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

Scott Hotaling[1,2], Joanna L. Kelley[1], and David W. Weisrock[2]

Affiliations:

[1] School of Biological Sciences, Washington State University, Pullman, WA, USA

[2] Department of Biology, University of Kentucky, Lexington, KY, USA

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

Abstract:

With more than 3,500 described species, stoneflies (Order Plecoptera) are an important component of global aquatic biodiversity. The meltwater stonefly *Lednia tumana* (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 a *de novo* assembly of the nuclear genome (~305 million bp) as well as the nearly complete mitochondrial genome (13,752 bp) for *L. tumana*. To our knowledge, this represents the first high-coverage (>10x) nuclear genome for the stonefly order (Plecoptera) and second mitochondrial genome for the family Nemouridae. Both assemblies provide important foundational resources for future stonefly research.

Keywords: nuclear genome, Glacier National Park, mitochondrial genome, conservation genetics, global change biology

Running head: Nuclear and mitochondrial genomes for *Lednia tumana*

Introduction:

Stoneflies are a diverse, globally distributed group of hemimetabolous insects that diverged from their closest relatives (e.g., Orthoptera, Dermaptera, Zoraptera) ~275 million years

32 ago in the early Permian (Misof et al., 2014). With more than 3,500 described species, stoneflies
33 account for a substantial portion of freshwater biodiversity (Fochetti and De Figueroa, 2008).
34 The meltwater stonefly, *Lednia tumana* (Plecoptera: Nemouridae), resides in alpine streams of
35 Glacier National Park (GNP), USA, where it is iconic of habitat loss due to climate change in the
36 region (Giersch, Hotaling, Kovach, Jones and Muhlfeld, 2017). *Lednia tumana* is one of four
37 species in the genus *Lednia* which all have alpine, cold-water distributions in western North
38 America (Baumann and Kondratieff, 2010). The majority of *L. tumana*'s habitat is supported by
39 seasonal melting of permanent ice and snow, a habitat type that is under considerable threat of
40 near-term loss as the global cryosphere recedes (Hotaling, Finn, Joseph Giersch, Weisrock and
41 Jacobsen, 2017). The recent evolutionary history of *L. tumana* is also closely tied to glacier
42 dynamics with present-day genetic clusters arising in parallel with ice sheet recession at the end
43 of the Pleistocene (~20,000 years ago, Hotaling et al., 2018). Genetic evidence has also
44 highlighted a possible loss of mitochondrial genetic diversity on even more recent, decadal
45 timescales for the species (Jordan et al., 2016). With such a narrow habitat niche in a small,
46 mountainous region of the northern Rocky Mountains, *L. tumana* has been recommended for
47 listing under the U.S. Endangered Species Act (US Fish & Wildlife Service, 2016). In this study,
48 we present an assembly of the nuclear genome for *L. tumana*, the first high-coverage (>10x)
49 nuclear genome for the order Plecoptera. This resource also represents one of only three high-
50 coverage genomes for any EPT taxon (Ephemeroptera, Plecoptera, and Trichoptera), a globally
51 important group of aquatic organisms commonly used for biological monitoring (e.g., Tronstad,
52 Hotaling and Bish, 2016). We also present a mitochondrial genome assembly (mitogenome) for
53 *L. tumana*, the second for the stonefly family Nemouridae after *Nemoura nankinensis* (Chen and
54 Du, 2017).
55
56 **Materials and Methods:**
57 Genomic DNA was extracted using a Qiagen DNeasy Blood & Tissue Kit from a single
58 *L. tumana* nymph collected in 2013 from Lunch Creek in GNP. Prior to extraction, both the head
59 and as much of the digestive tract as possible were removed. A whole-genome shotgun
60 sequencing library targeting a 250-bp fragment size was constructed and sequenced by the
61 Florida State University Center for Genomics. The library was sequenced twice on 50% of an
62 Illumina HiSeq2500 flow cell each time with paired-end, 150-bp chemistry, resulting in

242,208,840 total reads. The size of the nuclear genome for *L. tumana* was estimated to be 536.7 mega basepairs (Mb) 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 as well as two other publicly available stonefly genomes (Macdonald, Cunha and Bruford, 2016; Macdonald, Ormerod and Bruford, 2017) was assessed by calculating the number of conserved single copy orthologs (BUSCOs) in the assemblies using BUSCO v3 and the 1,658 "insecta_ob9" set of reference genes (Simão, Waterhouse, Ioannidis, Kriventseva and Zdobnov, 2015). These BUSCO analyses provided a proxy for how complete genic regions are in both our own and the two existing stonefly genomes.

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. The mitogenome was annotated through a combination of the MITOS web server with default settings (Bernt et al., 2013) and comparison to the *Nemoura nankinensis* mitogenome (Plecoptera: Nemouridae; Chen and Du, 2017).

Results and Discussion:

The *L. tumana* nuclear genome assembly comprises just over 304.5 Mb with 50% of the assembly in contigs equal to or larger than 4.69 kilobases (kb; Table 1). The size of the genome is in line with two other publicly available stonefly genomes, *Isoperla grammatica* (509.5 Mb; Family Perlodidae) and *Amphinemura sulcicollis* (271.9 Mb; Family Nemouridae) which are both deposited under GenBank BioProject PRJNA315680 (Macdonald et al., 2016; Macdonald et al., 2017). 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 (58.2-59.9%; 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). Our assembly is more intact in terms of genic regions than both existing stonefly assemblies with 92.8% of BUSCOs 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; Table 1). However, our assembly is still highly fragmented, making it difficult to draw more nuanced conclusions about its structure and evolution.

The mitogenome is also nearly complete, covering 13,752 bp, including all 13 protein-coding genes, 21 tRNA genes, the 12S rRNA gene, and only missing the 16S rRNA gene and control region (Figure 1). 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 regions missing from the *L. tumana* mitogenome assembly are ~3 kb, indicating that the complete *L. tumana* mitogenome is likely around 16.7 kb, which falls in line with other stonefly mitogenomes (Chen and Du, 2017). Our inability to resolve the control region is perhaps unsurprising as the *N. nankinensis* control region contains a large, ~1 kb repeat region which is difficult to resolve without targeted long-range PCR re-sequencing or long-read high-throughput sequencing.

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* and related stoneflies. Still, the resources included here provide 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 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'.

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.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, 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.

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.

Chen, Z.-T. and Du, Y.-Z. (2017), 'First mitochondrial genome from Nemouridae (Plecoptera) reveals novel features of the elongated control region and phylogenetic implications', *International journal of molecular sciences*, 18, 996.

Dierckxsens, N., Mardulyn, P. and Smits, G. (2016), 'NOVOPlasty: de novo assembly of organelle genomes from whole genome data', *Nucleic Acids Res*, 45, e18-e18.

Fochetti, R. and De Figueroa, J.M.T. (2008), 'Global diversity of stoneflies (Plecoptera; Insecta) in freshwater', *Hydrobiologia*, 595, 365-377.

Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D.F., Smith, C.R., Suen, G., Wurm, Y. and Smith, C.D. (2012) 'The genomic impact of 100 million years of social evolution in seven ant species', *Trends in Genetics*, 28, 14-21.

Giersch, J.J., Hotaling, S., Kovach, R.P., Jones, L.A. and Muhlfeld, C.C. (2017), 'Climate-induced glacier and snow loss imperils alpine stream insects', *Global Change Biology*, 23, 2577-2589.

Honeybee Genome Sequencing Consortium. (2006), 'Insights into social insects from the genome of the honeybee Apis mellifera', *Nature*, 443, 931.

Hotaling, S., Finn, D.S., Joseph Giersch, J., Weisrock, D.W. and Jacobsen, D. (2017), 'Climate change and alpine stream biology: progress, challenges, and opportunities for the future', *Biological Reviews*, 92, 2024-2045.

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

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

Krueger, F. (2015), 'Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files'.

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

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

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspock, U., Aspock, H., Bartel, D., Blanke, A., Berger, S., Bohm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schutte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M. and Zhou, X. (2014), 'Phylogenomics resolves the timing and pattern of insect evolution', *Science*, 346, 763-767.

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

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

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

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

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

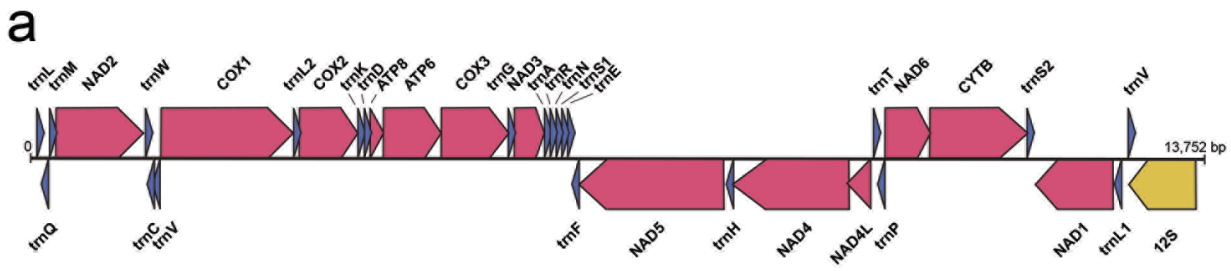
203 **Tables:**

204 Table 1. Assembly statistics for the nuclear genome of *Lednia tumana* (Plecoptera: Nemouridae)
205 and two other stonefly species, *Amphinemura sulcicollis* and *Isoperla grammatica* (Macdonald et
206 al., 2016; Macdonald et al., 2017). For *L.* tumana, only contigs larger than 500-bp were included.
207 "BUSCOs" are single-copy, orthologous genes known to be highly conserved among insects. A
208 total of 1,658 BUSCOs were searched.

	L. tumana	*A. sulcicollis*	*I. grammatica*
Estimated genome size	536,700,000	n/a	n/a
Assembly size	304,502,267	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	1172 (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%)

209

210 **Figures:**



b **tRNAs**

Name	Range	Strand	Length
trnL	176-241	F	66
trnQ	239-307	R	69
trnM	318-385	F	68
trnW	1434-1502	F	69
trnC	1495-1557	R	63
trnY	1561-1627	R	67
trnL2	3160-3225	F	66
trnK	3916-3985	F	70
trnD	3986-4052	F	67
trnG	5670-5735	F	66
trnA	6088-6151	F	64
trnR	6152-6214	F	63
trnN	6217-6283	F	67
trnS1	6284-6351	F	68
trnE	6352-6416	F	65
trnF	6415-6481	R	67
trnH	8217-8282	R	66
trnT	9918-9982	F	65
trnP	9983-10048	R	66
trnS2	11709-11778	F	70
trnL1	12751-12816	R	66
trnV	12891-12956	R	66

c **PCGs and rRNAs**

Name	Range	Strand	Length
NAD2	386-1393	F	1008
COX1	1620-3140	F	1521
COX2	3228-3908	F	681
ATP8	4053-4208	F	156
ATP6	4205-4873	F	669
COX3	4882-5667	F	786
NAD3	5739-6083	F	345
NAD5	6489-8168	R	1680
NAD4	8294-9625	R	1332
NAD4L	9622-9885	R	264
NAD6	10062-10565	F	504
CYTB	10574-11704	F	1131
NAD1	11808-12710	R	903
12S	12897-13680	R	784

Regions absent: 16S, control region

Protein coding gene (PCG)
rRNA gene
tRNA gene

211

212 Figure 1. (a) The nearly complete mitogenome of *Lednia tumana* (Plecoptera: Nemouridae). (b)
213 Location of tRNAs in the *L. tumana* mitogenome. (c) Location of protein coding genes (PCGs)
214 and rRNA genes in the *L. tumana* mitogenome.