Journal of Heredity, 2024, **XX**, 1–9
https://doi.org/10.1093/jhered/esae044
Advance access publication 22 August 2024
Genome Resources

Genome Resources

An unusually large genome from an unusually large stonefly: A chromosome-length genome assembly for the giant salmonfly, *Pteronarcys californica* (Plecoptera: Pteronarcyidae)

Anna Eichert[1,2,*,†], John Sproul[3,†], Ethan R. Tolman[1,4], Jackson Birrell[5], Jared Meek[6], Jacqueline Heckenhauer[7,8], Charles Riley Nelson[3], Olga Dudchenko[9], Jiyun Jeong[9], David Weisz[9], Erez Lieberman Aiden[9], Scott Hotaling[10], Jessica L. Ware[1] and Paul B. Frandsen[11,*]

[1]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, United States,
[2]Richard Gilder Graduate School, American Museum of Natural History, New York, NY, United States,
[3]Department of Biology and Bean Life Science Museum, Brigham Young University, Provo, UT, United States,
[4]Department of Biological Sciences, Virginia Tech University, Blacksburg, VA, United States,
[5]The Salmonfly Project, Missoula, MT, United States,
[6]Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY, United States,
[7]Department of Terrestrial Zoology, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Hesse, Germany,
[8]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Hesse, Germany,
[9]Department of Molecular and Human Genetics, DNAZoo, The Center for Genome Architecture, Baylor College of Medicine, Houston, TX, United States,
[10]Department of Watershed Sciences, Utah State University, Logan, UT, United States,
[11]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, United States

[†]Equally contributing author.
[*]Corresponding authors: Anna Eichert, Richard Gilder Graduate School, American Museum of Natural History, New York, NY, USA, aeichert@amnh.org; Paul Frandsen, Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84602, USA, paul_frandsen@byu.edu

Corresponding Editor: Beth Shapiro

Abstract

Pteronarcys californica (Newport 1848) is commonly referred to as the giant salmonfly and is the largest species of stonefly (Insecta: Plecoptera) in the western United States. Historically, it was widespread and abundant in western rivers, but populations have experienced a substantial decline in the past few decades, becoming locally extirpated in numerous rivers in Utah, Colorado, and Montana. Although previous research has explored the ecological variables conducive to the survivability of populations of the giant salmonfly, a lack of genomic resources hampers exploration of how genetic variation is spread across extant populations. To accelerate research on this imperiled species, we present a de novo chromosomal-length genome assembly of *P. californica* generated from PacBio HiFi sequencing and Hi-C chromosome conformation capture. Our assembly includes 14 predicted pseudo chromosomes and 98.8% of Insecta universal core orthologs. At 2.40 gigabases, the *P. californica* assembly is the largest of available stonefly assemblies, highlighting at least a 9.5-fold variation in assembly size across the order. Repetitive elements account for much of the genome size increase in *P. californica* relative to other stonefly species, with the content of Class I retroelements alone exceeding the entire assembly size of all but two other species studied. We also observed preliminary suborder-specific trends in genome size that merit testing with more robust taxon sampling.

Key words: aquatic insects, conservation, long-read assembly, Plecoptera

1. Introduction

Plecoptera (stoneflies) are a cosmopolitan order of polyneopteran aquatic insects that inhabit streams, rivers, and lakes on every continent except Antarctica. There are currently more than 4,000 recognized species belonging to 17 families (Ding et al. 2019; DeWalt et al. 2024). To date, genome assemblies are publicly available for nine stonefly species, six of which include chromosome-length scaffolding. Here, we describe genome assembly for *Pteronarcys californica*, the first from the stonefly family Pteronarcyidae.

Pteronarcys californica, known to the fly-fishing community as the "giant salmonfly," typically dwells in large, cold,

Received May 28, 2024; Accepted August 20, 2024

fast-flowing rivers underneath medium-large rocks among the benthic substrate (Fig. 1). They are ecologically, culturally, and economically important (Albertson et al. 2022). For example, due to their large body size (>6 cm in length) and frequent high densities, they are a common prey item for aquatic consumers like trout (Nehring 2011; Anderson et al. 2019), and their synchronous emergence to adulthood provides an important seasonal food resource for terrestrial consumers, including birds, spiders, amphibians, and small mammals (Walters et al. 2018; Albertson et al. 2022). They are also prized among anglers, as their synchronized adult emergences create world-renowned fly-fishing opportunities, supporting local fishing economies.

Giant salmonflies are merovoltine, with a 3- to 5-year aquatic nymphal life cycle (DeWalt and Stewart 1995) and are sensitive to pollution, warming temperatures, flow modification, land-use change, sedimentation, and other environmental stressors (Anderson et al. 2019; Birrell et al. 2019, Kowalski and Richer 2020; Frakes et al. 2021, 2022). Anthropogenic influences such as water diversion, environmental degradation, the introduction of invasive species, and the accumulation of pollutants have led to the decline of available habitat for stoneflies in the United States (Morse et al. 1993; DeWalt et al. 2005; Birrell et al. 2019). *Pteronarcys californica* populations have experienced a substantial decline in the past few decades—undergoing reductions in population size or regional extinction in numerous rivers in Utah, Montana, and Colorado of the western United States (Fig. 1) (Colborn 1987; Stagliano 2010; Nehring et al. 2012; Anderson et al. 2019; Birrell et al. 2019; Kowalski and Richer 2020). To mitigate this loss and prevent extinction, increased conservation efforts, and population studies are essential. However, while ecological variables pertaining to the persistence of *P. californica* populations have been relatively well-recorded, the genomic features of this species (and family) have remained unexplored.

2. Materials and methods

2.1 Biological materials

Multiple *P. californica* nymphs were collected from the Diamond Fork River in Spanish Fork Canyon, Utah (N 40.067222°, W −111.439722°) and immediately flash-frozen with liquid nitrogen on 29 March 2022. The specimens were then stored in a −80 °C freezer prior to extraction. One specimen was used for whole genome sequencing and one specimen was used for Hi-C library preparation and sequencing.

2.2 DNA extraction and sequencing

We extracted high molecular weight DNA from thoracic tissue using a Qiagen Genomic-tip DNA extraction kit. We then sheared extracted DNA to 18 kbp using a Diagenode Megaruptor and prepared sequencing libraries with the PacBio SMRTbell Express Template Prep Kit 2.0 following manufacturer instructions. One sequencing library was sequenced across four PacBio Sequel II 30-h Single Molecule Real Time (SMRT) cells in circular consensus sequencing (CCS) mode. To ensure high genome coverage, we prepared an additional library from the same individual and sequenced the library on a single PacBio Revio SMRT cell (Pacbio. 2024b). To scaffold contigs into pseudochromosomes, we used high throughput chromosome conformation capture (Hi-C) (Van Burkum et al. 2010; Belton et al. 2012). In situ Hi-C data was generated following the protocol described in Rao et al. (2014). The resulting data was used to scaffold the hifiasm draft to chromosome length using the methodology described in www.dnazoo.org/methods.

2.3 Nuclear genome assembly

We estimated genome size and repetitive content from kmer frequencies using KMC 3 (Kokot et al. 2017) and GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) following GenomeScope 2.0 documentation (all software and commands used to perform these analyses are provided in Supplementary Table 1). HiFi reads generated using PacBio Sequel II was computed from raw subreads using PacBio SMRTlink v13.1 software (available at https://www.pacb.com/smrt-link/). To generate additional HiFi reads, we used DeepConsensus v.1.0.0 (Baid et al. 2022). In short, to generate a draft consensus, we ran pbbioconda ccs 6.4.0, which generates circular consensus reads and, in the process, removes adapters, with the --min-rq = 0.88 flag (Pacbio. 2024a). Next, we used actc (Töpfer and Kronenberg 2022) to align subreads to the draft consensus sequence. Then, we used DeepConsensus to generate polished reads by using gap-aware sequence transformers to correct errors in the CCS reads. The default setting is for sequences generated on the Revio instrument to be processed with DeepConsensus. The HiFi reads were then assembled into contigs using default parameters in hifiasm v0.16.1-r375 (Cheng et al. 2021). Following assembly in hifiasm, we further removed putative duplicated contigs using the purge_dups pipeline (Guan et al. 2020). The Hi-C data was processed using Juicer (Durand et al. 2016), and used as input into the 3D-DNA pipeline (Dudchenko et al. 2017) to produce a candidate chromosome-length genome assembly. We performed additional manual curation on the scaffolds using Juicebox Assembly Tools (Durand et al. 2016; Dudchenko et al. 2018). The contact matrices generated by aligning the Hi-C data to the genome assembly before and after the Hi-C scaffolding are available for browsing at multiple resolutions at https://www.dnazoo.org/assemblies/pteronarcys_californica using Juicebox.js, a cloud-based visualization system for Hi-C data (Robinson et al. 2018).

We evaluated contamination in the genome using two methods 1) the NCBI Foreign Contamination Screening tools (Astashyn et al. 2024) and 2) blobtools (Laetsch and Blaxter 2017). We then removed contigs that were assigned to Bacteria, plants, or Chordata (see Supplementary Fig. 1).

2.4 Annotation

To characterize repetitive elements (REs) in the *P. californica* assembly we identified and classified repeats using RepeatModeler v2.0.5 (Flynn et al. 2020), and then annotated REs in the assembly using RepeatMasker v4.1.5 (Smit and Hubley 2019). The library used by RepeatMasker during annotation (i.e. "-lib" flag) combined the species-specific repeat library output by RepeatModeler (i.e. the consensi.fa.classified file) with a database of Arthropoda REs which we extracted from RepeatMasker's internal library using the queryRepeatDatabase.pl script in the RepeatMasker/util directory. We generated a transposable element (TE) landscape plot by first processing the ".align" file produced by RepeatMasker with the calcDivergenceFromAlign.pl script (RepeatMasker/util) and then plotting the resulting output

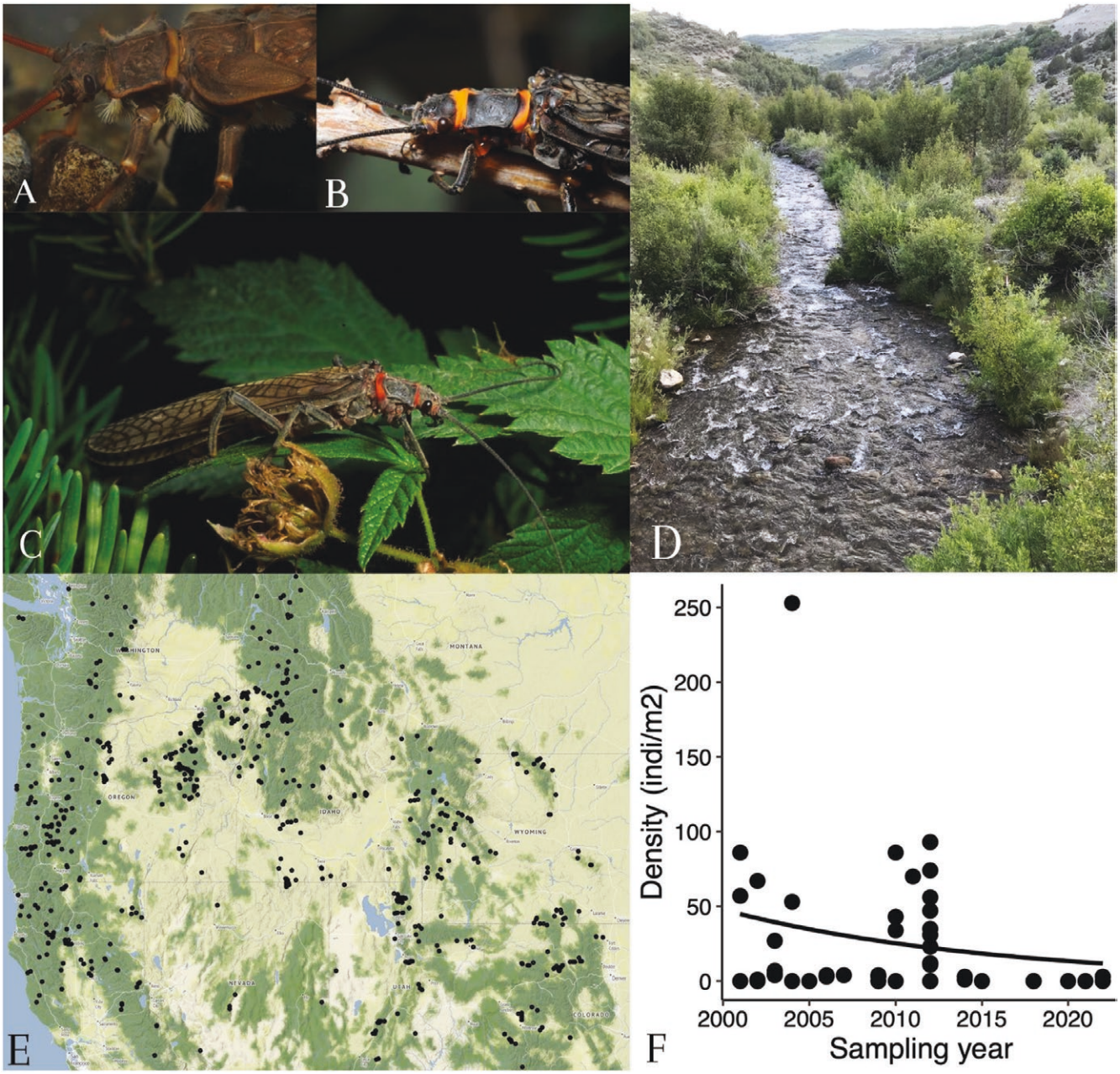

Fig. 1. Giant salmonfly photographs, habitat, distribution, and decline. A) Photos of a nymph and B) an adult *Pteronarcys californica*; C) full-body photo of an adult; D) photo of Diamond Fork River, UT, where the individuals sequenced for this study were collected; E) map of species distribution; and F) plot of individual densities over time within Utah, United States, based on sampling data from the National Aquatic Monitoring Center (NAMC) (https://namc-usu.org/data) and fitted to a negative binomial generalized linear model.

in R v4.3.1 (R Core Team 2023). We also characterized REs in the assemblies of nine additional stonefly species for which assemblies were available on NCBI: *Amphinemura sulcicollis* (GCA_001676325.1; MacDonald et al. 2016), *Brachyptera putata* (GCA_907164805.1), *Isoperla grammatica* (GCA_945910005.1; MacDonald et al. 2016), *Lednia tumana* (GCA_003287335.1; Hotaling et al. 2019), *Leuctra nigra* (GCA_934045905.1; Clifford et al. 2023), *Nemoura dubitans* (GCA_921293005.1; Farr and Macadam 2023), *Nemurella pictetii* (GCA_921293315.2; Macadam et al. 2022), *Protonemura montana* (GCA_947568835.1; Dixon and Macadam 2023), and *Sweltsa coloradensis* (GCA_024449915.1; Malison et al. 2022). For each additional species, we followed the above protocol for RE analysis except that the species-specific RepeatModeler library for the respective species was merged with the Arthropoda repeat database prior to annotation in RepeatMasker.

To annotate gene sequences we assembled the transcriptome of *Pteronarcys scotti* using publicly available RNA sequencing reads (South et al. 2021) with Trinity v2.8.5 (Grabherr et al. 2011). We then translated the transcriptome assembly into amino acid sequences with TransDecoder (Haas 2024) and used translated sequences as evidence for gene prediction with GALBA (Brůna et al. 2023) on the softmasked genome assembly. Following gene prediction in GALBA, we used blastP 2.9.0 + (Camacho et al. 2009) with a filter of 1e−25

to identify annotated genes that had orthologues in the annotated protein sets of the polyneopteran species *Bacillus rossius* (Lavanchy et al. 2024), *Dryococelus australis* (Stuart et al. 2023), *Coptotermes formosanus* (Itakura et al. 2020), *Diploptera punctata* (Fouks et al. 2023), *Periplaneta americana* (Li et al. 2018), *Schistocerca americana* (Childers et al. 2021), *Schistocerca gregaria* (Verlinden et al. 2021), and *Timema podura* (Comeault et al. 2016). This filtering approach has previously been shown to increase the specificity of GALBA annotations by removing spurious gene annotations, with a minimal reduction of sensitivity (Tolman et al. 2023). We evaluated the completeness of the annotation with the insecta_odb10 dataset with BUSCO 4.1.4 (Kriventseva et al. 2019; Seppey et al. 2019; Manni et al. 2021), and assigned gene ontology (GO) terms for the recovered amino acid sequences with InterProScan 1.8.0_402 (Jones et al. 2014).

3. Results

The GenomeScope 2.0 profile indicated that the genome of *P. californica* is substantially repetitive (49%) with an estimated genome length of 1.7 Gbp (Fig. 2). The final duplicate-purged and contaminant-filtered genome assembly length, however, was 2.40 Gbp, which is the largest for any stonefly species sequenced to date. The assembly length was much larger than the estimate from GenomeScope 2.0, likely due to the undercounting of repetitive kmers as demonstrated previously on large insect genomes (Pflug et al. 2020). We recovered 98.83% of Insecta universal core orthologs, suggesting high gene completeness (Table 1). We annotated a set of ~25,000 genes, containing 96.8% of the universal core orthologs. Approximately 14,000 of the genes were assigned GO terms by InterProScan.

Following repeat masking, we found that REs accounted for 61.4% of the *P. californica* assembly. Of those REs that could be classified, Class I retroelements comprised 11.4% of the assembly with long interspersed nuclear elements (LINEs, 7.82%) and long terminal repeats (LTRs, 2.64%) being the most abundant Class I elements, followed by Penelope (1.17%) and small interspersed nuclear elements (SINEs, 0.91%). Class II elements (transposons) accounted for 7.1% of the assembly with hAT elements being the most abundant Class II superfamily (2.64%) followed by CMC (1.23%) and Tc1/mariner (0.79%) superfamilies. Simple repeats were the largest category of non-interspersed repeats at 4.1% of the assembly. The large majority of REs (60.6%) were unclassified, accounting for 37.2% of the overall assembly.

Across publicly available stonefly assemblies, REs accounted for much of the variation in genome size and were particularly abundant in *P. californi*ca. For example, the LINE + LTR content of *P. californica* alone exceeded the entire assembly size of seven of the nine additional stonefly species sampled (Fig. 3). DNA transposons showed the largest increase in genomic proportion in *P. californica* compared to the relative abundance in all other species. The fraction of unclassified repeats was high for all species (average fraction of all repeats = 70.1%, range = 60.6% to 83.5%), a trend consistent with prior observations in Plecoptera (Sproul et al. 2023). Although the resolution of TE landscape plots is limited by the fraction of unclassified repeats, those REs that could be classified as well as the varied distribution of RE peaks along the *x* axis suggest dynamic evolutionary histories of REs in stoneflies. For example, they show evidence of major recent TE expansions (e.g. LTRs in *L. nigra*, Fig. 3) to various historical bursts of activity (e.g. SINEs in *L. tumana* and other repeats (Helitrons) in *Protonemura montana*).

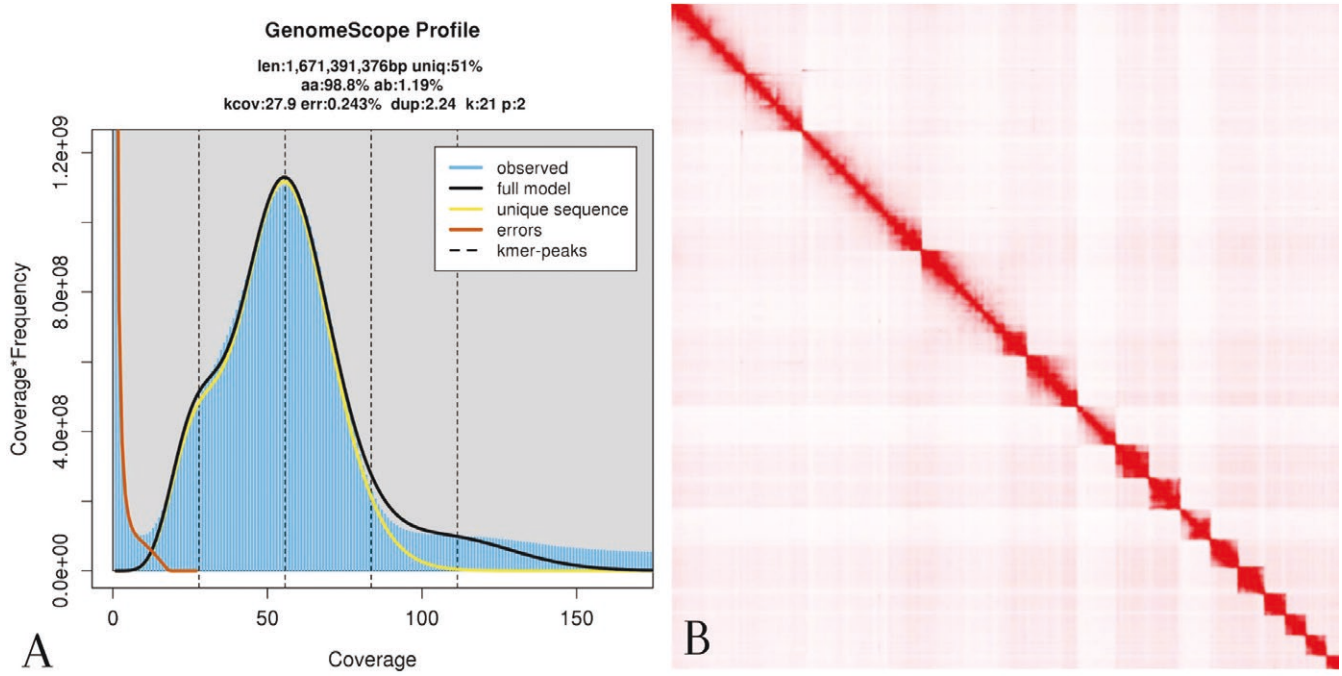

Fig. 2. Kmer histogram and Hi-C contact map. A) Genomescope Profile of kmers derived from HiFi reads. B) Hi-C contact map of the 14 chromosome-length scaffolds, numbered by size for convenience https://www.dnazoo.org/assemblies/pteronarcys_californica.

Table 1. Current stonefly genome assemblies and quality statistics as of 2024.

NCBI accession	Family	Species	Contig N50 (Mb)	Contig L50	Genome size (Mb)	Type of genome	Coverage	BUSCO
GCA_947568835.1	Nemouridae	*Protonemura montana*	6.1	13	258.5	Chromosome	83×	C: 99.42% [S: 98.76%, D: 0.66%], F: 0.07%, M: 0.51%
GCA_001676325.1	Nemouridae	*Amphinemura sulcicollis*	0.000846	66,116	271.9	Contig	1.4×	C: 51.94% [S: 51.5%, D: 0.44%], 35.41%, M: 11.12%
GCA_921293315.2	Nemouridae	*Nemurella pictetii*	9.6	9	257.0	Chromosome	68×	C: 99.64% [S: 98.98%, D: 0.66%], F: 0.15%, M: 0.22%
GCA_921293005.1	Nemouridae	*Nemoura dubitans*	5.2	19	321.0	Chromosome	62×	C: 99.49% [S: 98.61%, D: 0.88%], F: 0.07%, M: 0.44%
GCA_003287335.1	Nemouridae	*Lednia tumana*	4.2	18,613	304.5	Scaffold	60×	C: 72.06% [S: 71.40%, D: 0.66%], F:20.99%, M: 5.93%
GCA_907164805.1	Taeniopterygidae	*Brachyptera putata*	1.5	61	436.5	Chromosome	64×	C: 98.54% [S: 98.10%, D: 0.44%], F: 0.59%, M: 0.88%
GCA_024449915.1	Chloroperlidae	*Sweltsa coloradensis*	0.254	1,257	1,472	Contig	35×	C: 97.14% [S: 93.12%, D: 4.02%], F: 1.90%, M: 0.95%, n: 1367
GCA_934045905.1	Leuctridae	*Leuctra nigra*	0.587	277	536.3	Chromosome	32×	C: 99.56% [S: 97.95%, D: 1.61%], F: 0.15%, M: 0.29%
GCA_945910005.1	Perlodidae	*Isoperla grammatica*	0.653	362	874.6	Chromosome	34×	C: 99.42% [S: 97.66%, D: 1.76%], F: 0.22%, M: 0.37%
JBCHJG000000000	Pteronarcyidae	*Pteronarcys californica*	3.15	203	2,402	Chromosome	55×	C: 98.83% [S: 96.93%, D: 1.9%], F:0.66%, M:0.51%

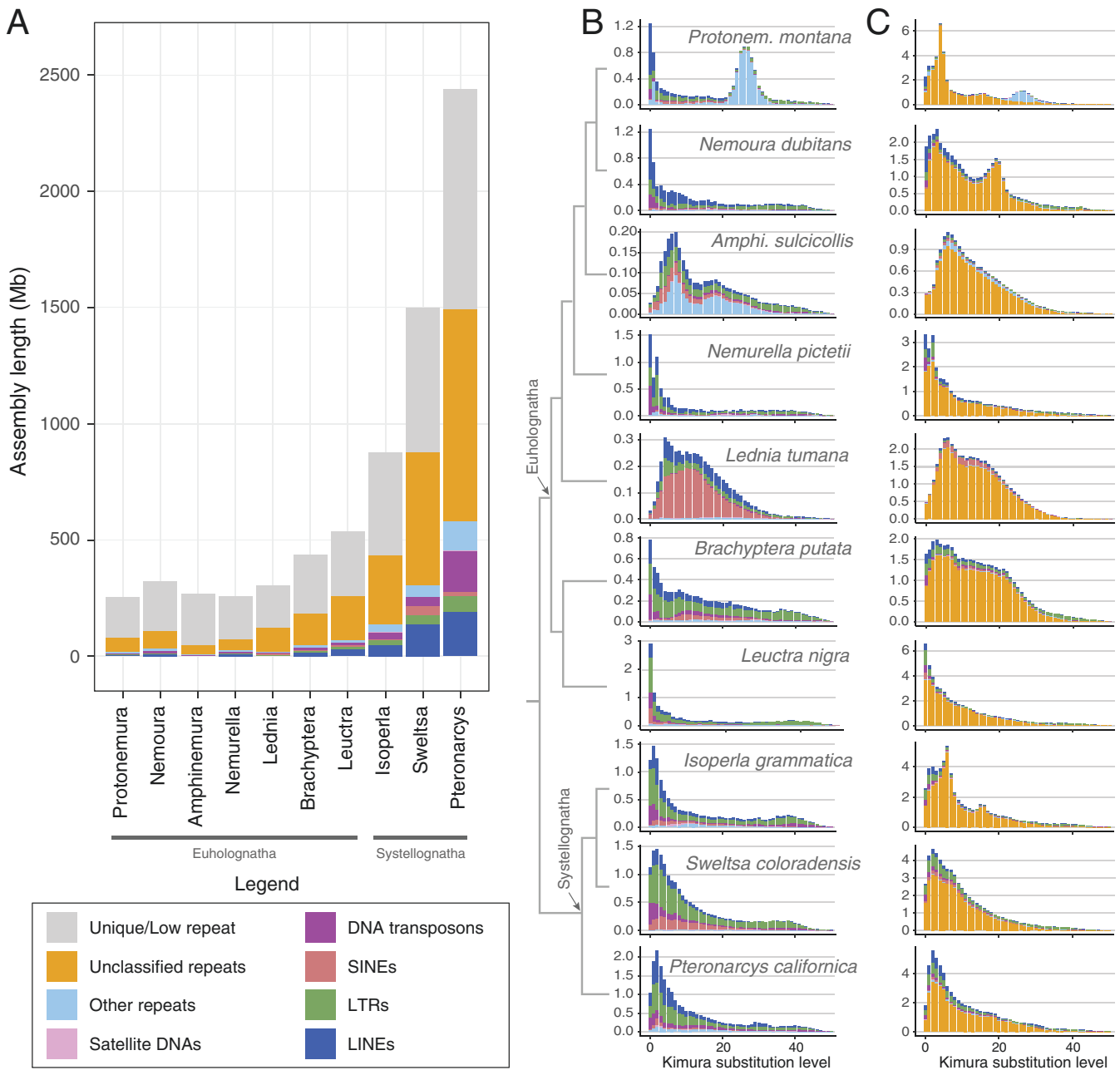

Fig. 3. Repetitive element content in stonefly genome assemblies. A) Summary of assembly content for 10 stonefly species with repetitive element content shown by different colors; B) transposable element (TE) landscape plots of classified TEs in the same 10 stonefly species as are summarized in with phylogenetic relationships indicated by the tree to the left of the plots. The *x* axis indicates sequence divergence (CpG adjusted Kimura distance) relative to consensus sequences for abundant TE categories. The *y* axis indicates TE abundance as a percentage of the assembly. Peaks in TE abundance shifted to the left of the *x* axis show less relative divergence from the consensus and are presumed to represent more recent historical TE expansion whereas abundance peaks that are right-skewed are presumed to represent relatively older historical TE expansions. C) See the previous explanation for (B), except that the unclassified repeats are included in TE landscape plots.

4. Discussion

Here we present a chromosome-length genome assembly for *P. californica*; the first species sequenced in the stonefly family Pteronarcyidae. Our aim was to provide a genomic resource to complement ongoing ecological research into an organism that is disappearing from rivers across western North America.

At 2.40 Gbp, the *P. californica* genome assembly was the largest stonefly genome sequenced to date. We provide evidence that the expansion of REs accounts for much of the genome size increase in *P. californica* relative to other stonefly lineages (Fig. 3, Table 1). In Plecoptera, genome sizes vary widely due to RE expansion (Fig. 3, Table 1), similar to many groups of insects and other eukaryotes (Blommaert 2020; Cong et al. 2022; Heckenhauer et al. 2022). Our findings reveal a preliminary trend in which species belonging to infraorder Euholognatha (i.e. families Capniidae, Leuctridae, Nemouridae, Notomemouridae, and Taeniopterygidae) have smaller, less repetitive genomes than species in infraorder Systellognatha (i.e. families Perlodidae, Chloroperlidae, and

Pteronarcyidae), but additional sequenced samples in both infraorders are needed to verify this trend. Additional sampling would also allow for testing of whether ecological (e.g. habitat use) and life history strategies (e.g. voltinism, body size), or population demographics (e.g. effective population size), are correlated with patterns of genome size across the order.

For all stonefly species, a high proportion of REs were unclassified (average = 70.1% stdev = 6.7%), which limits resolution regarding which REs groups have been the most important drivers of genome size evolution. A similarly high proportion of unclassified REs was previously noted within Plecoptera (*Lednia*) and other early branching insect lineages (e.g. Ephemeroptera) and correlates with poor representation of curated REs for these groups in RE databases (e.g. Repbase/Dfam) (Sproul et al. 2023, see Fig 3D and Supplementary Table S1). As a critical mass of stonefly assemblies accumulates, efforts to build a curated RE library for the group can enable higher-resolution insights from REs and accelerate our understanding of their genome evolution.

While thoroughly addressing genome size evolution in *Pteronarcys* requires additional study within the species and order, we present a brief discussion here to guide future work. Two observations that correlate with the large genome size of *P. californica* are its substantial body size (among the largest of all stonefly species) and its merovoltine life cycle. It is conceivable that both factors may lead to a relatively reduced effective population size for *P. californica* compared to other stonefly species, particularly those with genome assemblies already available. For example, the large body size of benthic invertebrates may limit their potential habitat as they require larger substrates with sufficient interstitial space to avoid predation or dislodging due to water currents (Douglas and Lake 1994; Buffagni et al. 1995). Merovoltinism exposes cohorts to the risk of predation/death over 3 to 5 years as nymphs, before their emergence as adults. As effective population size (N_e) decreases, the efficacy of natural selection at purging slightly deleterious mutations (e.g. repetitive element accumulation), also decreases. Based on this N_e-mediated balance between selection and drift, and the broad link between body size and N_e observed across eukaryotes, it has been proposed that body size is an important driver of genome size and complexity (Lynch and Conery 2003).

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

We express gratitude to the BYU DNA Sequencing Center which provided the PacBio HiFi sequencing service to help us obtain the whole genome assembly. Genome assembly was done in collaboration with the DNA Zoo Consortium (www.dnazoo.org).

Funding

This study was funded by start-up funding to Jessica Ware from the American Museum of Natural History, start-up funding to Paul Frandsen from the BYU College of Life Sciences, and funding from the Charles Redd Center for Western Studies.

Conflict of interest statement. The authors declare no conflict of interest.

Author Contributions

Anna Eichert (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing), John Sproul (Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing), Ethan R. Tolman (Data curation, Formal analysis, Methodology, Software, Writing - review & editing), Jackson Birrell (Conceptualization, Visualization, Writing - review & editing), Jared Meek (Visualization, Writing - original draft), Jacqueline Heckenhauer (Formal analysis, Methodology, Writing - original draft, Writing - review & editing), C. Riley Nelson (Conceptualization), Olga Dudchenko (Data curation, Formal analysis, Methodology, Visualization, Writing - review & editing), Jiyun Jeong (Data curation, Formal analysis, Methodology), David Weisz (Data curation, Formal analysis), Erez Lieberman Aiden (Data curation, Formal analysis), Scott Hotaling (Conceptualization, Data curation, Methodology, Resources, Writing - original draft, Writing - review & editing), Jessica L. Ware (Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing), and Paul B. Frandsen (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing)

Data Availability

The genome assembly of *Pteronarcys californica* and associated reads are available from the NCBI BioProject PRJNA1090442. Hi-C contact maps for *P. californica* are interactively available at https://www.dnazoo.org/assemblies/pteronarcys_californica. Supplementary files are available at https://figshare.com/s/63e6ea20401fcf72742f for review and will be made public upon manuscript acceptance.

References

Albertson LK, Briggs MA, Maguire Z, Swart S, Cross WF, Twining CW, Wesner JS, Baxter CV, Walters DM. Dietary composition and fatty acid content of giant salmonflies (*Pteronarcys californica*) in two Rocky Mountain rivers. Ecosphere. 2022:13:e3904. https://doi.org/10.1002/ecs2.3904

Anderson HE, Albertson LK, Walters DM. Water temperature drives variability in salmonfly abundance, emergence timing, and body size. River Res Appl. 2019:35:1013–1022. https://doi.org/10.1002/rra.3464

Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strope PK, Sylla PM, Wagner L, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. 2024:25:60. https://doi.org/10.1186/s13059-024-03198-7

Baid G, Cook DE, Shafin K, Yun T, Llinares-López F, Berthet Q, Belyaeva A, Töpfer A, Wenger AM, Rowell WJ, et al. DeepConsensus improves the accuracy of sequences with a gap-aware sequence

transformer. Nat Biotechnol. 2023:41:232–238. https://doi.org/10.1038/s41587-022-01435-7
Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods. 2012:58:268–276. https://doi.org/10.1016/j.ymeth.2012.05.001
Birrell JH, Meek JB, Nelson CR. Decline of the giant salmonfly *Pteronarcys californica* newport, 1848 (Plecoptera: Pteronarcyidae) in the Provo river. Utah, USA: Illiesia; 2019.
BlobTools: Interrogation of genome assemblies | F1000Research. Retrieved 2 July 2024. 2017. https://f1000research.com/articles/6-1287/v1
Blommaert J. Genome size evolution: towards new model systems for old questions. Proc Biol Sci. 2020:287:20201441. https://doi.org/10.1098/rspb.2020.1441
Brůna T, Li H, Guhlin J, Honsel D, Herbold S, Stanke M, Nenasheva N, Ebel M, Gabriel L, Hoff KJ. Galba: genome annotation with miniprot and AUGUSTUS. BMC Bioinf. 2023:24:327. https://doi.org/10.1186/s12859-023-05449-z
Buffagni A, Crosa G, Marchetti R. Size-related shifts in the physical habitat of two mayfly species (Ephemeroptera). Freshw Biol. 1995:34:297–302. https://doi.org/10.1111/j.1365-2427.1995.tb00889.x
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10:421. https://doi.org/10.1186/1471-2105-10-421
Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5
Childers AK, Geib SM, Sim SB, Poelchau MF, Coates BS, Simmonds TJ, Scully ED, Smith TPL, Childers CP, Corpuz RL, et al. The USDA-ARS Ag100Pest initiative: high-quality genome assemblies for agricultural pest arthropod research. Insects. 2021:12:626. https://doi.org/10.3390/insects12070626
Clifford C, Macadam CR, Price BW; Natural History Museum Genome Acquisition Lab. The genome sequence of the black needle fly, *Leuctra nigra* (Olivier, 1811). Wellcome Open Res. 2023:8:95. https://doi.org/10.12688/wellcomeopenres.19082.1
Colborn T. *Pteronarcys californica* Newport: mortality and cadmium accumulation at three stations on the Gunnison River, Gunnison County, Colorado. Heavy Metal Environ. 1987:2:439–441.
Comeault AA, Carvalho CF, Dennis S, Soria-Carrasco V, Nosil P. Color phenotypes are under similar genetic control in two distantly related species of Timema stick insect. Evolution. 2016:70:1283–1296. https://doi.org/10.1111/evo.12931
Cong Y, Ye X, Mei Y, He K, Li F. Transposons and non-coding regions drive the intrafamily differences of genome size in insects. iScience. 2022:25:104873. https://doi.org/10.1016/j.isci.2022.104873
DeWalt RE, Favret C, Webb DW. Just how imperiled are aquatic insects? A case study of stoneflies (Plecoptera) in Illinois. Ann Entomol Soc Am. 2005:98:941–950. https://doi.org/10.1603/0013-8746(2005)098[0941:JHIAAI]2.0.CO;2
DeWalt RE, Hopkins H, Neu-Becker U, and Stueber G Plecoptera species file [dataset]. Retrieved 22 February 2024. 2024. https://plecoptera.speciesfile.org/
DeWalt RE, Stewart KW. Life histories of stoneflies (plecoptera) in the Rio Conejos of Southern Colorado. Great Basin Nat. 1995:55:1–18.
Ding S, Li W, Wang Y, Cameron SL, Murányi D, Yang D. The phylogeny and evolutionary timescale of stoneflies (Insecta: Plecoptera) inferred from mitochondrial genomes. Mol Phylogenet Evol. 2019:135:123–135. https://doi.org/10.1016/j.ympev.2019.03.005
Dixon A, Macadam CR, Darwin Tree of Life Consortium. The genome sequence of the spring stonefly, Protonemura montana (Kimmins, 1941). Wellcome Open Res. 2023:8:333.
Douglas M, Lake PS. Species richness of stream stones: an investigation of the mechanisms generating the species-area relationship. Oikos. 1994:69:387–396. https://doi.org/10.2307/3545851
Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017:356:92–95. https://doi.org/10.1126/science.aal3327
Dudchenko O, Shamim MS, Batra SS, Durand, NC, Musial NT, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. BioRxiv. 2018. https://doi.org/10.1101/254797
Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016:3:95–98. https://doi.org/10.1016/j.cels.2016.07.002
Farr A, Macadam CR, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Darwin Tree of Life Consortium. The genome sequence of a stonefly, Nemoura dubitans (Morton, 1894). Wellcome Open Res. 2023:8:329.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117
Fouks B, Harrison MC, Mikhailova AA, Marchal E, English S, Carruthers M, Jennings EC, Chiamaka EL, Frigard RA, Pippel M, et al. Live-bearing cockroach genome reveals convergent evolutionary mechanisms linked to viviparity in insects and beyond. iScience. 2023:26:107832. https://doi.org/10.1016/j.isci.2023.107832
Frakes JI, Birrell JH, Shah AA, Woods HA. Flow increases tolerance of heat and hypoxia of an aquatic insect. Biol Lett. 2021:17:20210004. https://doi.org/10.1098/rsbl.2021.0004
Frakes JI, Malison RL, Sydor MJ, Arthur Woods H. Exposure to copper increases hypoxia sensitivity and decreases upper thermal tolerance of giant salmonfly nymphs (*Pteronarcys californica*). J Insect Physiol. 2022:143:104455. https://doi.org/10.1016/j.jinsphys.2022.104455
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011:29:644–652. https://doi.org/10.1038/nbt.1883. PMID: 21572440; PMCID: PMC3571712
Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36:2896–2898. https://doi.org/10.1093/bioinformatics/btaa025
Haas B. TransDecoder [Computer software]. GitHub. 2024. Retrieved 02 July 2024. https://github.com/TransDecoder/TransDecoder
Heckenhauer J, Frandsen PB, Sproul JS, Li Z, Paule J, Larracuente AM, Maughan PJ, Barker MS, Schneider JV, Stewart RJ, et al. Genome size evolution in the diverse insect order Trichoptera. GigaScience. 2022:11:giac011. https://doi.org/10.1093/gigascience/giac011
Hotaling S, Kelley JL, Weisrock DW. Nuclear and mitochondrial genomic resources for the meltwater stonefly (Plecoptera: Nemouridae), *Lednia tumana* (Ricker, 1952). Aquat Insects. 2019:40:362–369. https://doi.org/10.1080/01650424.2019.1639764
Itakura S, Yoshikawa Y, Togami Y, Umezawa K. Draft genome sequence of the termite, *Coptotermes formosanus*: genetic insights into the pyruvate dehydrogenase complex of the termite. J Asia Pac Entomol. 2020:23:666–674. https://doi.org/10.1016/j.aspen.2020.05.004
Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014:30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031
Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 2017:33:2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Kowalski DA, Richer EE. Quantifying the habitat preferences of the stonefly *Pteronarcys californica* in Colorado. River Res Appl. 2020:36:2043–2050. https://doi.org/10.1002/rra.3733

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2019:47:D807–D811. https://doi.org/10.1093/nar/gky1053

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research. 2017:6:1287. https://doi.org/10.12688/f1000research.12232.1

Lavanchy G, Brandt A, Bastardot M, Dumas Z, Labédan M, Massy M, Toubiana W, Tran Van P, Luchetti A, Scali V, et al. Evolution of alternative reproductive systems in Bacillus stick insects. Evolution. 2024:78:1109–1120. https://doi.org/10.1093/evolut/qpae045

Li S, Zhu S, Jia Q, Yuan D, Ren C, Li K, Liu S, Cui Y, Zhao H, Cao Y, et al. The genomic and functional landscapes of developmental plasticity in the American cockroach. Nat Commun. 2018:9:1008. https://doi.org/10.1038/s41467-018-03281-1

Lynch M, Conery JS. The origins of genome complexity. Science. 2003:302:1401–1404.

Macadam C, Clifford C, Price BW; Natural History Museum Genome Acquisition Lab. The genome sequence of a stonefly, *Nemurella pictetii* Klapalek, 1900. Wellcome Open Res. 2022:7:56. https://doi.org/10.12688/wellcomeopenres.17684.2

Macdonald HC, Cunha L, Bruford MW. Development of genomic resources for four potential environmental bioindicator species: Isoperla grammatica, *Amphinemura sulcicollis*, *Oniscus asellus* and *Baetis rhodani*. BioRxiv. 2016. https://doi.org/10.1101/046227

Malison RL, Hand BK, Winter E, Giersch JJ, Amish SJ, Whited D, Stanford JA, Luikart G. Landscape connectivity and genetic structure in a Mainstem and a Tributary Stonefly (*Plecoptera*) species using a novel reference genome. J Hered. 2022:113:453–471. https://doi.org/10.1093/jhered/esac025

Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021:1:e323. https://doi.org/10.1002/cpz1.323

Morse J, Stark B, McCafferty W. Southern Appalachian streams at risk: implications for mayflies, stoneflies, caddisflies, and other aquatic biota. Aquat Conserv Mar Freshw Ecosyst. 1993:3:293–303.

Nehring B. Colorado river aquatic resource investigations [Job Progress Report]. Colorado Parks & Wildlife; 2011. p. 1–9. https://www.researchgate.net/profile/Brian-Heinold/publication/281120443_Colorado_River_Aquatic_Resources_Investigations_Federal_Aid_Project_F-237R-18/links/55d775fb08aeb38e8a85ad0d/Colorado-River-Aquatic-Resources-Investigations-Federal-Aid-Project-F-237R-18.pdf

Nehring B. Colorado river aquatic resource investigations [Job Progress Report]. Colorado Parks & Wildlife; 2012. p. 1–9. 26 June 2024. https://cpw.state.co.us/Documents/Research/Aquatic/pdf/Publications/CORiverAquaticResourceInvestigations2012.pdf

Newport G. Proceedings of the Linnean Society of London. 1848. 20 June 2024. http://archive.org/details/proceedingsoflin00linn

Pacbio. pbbioconda ccs software. 2024a. 25 June 2024. https://github.com/PacificBiosciences/pbbioconda

Pacbio. SMRTLink V13.1 software. 2024b. 24 June 2024. https://www.pacb.com/smrt-link/

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3 (Bethesda, Md.). 2020:10:3047–3060. https://doi.org/10.1534/g3.120.401028

R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023. 29 June 2024. https://www.R-project.org/

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432. https://doi.org/10.1038/s41467-020-14998-3

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014:159:1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js provides a cloud-based visualization system for Hi-C data. Cell Syst. 2018:6:256–258.e1. https://doi.org/10.1016/j.cels.2018.01.001

Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor, Gene prediction: methods and protocols. New York: Springer; 2019. p. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Smit A, Hubley R. RepeatMasker (Open-4.1) [Computer software]. Software Downloads. 2019 [accessed 3 July 2024]. https://www.pacb.com/support/software-downloads/

South EJ, Skinner RK, DeWalt RE, Kondratieff BC, Johnson KP, Davis MA, Lee JJ, Durfee RS. Phylogenomics of the North American *Plecoptera*. Syst Entomol. 2021:46:287–305. https://doi.org/10.1111/syen.12462

Sproul JS, Hotaling S, Heckenhauer J, Powell A, Marshall D, Larracuente AM, Kelley JL, Pauls SU, Frandsen PB. Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. Genome Res. 2023:33:1708–1717. https://genome.cshlp.org/content/33/10/1708

Stagliano D. Evaluation of salmonflies in Montana's rivers: are statewide populations really declining? Report to Montana Natural Heritage Program, Helena, Montana; 2010. 29 pp.

Stuart OP, Cleave R, Magrath MJL, Mikheyev AS. Genome of the lord Howe island stick insect reveals a highly conserved phasmid X chromosome. Genome Biol Evol. 2023:15:evad104. https://doi.org/10.1093/gbe/evad104

Tolman ER, Beatty CD, Frandsen PB, Bush RJ, Burchim OR, Driever ES, Harding KM, Jordan D, Kohli MK, Park J, et al. Newly sequenced genomes reveal patterns of gene family expansion in select Dragonflies (Odonata: Anisoptera). bioRxiv. 2023:12.

Töpfer A, Kronenberg Z. actc, https://github.com/PacificBiosciences/actc. 2022.

van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. Hi-C: a method to study the three-dimensional architecture of fenomes. J Vis Exp. 2010:39:1869. https://doi.org/10.3791/1869

Verlinden H, Sterck L, Li J, Li Z, Yssel A, Gansemans Y, Verdonck R, Holtof M, Song H, Behmer ST, et al. First draft genome assembly of the desert locust, *Schistocerca gregaria*. F1000Research. 2021:9:775. https://doi.org/10.12688/f1000research.25148.2

Walters DM, Wesner JS, Zuellig RE, Kowalski DA, Kondratieff MC. Holy flux: spatial and temporal variation in massive pulses of emerging insect biomass from western U.S. rivers. Ecology. 2018:99:238–240.