carabidae and Trachypachidae). *Zool. Scr*. 2009;38(1):43–62. <https://doi.org/10.1111/j.1463-6409.2008.00359.x>.
- McKenna DD. Beetle genomes in the 21st century: prospects, progress and priorities. *Curr Opin Insect Sci*. 2018;25:76–82. <https://doi.org/10.1016/j.cois.2017.12.002>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Osawa S, Su ZH, Imura Y. Phylogeny and distribution of the subfamily Carabinae. In: Osawa S, Su ZH, Imura Y, editors. *Molecular phylogeny and evolution of carabid ground beetles*. Tokyo: Springer; 2004. p. 25–32.
- Ribera I, Mateu J, Bellés X. Phylogenetic relationships of *Dalyat mirabilis* Mateu, 2002, with a revised molecular phylogeny of ground beetles (Coleoptera, Carabidae). *J Zool Syst Evol Res*. 2005;43(4):284–296. <https://doi.org/10.1111/j.1439-0469.2005.00324.x>.
- Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editors. *Gene*

- prediction: methods and protocols. New York (NY): Springer New York; 2019. p. 227–245.
- Serrano J, Yadav JS. Chromosome numbers and sex-determining mechanisms in adephagan Coleoptera. *Coleopt Bull.* 1984;38:335–357. <https://www.jstor.org/stable/4008212>
- Sota T, Hori M, Scholtz C, Karagyan G, Liang HB, Ikeda H, Takami Y. The origin of the giant ground beetle *Aplothorax burchelli* on St Helena Island. *Biol J Linn Soc Lond.* 2020;131(1):50–60. <https://doi.org/10.1093/biolinnean/blaa093>.
- Sota T, Takamo Y, Ikeda H, Liang H, Karagyan G, Scholtz C, Hori M. Global dispersal and diversification in ground beetles of the subfamily Carabinae. *Mol Phylogenet Evol.* 2022;167:107355. <https://doi.org/10.1016/j.ympev.2021.107355>.
- Stork NE. How many species of insects and other terrestrial arthropods are there on Earth? *Annu Rev Entomol.* 2018;63(1):31–45. <https://doi.org/10.1146/annurev-ento-020117-043348>.
- Su Z-H, Imura Y, Osawa S. Evolutionary history of Calosomina ground beetles (Coleoptera, Carabidae, Carabinae) of the world as deduced from sequence comparisons of the mitochondrial ND 5 gene. *Gene.* 2005;360(2):140–150. <https://doi.org/10.1016/j.gene.2005.06.028>.
- Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* 2015;43(12):e78. <https://doi.org/10.1093/nar/gkv227>.
- Toussaint EFA, Gauthier J, Bilat J, Gillett CPDT, Gough HM, Lundkvist H, Blanc M, Muñoz-Ramírez CP, Alvarez N. HyRAD-X exome capture museomics unravels giant ground beetle evolution. *Genome Biol Evol.* 2021;13(7):evab112. <https://doi.org/10.1093/gbe/evab112>.
- Toussaint EFA, Gillett CPDT. Rekindling Jeannel's Gondwanan vision? Phylogenetics and evolution of Carabinae with a focus on Calosoma caterpillar hunter beetles. *Biol J Linn Soc Lond.* 2018;123(1):191–207. <https://doi.org/10.1093/biolinnean/blx128>.
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K; Darwin Tree of Life Consortium; Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, et al. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics.* 2023;24(1):288. <https://doi.org/10.1186/s12859-023-05385-y>.
- UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2023;51(D1):D523–D531. <https://doi.org/10.1093/nar/gkac1052>.
- Vangestel C, Swaegers J, De Corte Z, Dekoninck W, Gharbi K, Gillespie R, Vandekerckhove M, Van Belleghem S, Hendrickx F. Chromosomal inversions from an initial ecotypic divergence drive a gradual repeated radiation of Galápagos beetles. *Sci Adv.* 2024;10(22):eadk7906. <https://doi.org/10.1126/sciadv.adk7>.
- Vasilikopoulos A, Balke M, Kukowka S, Pflug JM, Martin S, Meusemann K, Hendrich L, Mayer C, Maddison DR, Niehuis O, et al. Phylogenomic analyses clarify the pattern of evolution of Adephaga (Coleoptera) and highlight phylogenetic artefacts due to model misspecification and excessive data trimming. *Syst Entomol.* 2021;46(4):991–1018. <https://doi.org/10.1111/syen.12508>.
- Vasimuddin M, Misra S, Li H, Aluru S. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Rio de Janeiro (Brazil): IEEE; 2019. p. 314–324.
- Whiting MF, Bradler S, Maxwell T. Loss and recovery of wings in stick insects. *Nature.* 2003;421(6920):264–267. <https://doi.org/10.1038/nature01313>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 2023;39(1):btac808. <https://doi.org/10.1093/bioinformatics/btac808>.

Associate editor: John Wang