**Nuclear and mitochondrial genomic resources for the meltwater stonefly, *Lednia tumana*
Ricker, 1952 (Plecoptera: Nemouridae)**

Scott Hotaling^{a,b*}, Joanna L. Kelley^a, and David W. Weisrock^b

^a School of Biological Sciences, Washington State University, Pullman, WA, USA; ^b Department
of Biology, University of Kentucky, Lexington, KY, USA

* Scott Hotaling, School of Biological Sciences, Washington State University, Pullman, WA,
99164, USA; Email: scott.hotaling@wsu.edu; Phone: (828) 507-9950; ORCID: 0000-0002-
5965-0986

Abstract:

With more than 3,700 described species, stoneflies (Order Plecoptera) are an important component of global aquatic biodiversity. The meltwater stonefly *Lednia tumana* (Ricker, 1952; Family Nemouridae) is endemic to alpine streams of Glacier National Park and has been petitioned for listing under the U.S. Endangered Species Act (ESA) due to climate change-induced loss of alpine glaciers and snowfields. Here, we present *de novo* assemblies of the nuclear (~520 million base pairs [bp]) and mitochondrial (15,014-bp) genomes for *L. tumana*. The *L. tumana* nuclear genome is the most complete stonefly genome reported to date, with ~71% of genes present in complete form and more than 4,600 contigs longer than 10-kilobases (kb). The *L. tumana* mitochondrial genome is the second for the family Nemouridae and the first from North America. Together, both genomes represent important foundational resources, setting the stage for future efforts to understand the evolution of *L. tumana*, stoneflies, and aquatic insects worldwide.

Keywords: Plecoptera; Nemouridae; *Lednia*; genomics; nuclear genome; USA

Introduction:

Stoneflies are a diverse, globally distributed group of hemimetabolous insects that diverged from their closest relatives (e.g., Orthoptera, Dermaptera, Zoraptera) at least 300 million years ago in the Carboniferous Period (Béthoux, Cui, Kondratieff, Stark, and Ren 2011). With more than 3,700 described species, stoneflies account for a substantial portion of freshwater biodiversity (DeWalt, Kondratieff, and Sandberg 2015). The meltwater stonefly, *Lednia tumana* (Ricker, 1952; Plecoptera: Nemouridae), resides in alpine streams of Glacier National Park (GNP), USA, where it is iconic of habitat loss due to climate change in the region (Muhlfeld et al. 2011; Giersch, Hotaling, Kovach, Jones and Muhlfeld 2017). *Lednia tumana* is one of four extant species in the genus *Lednia* which all exhibit alpine, cold-water distributions in western North America (Baumann and Kondratieff 2010; Baumann and Call 2012). The majority of *L. tumana*'s habitat is supported by seasonal melting of permanent ice and snow, a habitat type that is under considerable threat of near-term loss as the global cryosphere recedes (Hotaling, Finn, Giersch, Weisrock, and Jacobsen 2017; Hotaling et al. 2019). The recent evolutionary history of *L. tumana* is closely tied to glacier dynamics with present-day genetic

clusters arising in parallel with ice sheet recession at the end of the Pleistocene (~20,000 years ago, Hotaling et al. 2018). Genetic evidence has also highlighted a possible loss of mitochondrial genetic diversity for the species on even more recent, decadal timescales (Jordan et al. 2016). With such a narrow habitat niche in a small, mountainous region of the northern Rocky Mountains, *L. tumana* has been recommended for listing under the U.S. Endangered Species Act (US Fish & Wildlife Service 2016).

In this study, we present an assembly of the nuclear genome for *L. tumana*, the most complete nuclear genome for the order Plecoptera reported to date. This resource also represents one of only three high-coverage (> 50x) genomes for any EPT taxon (Ephemeroptera, *Ephemera danica*, Polechau et al. 2014; Plecoptera, this study; Trichoptera, *Stenopsyche tienmushanensis*, Luo, Tang, Frandsen, Stewart, and Zhou 2018), a globally important group of aquatic organisms commonly used for biological monitoring (e.g., Tronstad, Hotaling, and Bish 2016). We also present a nearly complete mitochondrial genome assembly (mitogenome) for *L. tumana*, the fourth for the stonefly family Nemouridae after two previous studies (Chen and Du 2017; Cao, Wang, Huang, and Li 2019).

Materials and Methods:

Genomic DNA was extracted using a Qiagen DNeasy Blood & Tissue Kit from a single *L. tumana* nymph collected in 2013 from Lunch Creek in GNP. Prior to extraction, both the head and as much of the digestive tract as possible were removed. A whole-genome shotgun sequencing library targeting a 250-bp fragment size was constructed and sequenced by the Florida State University Center for Genomics. The library was sequenced twice on 50% of an Illumina HiSeq2500 flow cell each time with paired-end, 150-bp chemistry, resulting in 242,208,840 total reads. The size of the *L. tumana* nuclear genome was estimated using a kmer-based approach in sga preQC (Simpson 2014). Read quality was assessed with fastQC (Andrews 2010) and low-quality reads were either trimmed or removed entirely using TrimGalore (Krueger 2015) with the flags: --stringency 3 --quality 20 --length 40. We assembled the nuclear genome using SPAdes v3.11.1 with default settings (Bankevich et al. 2012) and generated summary statistics with the Assemblathon2 perl script (assemblathon_stats.pl, Bradnam et al. 2013). The completeness of our nuclear assembly was assessed by calculating the number of conserved single copy orthologous genes [Benchmarking Universal Single-Copy Orthologs (BUSCOs)] in

the assembly using BUSCO v3 and the 1,658 “insecta_ob9” set of reference genes (Simão, Waterhouse, Ioannidis, Kriventseva, and Zdobnov 2015). To compare the completeness of the *L. tumana* genome in the context of other stoneflies, we downloaded the two other publicly available stonefly genomes for *Isoperla grammatica* (Poda, 1761; Perlodidae) and *Amphinemura sulciollis* (Stephens, 1836; Nemouridae) which are deposited under GenBank BioProject PRJNA315680 (Macdonald, Cunha, and Bruford 2016; Macdonald, Ormerod, and Bruford 2017). We performed the same BUSCO and Assemblathon2 analyses on the *I. grammatica* and *A. sulciollis* genomes as we did for *L. tumana* above.

We assembled the *L. tumana* mitogenome with NOVOPlasty v2.6.7 (Dierckxsens, Mardulyn, and Smits 2016) using an 872-bp segment of the *L. tumana* *cytb* gene (GenBank KX212756.1) as the “seed” sequence. After assembly, the mitogenome was annotated through a combination of the MITOS web server with default settings (Bernt et al. 2013) and comparison to the *Nemoura nanikensis* mitogenome (Plecoptera: Nemouridae; Chen and Du 2017). However, in this initial assembly, the 16S gene was fragmented and the 12S gene was missing entirely. To mitigate this, we extracted sequences for both genes from the *N. nanikensis* mitogenome (Plecoptera: Nemouridae; Chen and Du 2017) and mapped our raw reads to these reference sequences for each gene with BWA-MEM v0.7.12-r1039 (Li 2013) using the default settings. Next, we used bcftools v1.9 (Li, Orti, Zhang, and Lu 2009) to collect summary information on the read mapping and genotype likelihoods (‘mpileup’ with default settings), called consensus nucleotides (‘call’ with -m flag), and output the consensus sequence (‘consensus’ with default settings). Finally, we used samtools v1.7 (Li et al. 2009) to calculate coverage depth per nucleotide for each sequence and masked consensus bases with no coverage (i.e., bases with no information from the *L. tumana* read mapping). We manually integrated our 16S and 12S sequences into the *L. tumana* mitogenome through comparison with the *N. nanikensis* mitogenome (Chen and Du 2017) and re-annotated the assembly with the MITOS web server (Bernt et al. 2013) as described above.

Results and Discussion:

The size of the *L. tumana* nuclear genome was estimated to be 536.7-megabases (Mb) from raw sequence data. Our final *L. tumana* genome assembly was 520.2-Mb with 50% of the assembly in contigs ≥ 4.69 -kilobases (kb; Figure 1a, Table 1). The assembled genome size is in

line with the only other publicly available stonefly genomes, *I. grammatica* (509.5 Mb) and *A. sulcicollis* (271.9 Mb). The *L. tumana* genome assembly also includes ~3,800 more contigs > 10 kb than the *A. sulcicollis* genome and ~4,600 more than the *I. grammatica* assembly (Figure 1a, Table 1). All associated data for the resources detailed in this study, including both raw reads and assemblies, are available as part of GenBank BioProject PRJNA472568 (mitogenome: MH374046, nuclear genome: SAMN09295077, raw reads: SRP148706).

The meltwater stonefly's nuclear genome is similarly A/T-rich (58.4%) to the other stoneflies (58–60%; Table 1), ants (55–67%; Gadau et al. 2012), *Drosophila melanogaster* (58%), *Anopheles gambiae* (56%), and the honeybee, *Apis mellifera* (67%, The Honeybee Genome Sequencing Consortium 2006). However, the *L. tumana* genome is far more complete in terms of genic regions than both existing stonefly assemblies with 92.8% of BUSCO reference genes either complete (70.6%) or fragmented (22.2%) versus 80.8% for *A. sulcicollis* (50.5% complete, 31.3% fragmented) and just 50.1% for *I. grammatica* (13.3% complete, 36.8% fragmented; Figure 1b, Table 1).

The *L. tumana* mitogenome is nearly complete, covering 15,014-bp, including all 13 protein-coding genes, 21 tRNA genes, both rRNA genes, and is only missing the control region (Figure 2). The organization of the *L. tumana* mitogenome is similar to that of *N. nankinensis*, the only other mitogenome available for the family Nemouridae. In *N. nankinensis*, the control region is ~1-kb which indicates that the complete *L. tumana* mitogenome is likely around 16-kb. This predicted size would fall in line with mitogenome sizes reported for other stoneflies (Chen and Du 2017). Our inability to resolve the control region is also unsurprising. In *N. nankinensis*, the control region contains a large ~1-kb repeat region which is inherently difficult to resolve without targeted long-range PCR re-sequencing or longer read high-throughput sequencing.

Conclusion:

The increasing availability of genome assemblies for a wide array of organisms is rapidly expanding the scope and comparative power of modern genome biology (Hotelling and Kelley 2019). With more than 4,600 contigs longer than 10-kb and ~70% of genes assembled in their complete form, the *L. tumana* nuclear genome provides new opportunity for exploring genome evolution within Plecoptera, a highly diverse, globally distributed insect order, or at higher levels of taxonomic organization (e.g., across all insects). Specifically, the *L. tumana* genome could be

mined for genes for phylogenomic studies (e.g., Li et al. 2007; Borowiec, Lee, Chiu, and Plachetzki 2015) or more targeted, comparative assessments of specific genes or gene families across many taxa to clarify evolutionary shifts and/or copy number variation (e.g., Baalsrud et al. 2017).

Moreover, single-copy orthologous genes compared among species can provide a means for quantifying differences in evolutionary rates across diverse taxa (e.g., Honeybee Genome Sequencing Consortium 2006) and/or to identify rapidly evolving genes that underlie evolutionary transitions of interest. In the case of stoneflies and aquatic biodiversity in general, little is known of the evolutionary changes underlying the shift to an aquatic larval stage that is common among many orders (e.g., Plecoptera, Ephemeroptera, Trichoptera). With the addition of the *L. tumana* nuclear genome reported here to the recently published caddisfly (*S. tienmushanensis*, Order Trichoptera, Luo et al. 2018) and mayfly (*E. danica*, Order Ephemeroptera, Polechau et al. 2014) genomes, the stage is now set for broad, genome-scale investigations of how this major life history transition occurred across three globally distributed insect orders.

Future efforts to refine both assemblies, including the incorporation of longer reads (e.g., generated using Pacific Biosciences sequencing technology, Utturkar et al. 2014) will yield greater insight into the genome biology of *L. tumana*, stoneflies, and aquatic insects broadly. Still, the resources provided here, and particularly the most complete stonefly nuclear genome published thus far, represent an important step towards empowering modern stonefly research, a globally relevant group of aquatic insects that has been largely overlooked in the genomic age.

Acknowledgements:

We thank Alan Lemmon and Emily Lemmon for advice and assistance with sequencing. This research was supported by the University of Kentucky (UK) and National Science Foundation (DEB-0949532). Computational resources were provided by the UK Center for Computational Sciences and the Lipscomb High Performance Computing Cluster, as well as the Washington State University Center for Institutional Research Computing.

Declaration of Interest:

The authors declare no financial interest or benefit stemming from this research.

References:

- Andrews, S. (2010), 'FastQC: a quality control tool for high throughput sequence data',
<https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc>
- Baalsrud, H. T., Tørresen, O. K., Solbakken, M. H., Salzburger, W., Hanel, R., Jakobsen, K. S.,
and Jentoft, S. (2017), 'De novo gene evolution of antifreeze glycoproteins in codfishes
revealed by whole genome sequence data', *Molecular Biology and Evolution*, 35, 593–606.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
Nikolenko, S.I., Pham, S., and Prjibelski, A.D. (2012), 'SPAdes: a new genome assembly
algorithm and its applications to single-cell sequencing', *Journal of Computational Biology*,
19, 455–477.
- Baumann, R.W., and Kondratieff, B.C. (2010), 'The stonefly genus *Lednia* in North America
(Plecoptera: Nemouridae)', *Illiesia*, 6, 315–327.
- Baumman, R.W., and Call, R.G. (2012), '*Lednia tetonica*, a new species of stonefly from
Wyoming (Plecoptera: Nemouridae)', *Illiesia*, 8, 104–110.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J.,
Middendorf, M., and Stadler, P.F. (2013), 'MITOS: improved de novo metazoan
mitochondrial genome annotation', *Molecular Phylogenetics and Evolution*, 69, 313–319.
- Béthoux, O., Cui, Y., Kondratieff, B., Stark, B., and Ren, D. (2011), 'At last, a Pennsylvanian
stem-stonefly (Plecoptera) discovered', *BMC Evolutionary Biology*, 11, 248.
- Borowiec, M. L., Lee, E. K., Chiu, J. C., and Plachetzki, D. C. (2015), 'Extracting phylogenetic
signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister
to remaining Metazoa', *BMC Genomics*, 16, 987.
- Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S.,
Chapman, J.A., Chapuis, G., and Chikhi, R. (2013), 'Assemblathon 2: evaluating de novo
methods of genome assembly in three vertebrate species', *GigaScience*, 2, 10.

- 192 Cao, J. J., Wang, Y., Huang, Y. R., and Li, W. H. (2019), 'Mitochondrial genomes of the
193 stoneflies *Mesonemoura metafiligera* and *Mesonemoura tritaenia* (Plecoptera,
194 Nemouridae), with a phylogenetic analysis of Nemouroidea', *ZooKeys*, 835, 43.
- 195 Chen, Z.-T., and Du, Y.-Z. (2017), 'First mitochondrial genome from Nemouridae (Plecoptera)
196 reveals novel features of the elongated control region and phylogenetic implications',
197 *International Journal of Molecular Sciences*, 18, 996.
- 198 DeWalt R.E., Kondratieff B.C., and Sandberg J.B. 2015, 'Order Plecoptera', Thorp J., Rogers
199 D.C. (eds), *Ecology and General Biology: Freshwater Invertebrates*. – Academic Press,
200 Cambridge.
- 201 Dierckxsens, N., Mardulyn, P., and Smits, G. (2016), 'NOVOPlasty: de novo assembly of
202 organelle genomes from whole genome data', *Nucleic Acids Research*, 45, e18.
- 203 Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D.F., Smith, C.R., Suen, G., Wurm,
204 Y., and Smith, C.D. (2012), 'The genomic impact of 100 million years of social evolution
205 in seven ant species', *Trends in Genetics*, 28, 14–21.
- 206 Giersch, J.J., Hotaling, S., Kovach, R.P., Jones, L.A., and Muhlfeld, C.C. (2017), 'Climate-
207 induced glacier and snow loss imperils alpine stream insects', *Global Change Biology*, 23,
208 2577–2589.
- 209 Honeybee Genome Sequencing Consortium. (2006), 'Insights into social insects from the
210 genome of the honeybee *Apis mellifera*', *Nature*, 443, 931.
- 211 Hotaling, S., Finn, D.S., Giersch, J.J., Weisrock, D.W., and Jacobsen, D. (2017), 'Climate change
212 and alpine stream biology: progress, challenges, and opportunities for the future',
213 *Biological Reviews*, 92, 2024–2045.
- 214 Hotaling, S., Giersch, J.J., Finn, D.S., Tronstad, L.M., Jordan, S., Serpa, L.E., Call, R.G.,
215 Muhlfeld, C.C., and Weisrock, D.W. (2019), 'Congruent population genetic structure but
216 differing depths of divergence for three alpine stoneflies with similar ecology and
217 geographic distributions', *Freshwater Biology*, 64, 335–347.

218 Hotaling, S., and Kelley, J.L. (2019), 'The rising tide of high-quality genomic resources',
219 *Molecular Ecology Resources*, 19, 567–569.

220 Hotaling, S., Muhlfeld, C.C., Giersch, J.J., Ali, O.A., Jordan, S., Miller, M.R., Luikart, G., and
221 Weisrock, D.W. (2018), 'Demographic modelling reveals a history of divergence with gene
222 flow for a glacially tied stonefly in a changing post-Pleistocene landscape', *Journal of*
223 *Biogeography*, 45, 304–317.

224 Jordan, S., Giersch, J.J., Muhlfeld, C.C., Hotaling, S., Fanning, L., Tappenbeck, T.H., and
225 Luikart, G. (2016), 'Loss of genetic diversity and increased subdivision in an endemic
226 alpine stonefly threatened by climate change', *PLoS One*, 11, e0157386.

227 Krueger, F. (2015), 'Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently
228 apply quality and adapter trimming to FastQ files'.
229 https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

230 Li, C., Ortí, G., Zhang, G., and Lu, G. (2007), 'A practical approach to phylogenomics: the
231 phylogeny of ray-finned fish (Actinopterygii) as a case study', *BMC Evolutionary*
232 *Biology*, 7, 44.

233 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
234 and Durbin, R. (2009) 'The sequence alignment/map format and SAMtools',
235 *Bioinformatics*, 25, 2078–2079.

236 Li, H. (2013), 'Aligning sequence reads, clone sequences and assembly contigs with BWA-
237 MEM', *arXiv*, 1303, 3997.

238 Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., and Zhou, X. (2018), 'The genome of an
239 underwater architect, the caddisfly *Stenopsyche tienmushanensis* Hwang (Insecta:
240 Trichoptera)', *GigaScience*, Doi: 10.1093/gigascience/giy143.

241 Macdonald, H.C., Cunha, L., and Bruford, M.W. (2016), 'Development of genomic resources for
242 four potential environmental bioindicator species: *Isoperla grammatica*, *Amphinemura*
243 *sulcicollis*, *Oniscus asellus* and *Baetis rhodani*', *bioRxiv*, 046227.

244 Macdonald, H.C., Ormerod, S.J., and Bruford, M.W. (2017), 'Enhancing capacity for freshwater
245 conservation at the genetic level: a demonstration using three stream
246 macroinvertebrates', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 27, 452–
247 461.

248 Muhlfeld, C.C., Giersch, J.J., Hauer F.R., Pederson, G.T., Luikart, G., Peterson, D.P., Downs,
249 C.C., and Fagre, D.P. (2011), 'Climate change links fate of glaciers and an endemic alpine
250 invertebrate', *Climatic Change*, 106, 337–345.

251 Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., and Hackett, K.
252 (2014), 'The i5k Workspace@ NAL—enabling genomic data access, visualization and
253 curation of arthropod genomes', *Nucleic Acids Research*, 43, D714–D719.

254 Ricker, W. E. (1952), *'Systematic studies in Plecoptera'*, Indiana University, Bloomington.

255 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015),
256 'BUSCO: assessing genome assembly and annotation completeness with single-copy
257 orthologs', *Bioinformatics*, 31, 3210–3212.

258 Simpson, J.T. (2014), 'Exploring genome characteristics and sequence quality without a
259 reference', *Bioinformatics*, 30, 1228–1235.

260 Tronstad, L.M., Hotaling, S., and Bish, J.C. (2016), 'Longitudinal changes in stream invertebrate
261 assemblages of Grand Teton National Park, Wyoming', *Insect Conservation and Diversity*,
262 9, 320–331.

263 US Fish & Wildlife Service. (2016), 'Endangered and threatened wildlife and plants; 12-month
264 finding on a petition to list the western glacier stonefly as an endangered or threatened
265 species; proposed threatened species status for Meltwater Lednian Stonefly and Western
266 Glacier Stonefly', *Federal Register*, 81, 68379–68397.

267 Utturkar, S.M., Klingeman, D.M., Land, M.L., Schadt, C.W., Doktycz, M.J., Pelletier, D.A., and
268 Brown, S.D. (2014), 'Evaluation and validation of de novo and hybrid assembly techniques
269 to derive high-quality genome sequences', *Bioinformatics*, 30, 2709–2716.

Tables:

Table 1. Assembly statistics for the nuclear genome of *Lednia tumana* (Ricker, 1952; Plecoptera: Nemouridae) and two other stonefly species, *Amphinemura sulcicollis* (Stephens, 1836) and *Isoperla grammatica* (Poda, 1761; Macdonald et al. 2016; Macdonald et al. 2017). BUSCOs: Single-copy, orthologous genes known to be highly conserved among insects. A total of 1,658 BUSCOs were searched.

	<i>L. tumana</i>	<i>A. sulcicollis</i>	<i>I. grammatica</i>
Estimated genome size	536,700,000	n/a	n/a
Assembly size	520,200,814	271,924,966	509,522,935
Coverage	~60x	~1.4x	~0.7x
Contigs > 1-kb	74,445	51,555	53,204
Contigs > 10-kb	4,608	849	4
Contig N50	4.69-kb	0.85-kb	0.46-kb
%A/T	58.4	58.2	59.9
%G/C	41.5	42.8	40.1
%N	0.1	0	0
Complete BUSCOs	1,172 (70.6%)	837 (50.5%)	221 (13.3%)
Fragmented BUSCOs	368 (22.2%)	519 (31.3%)	610 (36.8%)
Missing BUSCOs	118 (7.2%)	302 (18.2%)	827 (49.9%)

277 **Figures:**

278

279 Figure 1. Comparisons of the gene content and contiguity of the *Lednia tumana* (Ricker, 1952;
 280 Plecoptera: Nemouridae) nuclear genome to the two other previously published stonefly
 281 genomes for *Amphinemura sulcicollis* (Stephens, 1836) and *Isoperla grammatica* (Poda, 1761;
 282 Macdonald et al. 2016; Macdonald et al. 2017). (a) The number of contigs > 1-kb and > 10-kb
 283 across the three stonefly genomes. For *I. grammatica*, the assembly contains just four contigs
 284 greater than 10-kb. (b) Presence of highly conserved, single-copy orthologous genes (BUSCOs)
 285 across the three stonefly genomes.

286

287 Figure 2. (a) The mitogenome of *Lednia tumana* (Ricker, 1952; Plecoptera: Nemouridae). (b)

288 Locations of tRNAs. (c) Locations of protein coding (PCGs) and rRNA genes.