- 1 Nuclear and mitochondrial genomic resources for the meltwater stonefly, *Lednia tumana*
- 2 Ricker, 1952 (Plecoptera: Nemouridae)
- 4 Scott Hotaling<sup>a,b</sup>\*, Joanna L. Kelley<sup>a</sup>, and David W. Weisrock<sup>b</sup>
- <sup>a</sup> School of Biological Sciences, Washington State University, Pullman, WA, USA; <sup>b</sup> Department
- 7 of Biology, University of Kentucky, Lexington, KY, USA
- 9 \* Scott Hotaling, School of Biological Sciences, Washington State University, Pullman, WA,
- 10 99164, USA; Email: scott.hotaling@wsu.edu; Phone: (828) 507-9950; ORCID: 0000-0002-
- 11 5965-0986

5

12 **Abstract:** 13 With more than 3,700 described species, stoneflies (Order Plecoptera) are an important 14 component of global aquatic biodiversity. The meltwater stonefly *Lednia tumana* (Ricker, 1952; 15 Family Nemouridae) is endemic to alpine streams of Glacier National Park and has been 16 petitioned for listing under the U.S. Endangered Species Act (ESA) due to climate change-17 induced loss of alpine glaciers and snowfields. Here, we present de novo assemblies of the 18 nuclear (~520 million base pairs [bp]) and mitochondrial (15,014-bp) genomes for *L. tumana*. 19 The L. tumana nuclear genome is the most complete stonefly genome reported to date, with 20 ~71% of genes present in complete form and more than 4,600 contigs longer than 10-kilobases 21 (kb). The L. tumana mitochondrial genome is the second for the family Nemouridae and the first 22 from North America. Together, both genomes represent important foundational resources, setting 23 the stage for future efforts to understand the evolution of L. tumana, stoneflies, and aquatic 24 insects worldwide. 25 26 **Keywords:** Plecoptera; Nemouridae; *Lednia*; genomics; nuclear genome; USA 27 28 **Introduction:** 29 Stoneflies are a diverse, globally distributed group of hemimetabolous insects that 30 diverged from their closest relatives (e.g., Orthoptera, Dermaptera, Zoraptera) at least 300 31 million years ago in the Carboniferous Period (Béthoux, Cui, Kondratieff, Stark, and Ren 2011). 32 With more than 3,700 described species, stoneflies account for a substantial portion of 33 freshwater biodiversity (DeWalt, Kondratieff, and Sandberg 2015). The meltwater stonefly, 34 Lednia tumana (Ricker, 1952; Plecoptera: Nemouridae), resides in alpine streams of Glacier 35 National Park (GNP), USA, where it is iconic of habitat loss due to climate change in the region 36 (Muhlfeld et al. 2011; Giersch, Hotaling, Kovach, Jones and Muhlfeld 2017). Lednia tumana is 37 one of four extant species in the genus *Lednia* which all exhibit alpine, cold-water distributions 38 in western North America (Baumann and Kondratieff 2010; Baumann and Call 2012). The 39 majority of L. tumana's habitat is supported by seasonal melting of permanent ice and snow, a 40 habitat type that is under considerable threat of near-term loss as the global cryosphere recedes 41 (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

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58 59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

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 kmerbased 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

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

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 sulcicollis (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. sulcicollis 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 beftools 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  $\geq 4.69$ -kilobases (kb; Figure 1a, Table 1). The assembled genome size is in

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

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 (Hotaling 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

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

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.

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

**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)', *Illesia*, 8, 104–110. 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. 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.

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

Cao, J. J., Wang, Y., Huang, Y. R., and Li, W. H. (2019), 'Mitochondrial genomes of the stoneflies Mesonemoura metafiligera and Mesonemoura tritaenia (Plecoptera, Nemouridae), with a phylogenetic analysis of Nemouroidea', ZooKeys, 835, 43. 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. DeWalt R.E., Kondratieff B.C., and Sandberg J.B. 2015, 'Order Plecoptera', Thorp J., Rogers D.C. (eds), Ecology and General Biology: Freshwater Invertebrates. – Academic Press, Cambridge. Dierckxsens, N., Mardulyn, P., and Smits, G. (2016), 'NOVOPlasty: de novo assembly of organelle genomes from whole genome data', Nucleic Acids Research, 45, e18. 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), 'Climateinduced 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., Giersch, J.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., Giersch, J.J., Finn, D.S., Tronstad, L.M., Jordan, S., Serpa, L.E., Call, R.G., Muhlfeld, C.C., and Weisrock, D.W. (2019), 'Congruent population genetic structure but differing depths of divergence for three alpine stoneflies with similar ecology and geographic distributions', Freshwater Biology, 64, 335–347.

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

Hotaling, S., and Kelley, J.L. (2019), 'The rising tide of high-quality genomic resources', *Molecular Ecology Resources*, 19, 567–569. 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'. https://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/ Li, C., Ortí, G., Zhang, G., and Lu, G. (2007), 'A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study', BMC Evolutionary Biology, 7, 44. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin. R. (2009) 'The sequence alignment/map format and SAMtools', Bioinformatics, 25, 2078–2079. Li, H. (2013), 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', arXiv, 1303, 3997. Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., and Zhou, X. (2018), 'The genome of an underwater architect, the caddisfly Stenopsyche tienmushanensis Hwang (Insecta: Trichoptera)', GigaScience, Doi: 10.1093/gigascience/giy143. 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.

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

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. Muhlfeld, C.C., Giersch, J.J., Hauer F.R., Pederson, GT., Luikart, G., Peterson, D.P., Downs, C.C., and Fagre, D.P. (2011), 'Climate change links fate of glaciers and an endemic alpine invertebrate', Climatic Change, 106, 337–345. Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., and Hackett, K. (2014), 'The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes', *Nucleic Acids Research*, 43, D714–D719. Ricker, W. E. (1952), 'Systematic studies in Plecoptera', Indiana University, Bloomington. 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. Utturkar, S.M., Klingeman, D.M., Land, M.L., Schadt, C.W., Doktycz, M.J., Pelletier, D.A., and Brown, S.D. (2014), 'Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences', *Bioinformatics*, 30, 2709–2716.

## **Tables:**

270

271

- 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
- 273 Isoperla grammatica (Poda, 1761; Macdonald et al. 2016; Macdonald et al. 2017). BUSCOs:
- 274 Single-copy, orthologous genes known to be highly conserved among insects. A 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	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%)

## Figures:

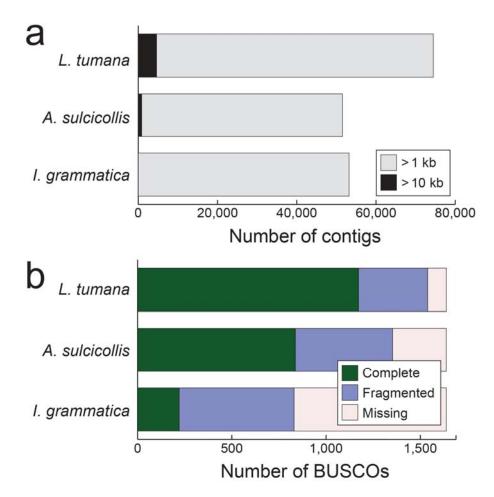


Figure 1. Comparisons of the gene content and contiguity of the *Lednia tumana* (Ricker, 1952; Plecoptera: Nemouridae) nuclear genome to the two other previously published stonefly genomes for *Amphinemura sulcicollis* (Stephens, 1836) and *Isoperla grammatica* (Poda, 1761; Macdonald et al. 2016; Macdonald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

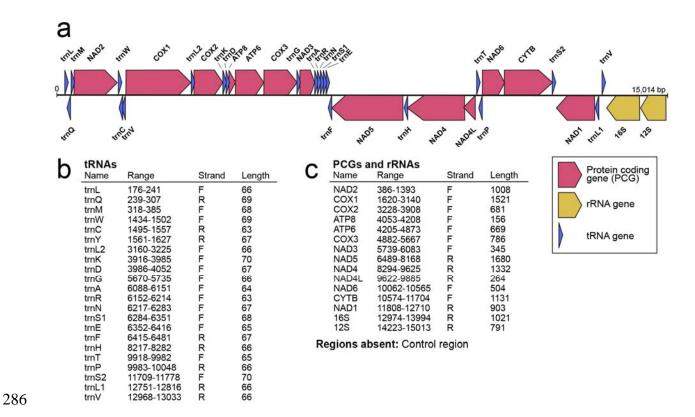


Figure 2. (a) The mitogenome of *Lednia tumana* (Ricker, 1952; Plecoptera: Nemouridae). (b) Locations of tRNAs. (c) Locations of protein coding (PCGs) and rRNA genes.