

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

3

4 Scott Hotaling^{1,2}, Joanna L. Kelley¹, and David W. Weisrock²

5

6 **Affiliations:**

7 ¹ School of Biological Sciences, Washington State University, Pullman, WA, 99164, USA

8 ² Department of Biology, University of Kentucky, Lexington, KY, USA

9

10 **Correspondence:** Scott Hotaling, School of Biological Sciences, Washington State University,
11 Pullman, WA, 99164, USA; Email: scott.hotaling@wsu.edu; Phone: (828) 507-9950; ORCID:
12 0000-0002-5965-0986

13

14 **Abstract:**

15 With more than 3,700 described species, stoneflies (Order Plecoptera) are an important
16 component of global aquatic biodiversity. The meltwater stonefly *Lednia tumana* (Family
17 Nemouridae) is endemic to alpine streams of Glacier National Park and has been petitioned for
18 listing under the U.S. Endangered Species Act (ESA) due to climate change-induced loss of
19 alpine glaciers and snowfields. Here, we present *de novo* assemblies of the nuclear (~520 million
20 base pairs [bp]) and mitochondrial (13,752 bp) genomes for *L. tumana*. The *L. tumana* nuclear
21 genome is the most complete stonefly genome reported to date, with ~71% of genes present in
22 complete form and > 4,600 contigs longer than 10 kilobases (kb). The *L. tumana* mitochondrial
23 genome is the second for the family Nemouridae and the first from North America. Together,
24 both genomes represent important foundational resources, setting the stage for future efforts to
25 understand the evolution of *L. tumana*, stoneflies, and aquatic insects worldwide.

26

27 **Keywords:** stonefly genome, nuclear genome, mitochondrial genome, genome assembly, global
28 change biology, alpine stream

29

30 **Running head:** Genomic resources for the stonefly, *Lednia tumana*

31

32 **Introduction:**

33 Stoneflies are a diverse, globally distributed group of hemimetabolous insects that
34 diverged from their closest relatives (e.g., Orthoptera, Dermaptera, Zoraptera) at least 300
35 million years ago in the Carboniferous Period (Béthoux, Cui, Kondratieff, Stark and Ren, 2011).
36 With more than 3,700 described species, stoneflies account for a substantial portion of
37 freshwater biodiversity (DeWalt, Kondratieff and Sandberg, 2015). The meltwater stonefly,
38 *Lednia tumana* (Plecoptera: Nemouridae), resides in alpine streams of Glacier National Park
39 (GNP), USA, where it is iconic of habitat loss due to climate change in the region (Giersch,
40 Hotaling, Kovach, Jones and Muhlfeld, 2017). *Lednia tumana* is one of four extant species in the
41 genus *Lednia* which all have alpine, cold-water distributions in western North America
42 (Baumann and Kondratieff, 2010; Baumann and Call, 2012). The majority of *L. tumana*'s habitat
43 is supported by seasonal melting of permanent ice and snow, a habitat type that is under
44 considerable threat of near-term loss as the global cryosphere recedes (Hotaling, Finn, Giersch,
45 Weisrock and Jacobsen, 2017; Hotaling et al., in press). The recent evolutionary history of *L.*
46 *tumana* is closely tied to glacier dynamics with present-day genetic clusters arising in parallel
47 with ice sheet recession at the end of the Pleistocene (~20,000 years ago, Hotaling et al., 2018).
48 Genetic evidence has also highlighted a possible loss of mitochondrial genetic diversity for the
49 species on even more recent, decadal timescales (Jordan et al., 2016). With such a narrow habitat
50 niche in a small, mountainous region of the northern Rocky Mountains, *L. tumana* has been
51 recommended for listing under the U.S. Endangered Species Act (US Fish & Wildlife Service,
52 2016).

53 In this study, we present an assembly of the nuclear genome for *L. tumana*, the most
54 complete nuclear genome for the order Plecoptera reported to date. This resource also represents
55 one of only three high-coverage genomes for any EPT taxon (Ephemeroptera, Plecoptera, and
56 Trichoptera), a globally important group of aquatic organisms commonly used for biological
57 monitoring (e.g., Tronstad, Hotaling and Bish, 2016). We also present a nearly complete

58 mitochondrial genome assembly (mitogenome) for *L. tumana*, the second for the stonefly family
59 Nemouridae after *Nemoura nankinensis* (Chen and Du, 2017).

60

61 **Materials and Methods:**

62 Genomic DNA was extracted using a Qiagen DNeasy Blood & Tissue Kit from a single
63 *L. tumana* nymph collected in 2013 from Lunch Creek in GNP. Prior to extraction, both the head
64 and as much of the digestive tract as possible were removed. A whole-genome shotgun
65 sequencing library targeting a 250 bp fragment size was constructed and sequenced by the
66 Florida State University Center for Genomics. The library was sequenced twice on 50% of an
67 Illumina HiSeq2500 flow cell each time with paired-end, 150 bp chemistry, resulting in
68 242,208,840 total reads. The size of the *L. tumana* nuclear genome was estimated using a kmer-
69 based approach in sga preQC (Simpson, 2014). Read quality was assessed with fastQC
70 (Andrews, 2010) and low-quality reads were either trimmed or removed entirely using
71 TrimGalore (Krueger, 2015) with the flags: --stringency 3 --quality 20 --length 40. We
72 assembled the nuclear genome using SPAdes v3.11.1 with default settings (Bankevich et al.,
73 2012) and generated summary statistics with the Assemblathon2 perl script
74 (assemblathon_stats.pl, Bradnam et al., 2013). The completeness of our nuclear assembly was
75 assessed by calculating the number of conserved single copy orthologs (BUSCOs) in the
76 assemblies using BUSCO v3 and the 1,658 “insecta_ob9” set of reference genes (Simão,
77 Waterhouse, Ioannidis, Kriventseva and Zdobnov, 2015). These BUSCO analyses provided a
78 proxy for how complete genic regions are in both our own and the two existing stonefly
79 genomes. To compare the completeness of the *L. tumana* genome in the context of other
80 stoneflies, we downloaded the two other publicly available stonefly genomes for *Isoperla*
81 *grammatica* (Family Perlodidae) and *Amphinemura sulcicollis* (Family Nemouridae) which are
82 deposited under GenBank BioProject PRJNA315680 (Macdonald et al., 2016; Macdonald et al.,
83 2017) and ran the same Assemblathon2 and BUSCO analyses.

84 We assembled the *L. tumana* mitogenome with NOVOPlasty v2.6.7 (Dierckxsens,
85 Mardulyn and Smits, 2016) using an 872 bp segment of the *L. tumana cytb* gene (GenBank
86 KX212756.1) as the “seed” sequence. After assembly, the mitogenome was annotated through a

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

89

90 **Results and Discussion:**

91 The size of the *L. tumana* nuclear genome was estimated to be 536.7 megabases (Mb)
92 from raw sequence data. Our final *L. tumana* genome assembly was 520.2 Mb with 50% of the
93 assembly in contigs \geq 4.69 kilobases (kb; Figure 1a, Table 1). The assembled genome size is in
94 line with the only other publicly available stonefly genomes, *I. grammatica* (509.5 Mb) and *A.*
95 *sulcicollis* (271.9 Mb). The *L. tumana* genome assembly also includes ~3,800 more contigs $>$ 10
96 kb than the *A. sulcicollis* genome and ~4,600 more than the *I. grammatica* assembly (Figure 1a,
97 Table 1). All associated data for the resources detailed in this study, including both raw reads and
98 assemblies, are available as part of GenBank BioProject PRJNA472568 (mitogenome:
99 MH374046, nuclear genome: SAMN09295077, raw reads: SRP148706).

100 The meltwater stonefly's nuclear genome is similarly A/T-rich (58.4%) to the other
101 stoneflies (58.2-59.9%; Table 1), ants (55-67%; Gadau et al., 2012), *Drosophila melanogaster*
102 (58%), *Anopheles gambiae* (56%), and the honeybee, *Apis mellifera* (67%, The Honeybee
103 Genome Sequencing Consortium, 2006). However, the *L. tumana* genome is far more complete
104 in terms of genic regions than both existing stonefly assemblies with 92.8% of BUSCO reference
105 genes either complete (70.6%) or fragmented (22.2%) versus 80.8% for *A. sulcicollis* (50.5%
106 complete, 31.3% fragmented) and just 50.1% for *I. grammatica* (13.3% complete, 36.8%
107 fragmented; Figure 1b, Table 1).

108 The *L. tumana* mitogenome is nearly complete, covering 13,752 bp, including all 13
109 protein-coding genes, 21 tRNA genes, the 12S rRNA gene, and is only missing the 16S rRNA
110 gene and control region (Figure 2). The organization of the *L. tumana* mitogenome is similar to
111 that of *N. nankinensis*, the only other mitogenome available for the family Nemouridae. In *N.*
112 *nankinensis*, the regions missing from the *L. tumana* mitogenome assembly are ~3 kb, indicating
113 that the complete *L. tumana* mitogenome is likely around 16.7 kb, which is similar to
114 mitogenome sizes reported for other stoneflies (Chen and Du, 2017). Our inability to resolve the
115 control region is unsurprising as the *N. nankinensis* control region contains a large, ~1 kb repeat

116 region which is difficult to resolve without targeted long-range PCR re-sequencing or longer
117 read high-throughput sequencing.

118

119 **Conclusion:**

120 The increasing availability of genome assemblies for a wide array of organisms is rapidly
121 expanding the scope and comparative power of modern genome biology (Hotaling and Kelley,
122 2018). With more than 4,600 contigs longer than 10 kb and ~70% of genes assembled in their
123 complete form, the *L. tumana* nuclear genome provides new opportunity for exploring genome
124 evolution within Plecoptera, a highly diverse, globally distributed insect order, or at higher levels
125 of taxonomic organization (e.g., across all insects). Specifically, the *L. tumana* genome could be
126 mined for genes for phylogenomic studies (e.g., Li et al., 2007, Boroweic et al., 2015) or more
127 targeted, comparative assessments of specific genes or gene families across many taxa to clarify
128 evolutionary shifts and/or copy number variation (e.g. Baalsrud et al., 2017).

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

140 Future efforts to refine both assemblies, including the incorporation of longer reads (e.g.,
141 generated using Pacific Biosciences sequencing technology, Utturkar et al., 2014) will yield
142 greater insight into the genome biology of *L. tumana*, stoneflies, and aquatic insects broadly.
143 Still, the resources provided here, and particularly the most complete stonefly nuclear genome

144 published thus far, represent an important step towards empowering modern stonefly research, a
145 globally relevant group of aquatic insects that has been largely overlooked in the genomic age.

146

147 **Acknowledgements:**

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

153

154 **Declaration of Interest:**

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

156 **References:**

- 157 Andrews, S. (2010), 'FastQC: a quality control tool for high throughput sequence data'.
158 <https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc>
- 159 Baalsrud, H. T., Tørresen, O. K., Solbakken, M. H., Salzburger, W., Hanel, R., Jakobsen, K. S.,
160 & Jentoft, S. (2017), 'De novo gene evolution of antifreeze glycoproteins in codfishes
161 revealed by whole genome sequence data', *Molecular Biology and Evolution*, 35, 593-606.
- 162 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
163 Nikolenko, S.I., Pham, S. and Prjibelski, A.D. (2012), 'SPAdes: a new genome assembly
164 algorithm and its applications to single-cell sequencing', *Journal of Computational Biology*,
165 19, 455-477.
- 166 Baumann, R.W. and Kondratieff, B.C. (2010), 'The stonefly genus *Lednia* in North America
167 (Plecoptera: Nemouridae)', *Illiesia*, 6, 315-327.
- 168 Baumann, R.W. and Call, R.G. (2012), '*Lednia tetonica*, a new species of stonefly from
169 Wyoming (Plecoptera: Nemouridae)', *Illiesia*, 8, 104-110.
- 170 Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J.,
171 Middendorf, M. and Stadler, P.F. (2013), 'MITOS: improved de novo metazoan
172 mitochondrial genome annotation', *Molecular Phylogenetics and Evolution*, 69, 313-319.
- 173 Béthoux, O., Cui, Y., Kondratieff, B., Stark, B. and Ren, D. (2011), 'At last, a Pennsylvanian
174 stem-stonefly (Plecoptera) discovered', *BMC Evolutionary Biology*, 11, 248.
- 175 Borowiec, M. L., Lee, E. K., Chiu, J. C. and Plachetzki, D. C. (2015), 'Extracting phylogenetic
176 signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister
177 to remaining Metazoa', *BMC Genomics*, 16, 987.
- 178 Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S.,
179 Chapman, J.A., Chapuis, G. and Chikhi, R. (2013), 'Assemblathon 2: evaluating de novo
180 methods of genome assembly in three vertebrate species', *GigaScience*, 2, 10.
- 181 Chen, Z.-T. and Du, Y.-Z. (2017), 'First mitochondrial genome from Nemouridae (Plecoptera)
182 reveals novel features of the elongated control region and phylogenetic implications',
183 *International Journal of Molecular Sciences*, 18, 996.
- 184 DeWalt R.E., Kondratieff B.C. and Sandberg J.B. 2015, 'Order Plecoptera', Thorp J., Rogers
185 D.C. (eds), *Ecology and General Biology: Freshwater Invertebrates*. – Academic Press,
186 Cambridge.
- 187 Dierckxsens, N., Mardulyn, P. and Smits, G. (2016), 'NOVOPlasty: de novo assembly of
188 organelle genomes from whole genome data', *Nucleic Acids Research*, 45, e18-e18.
- 189 Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D.F., Smith, C.R., Suen, G., Wurm,
190 Y. and Smith, C.D. (2012), 'The genomic impact of 100 million years of social evolution in
191 seven ant species', *Trends in Genetics*, 28, 14-21.

- 192 Giersch, J.J., Hotaling, S., Kovach, R.P., Jones, L.A. and Muhlfeld, C.C. (2017), 'Climate-
193 induced glacier and snow loss imperils alpine stream insects', *Global Change Biology*, 23,
194 2577-2589.
- 195 Honeybee Genome Sequencing Consortium. (2006), 'Insights into social insects from the
196 genome of the honeybee *Apis mellifera*', *Nature*, 443, 931.
- 197 Hotaling, S., Finn, D.S., Joseph Giersch, J., Weisrock, D.W. and Jacobsen, D. (2017), 'Climate
198 change and alpine stream biology: progress, challenges, and opportunities for the future',
199 *Biological Reviews*, 92, 2024-2045.
- 200 Hotaling, S., Giersch, J.J., Finn, D.S., Tronstad, L.M., Jordan, S., Serpa, L.E., Call, R.G.,
201 Muhlfeld, C.C., and Weisrock, D.W. (In press), 'Congruent population genetic structure
202 but differing depths of divergence for three alpine stoneflies with similar ecology and
203 geographic distributions', *Freshwater Biology*.
- 204 Hotaling, S. and Kelley, J.L. (In press), 'The rising tide of high-quality genomic resources',
205 *Molecular Ecology Resources*.
- 206 Hotaling, S., Muhlfeld, C.C., Giersch, J.J., Ali, O.A., Jordan, S., Miller, M.R., Luikart, G. and
207 Weisrock, D.W. (2018), 'Demographic modelling reveals a history of divergence with gene
208 flow for a glacially tied stonefly in a changing post-Pleistocene landscape', *Journal of*
209 *Biogeography*, 45, 304-317.
- 210 Jordan, S., Giersch, J.J., Muhlfeld, C.C., Hotaling, S., Fanning, L., Tappenbeck, T.H. and
211 Luikart, G. (2016), 'Loss of genetic diversity and increased subdivision in an endemic
212 alpine stonefly threatened by climate change', *PLoS One*, 11, e0157386.
- 213 Krueger, F. (2015), 'Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently
214 apply quality and adapter trimming to FastQ files'.
215 https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- 216 Li, C., Ortí, G., Zhang, G. and Lu, G. (2007), 'A practical approach to phylogenomics: the
217 phylogeny of ray-finned fish (Actinopterygii) as a case study', *BMC Evolutionary*
218 *Biology*, 7, 44.
- 219 Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J. and Zhou, X. (2018), 'The genome of an
220 underwater architect, the caddisfly *Stenopsyche tienmushanensis* Hwang (Insecta:
221 Trichoptera)', *GigaScience*, Doi: 10.1093/gigascience/giy143.
- 222 Macdonald, H.C., Cunha, L. and Bruford, M.W. (2016), 'Development of genomic resources for
223 four potential environmental bioindicator species: *Isoperla grammatica*, *Amphinemura*
224 *sulcicollis*, *Oniscus asellus* and *Baetis rhodani*', *bioRxiv*, 046227.
- 225 Macdonald, H.C., Ormerod, S.J. and Bruford, M.W. (2017), 'Enhancing capacity for freshwater
226 conservation at the genetic level: a demonstration using three stream
227 macroinvertebrates', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 27, 452-
228 461.

- 229 Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y. and Hackett, K.
230 (2014), 'The i5k Workspace@ NAL—enabling genomic data access, visualization and
231 curation of arthropod genomes', *Nucleic Acids Research*, 43, D714-D719.
- 232 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015),
233 'BUSCO: assessing genome assembly and annotation completeness with single-copy
234 orthologs', *Bioinformatics*, 31, 3210-3212.
- 235 Simpson, J.T. (2014), 'Exploring genome characteristics and sequence quality without a
236 reference', *Bioinformatics*, 30, 1228-1235.
- 237 Tronstad, L.M., Hotaling, S. and Bish, J.C. (2016), 'Longitudinal changes in stream invertebrate
238 assemblages of Grand Teton National Park, Wyoming', *Insect Conservation and Diversity*,
239 9, 320-331.
- 240 US Fish & Wildlife Service. (2016), 'Endangered and threatened wildlife and plants; 12-month
241 finding on a petition to list the western glacier stonefly as an endangered or threatened
242 species; proposed threatened species status for Meltwater Lednian Stonefly and Western
243 Glacier Stonefly', *Federal Register*, 81, 68379-68397.
- 244 Utturkar, S.M., Klingeman, D.M., Land, M.L., Schadt, C.W., Doktycz, M.J., Pelletier, D.A. and
245 Brown, S.D. (2014), 'Evaluation and validation of de novo and hybrid assembly techniques
246 to derive high-quality genome sequences', *Bioinformatics*, 30, 2709-2716.

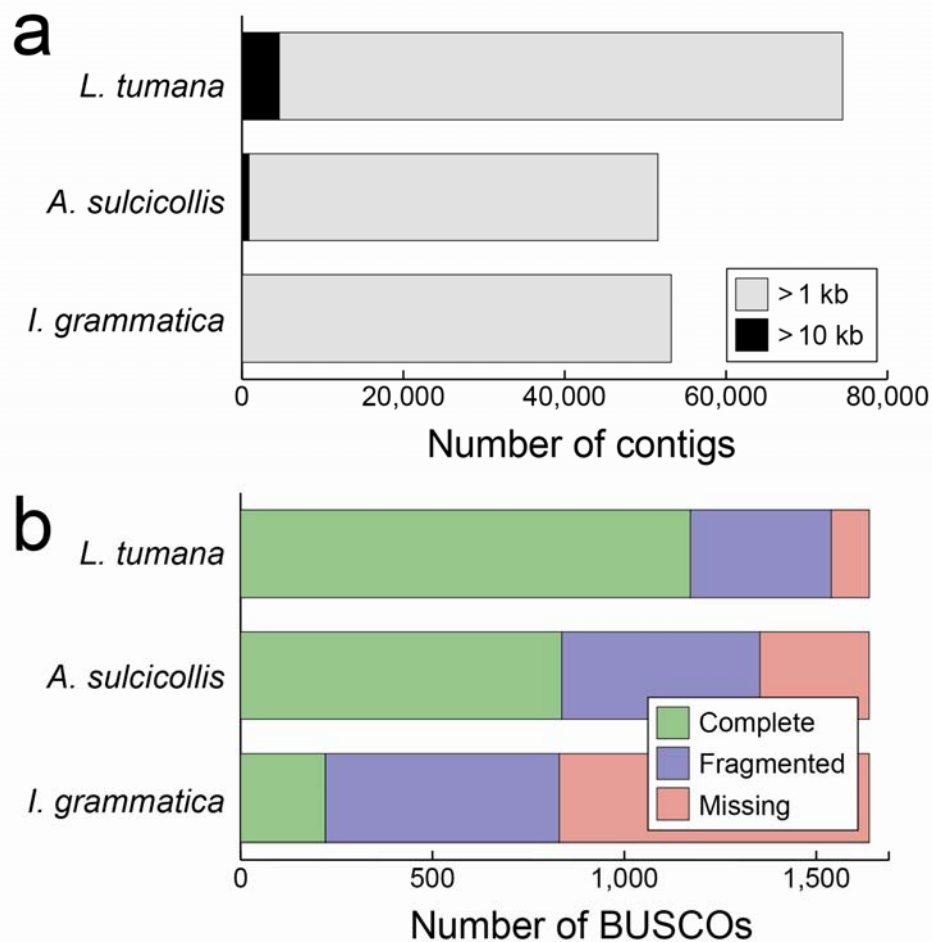
247 **Tables:**

248 Table 1. Assembly statistics for the nuclear genome of *Lednia tumana* (Plecoptera: Nemouridae)
249 and two other stonefly species, *Amphinemura sulcicollis* and *Isoperla grammatica* (Macdonald et
250 al., 2016; Macdonald et al., 2017). BUSCOs: Single-copy, orthologous genes known to be highly
251 conserved among insects. A total of 1,658 BUSCOs were searched. For *L. tumana*, the assembly
252 size was 304,502,267 bp when only contigs larger than 500 bp were included.

	<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	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%)

253

254 **Figures:**



255

256 Figure 1. Comparisons of the gene content and contiguity of the *Lednia tumana* nuclear genome
257 to the two other previously published stonefly genomes for *Amphinemura sulcicollis* and
258 *Isoperla grammatica* (Macdonald et al., 2016; Macdonald et al., 2017). (a) The number of
259 contigs > 1 kb and > 10 kb across the three stonefly genomes. For *I. grammatica*, the assembly
260 contains just four contigs > 10 kb. (b) Presence of highly conserved, single-copy orthologous
261 genes (BUSCOs) across the three stonefly genomes.

