Annotated Draft Genomes of Two Caddisfly Species *Plectrocnemia conspersa* CURTIS and *Hydropsyche tenuis* NAVAS (Insecta: Trichoptera)

Jacqueline Heckenhauer[1,2,*], Paul B. Frandsen[1,3,4], Deepak K. Gupta[1], Juraj Paule[5], Stefan Prost[1,6], Tilman Schell[1], Julio V. Schneider[2], Russell J. Stewart[7], and Steffen U. Pauls[1,2,8]

[1]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany

[2]Department of Terrestrial Zoology, Entomology III, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany

[3]Department of Plant & Wildlife Sciences, Brigham Young University, Provo, UT

[4]Data Science Lab, Smithsonian Institution, Washington, DC

[5]Department of Botany and Molecular Evolution, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany

[6]South African National Biodiversity Institute, National Zoological Gardens of South Africa, Pretoria, South Africa

[7]Department of Biomedical Engineering, University of Utah, Salt Lake City, UT

[8]Institute for Insect Biotechnology, Justus-Liebig-University, Gießen, Germany

*Corresponding author: E-mail: jacqueline.heckenhauer@senckenberg.de.

Accepted: November 25, 2019

Data deposition: This project has been deposited at GenBank under BioProject PRJNA558902 (BioSamples SAMN12501725, SAMN12501726, SRA: PRJNA558902). This project has been deposited at DDBJ/ENA/GenBank under the accession VTKK000000000 and VTON00000000. Specimens used are deposited in the Senckenberg Research Institute and Natural History Museum Frankfurt.

Abstract

Members of the speciose insect order Trichoptera (caddisflies) provide important ecosystem services, for example, nutrient cycling through breaking down of organic matter. They are also of industrial interest due to their larval silk secretions. These form the basis for their diverse case-making behavior that allows them to exploit a wide range of ecological niches. Only five genomes of this order have been published thus far, with variable qualities regarding contiguity and completeness. A low-cost sequencing strategy, that is, using a single Oxford Nanopore flow cell per individual along with Illumina sequence reads was successfully used to generate high-quality genomes of two Trichoptera species, *Plectrocnemia conspersa* and *Hydropsyche tenuis*. Of the de novo assembly methods compared, assembly of low coverage Nanopore reads (~18×) and subsequent polishing with long reads followed by Illumina short reads (~80–170× coverage) yielded the highest genome quality both in terms of contiguity and BUSCO completeness. The presented genomes are the shortest to date and extend our knowledge of genome size across caddisfly families. The genomic region that encodes for light (L)-chain fibroin, a protein component of larval caddisfly silk was identified and compared with existing L-fibroin gene clusters. The new genomic resources presented in this paper are among the highest quality Trichoptera genomes and will increase the knowledge of this important insect order by serving as the basis for phylogenomic and comparative genomic studies.

Key words: : de novo genome assembly, fibroin, insect genomics, silk genes, genome size, Trichoptera.

Introduction

With >16,000 extant species (Morse: Trichoptera World Checklist; available online: http://entweb.sites.clemson.edu/database/trichopt/; last accessed December 4, 2019) the holometabolous insect order Trichoptera (caddisflies) is the seventh most speciose order of all insects (Adler and Foottit 2017). Members of Trichoptera provide important ecosystem services (Macadam and Stockan 2015; Morse et al. 2019). They stabilize gravel bed sediments, play an important role in food webs, and are important biological indicator organisms used for assessing and monitoring water quality (Morse 2017). Similar to lepidopteran caterpillars, caddisfly larvae

Table 1
Comparison of Genome Assemblies among the Seven Published Caddisfly Genomes

Species	Accession	Suborder	Sequencing Platform (Coverage[a])	Assembly Length (bp)	Scaffold N50 (kb)	BUSCOs Present[b] (%)	Estimated Haploid Genome Size (1C)
Hydropsyche tenuis (this study)	VTFK00000000	Annulipalpia	Nanopore+Illumina (16.5× + 167.6×)[a]	229,663,394	2190	98.3	222 Mb[c] (FCM: 258 Mb[d])
Plectrocnemia conspersa (this study)	VTON00000000	Annulipalpia	Nanopore+Illumina (17.1× + 82.9×)[b]	396,695,105	869	98.6	315 Mb[c] (FCM: 455 Mb)
Stenopsyche tienmushanensis (Luo et al. 2018)	v1	Annulipalpia	PacBio+Illumina (153× + 150×)	451,494,475	1297	98.7	407[c] to 453 Mb[e]
Limnephilus lunatus (i5k Consortium)	Llun_2.0	Integripalpia	Illumina (80.1×)	1369,180,260	69.1	94.3	n.a.
Glossosoma conforme (Weigand et al. 2018)	ASM334726v1	Annulipalpia	Illumina (53×)	604,293,666	16.7	96.8	1.52 Gb[e]
Sericostoma sp. (Weigand et al. 2017)	ASM300347v1	Integripalpia	Illumina (43×)	1015,727,762	3.1	76.2	616 Mb[e]
Glyphotaelius pellucidus (Ferguson et al. 2014)	n.a.	Integripalpia	Illumina (8.12×)	757,289,448	1.47	62.2	n.a.

[a]Coverage of data used for genome assembly.
[b]$N_{Insecta}$ = 1,658; present = complete + fragmented.
[c]Based on GenomeScope (Vurture et al. 2017).
[d]Based on closely related species.
[e]Based on 17-mer analysis.
FCM, flow cytometry; n.a., not available.

produce silk in specially modified labial glands. Their silk secretion forms the basis for their diverse case-making behavior which, in turn, may allow caddisfly larvae to exploit a wide range of ecological niches. Caddisfly silk is an important target for interdisciplinary research because it is a tough, energy dissipating fiber that adheres to diverse substrates completely submerged in water (Stewart and Wang 2010; Ashton et al. 2013, 2016), making it an attractive model for tough and adhesive synthetics, including materials with potential applications in medicine (e.g., Brooks 2015; Lane et al. 2015). We report the first draft genomes of two caddisfly species, each generated with a single Oxford Nanopore flow cell along with Illumina sequencing data. This low-cost sequencing strategy was tested on two Trichoptera species from families for which genomic resources were previously not available: *Plectrocnemia conspersa* CURTIS (Polycentropodidae) and *Hydropsyche tenuis* NAVAS (Hydropsychidae) of the suborder Annulipalpia. The most closely related species for which a high-quality assembled genome is available is *Stenopsyche tienmushanensis* Hwang. It shared a common ancestor with *Hydropsyche* ~177 Ma and with *Plectrocnemia* ~166 Ma, representing large gaps in evolutionary time among groups (Malm et al. 2013; Luo et al. 2018). Despite their ecological importance and potential technological applications, caddisflies are underrepresented among publicly available genomes. Only five genomes have been published thus far, most with low-quality continuity and completeness (table 1). The new genomic resources presented in this paper will serve as a foundation for future genomic and applied research on caddisflies.

Materials and Methods

Genomic DNA was extracted from larvae (*H. tenuis*: 49°3′29.31″N, 13°17′44.02″E, *P. conspera*: 50°15′28.58″N, 9°25′1.23″E). Illumina paired-end sequencing was performed on a HiSeq 2000 sequencer. The Oxford Nanopore library was sequenced on a single flow cell using the MinION portable DNA sequencer. Further details and sequence read processing are described in supplementary note 1, Supplementary Material online.

Genome size estimates were conducted using flow cytometry (FCM) according to Otto (1990) using *Lycopersicon esculentum* cv. Stupické polnítyčkové rané (2C = 1.96 pg; Doležel et al. 1992) as internal standard and from the short-read sequence data with GenomeScope 2.0 (Ranallo-Benavidez et al. 2019) and backmap.pl v0.1 (Schell et al. 2017; supplementary note 2, Supplementary Material online). GC content was estimated based on comparison of nuclei fluorescence sample/standard ratios stained with the DNA intercalating propidium iodide and AT-specific DAPI using the protocols and the GC content calculation tool by Šmarda et al. (2008). We confirm with our genomic data that this method, which is usually applied in plant research can be successfully applied on animals.

Different de novo genome assembly methods (short-, long-read, and hybrid assembly approaches) were tested with the *P. conspersa* data set (supplementary note 3, Supplementary Material online). The pipeline, which yielded the best genome for *P. conspersa*, was used to assemble the *H. tenuis* genome. Quality of assemblies was evaluated based on continuity

[QUAST v5.0.2 (Gurevich et al. 2013)] and BUSCO completeness (Simão et al. 2015; Waterhouse et al. 2018). Genome assemblies were screened for potential contaminations with BlobTools v1.0 (supplementary note 4, Supplementary Material online).

Genome annotation is described in supplementary note 5, Supplementary Material online. In short, repeatModeler v1.0.11 (Smit and Hubley 2008–2015) was used to identify de novo repeat families. Masking and annotating repetitive elements were conducted using RepeatMasker v4.0.7 (Smit et al. 2013–2015). BLAST-like Alignment Tool v3.6 (BLAT, Kent 2002) was used to align the transcriptome of *Plectrocnemia* sp. (INSqiqTBXRAAPEI-18) to the *P. conspersa* and the transcriptome of *Arctopsyche palpata* (INSeqtTAARAAPEI-18) to the *H. tenuis* draft genome sequence. Transcriptomes were provided by 1KITE (http://www.1kite.org/; last accessed November 11, 2019). Protein-coding genes were annotated with AUGUSTUS v3.3 (Stanke et al. 2008). All protein sequences were assigned putative names by BlastP Protein–Protein BLAST 2.2.30+ searches (Camacho et al. 2009) and were functionally annotated using InterProScan v5 (Zdobnov and Apweiler 2001; Quevillon et al. 2005) and command line Blast2Go v1.3.3 (Götz et al. 2008).

Sequences of L-fibroin were extracted from the functional annotations and aligned with publicly available L-fibroin genes in Geneious v8.0.5. Distance and maximum likelihood trees were reconstructed as described in supplementary note 7, Supplementary Material online.

Results and Discussion

Genome Assembly

These two Trichoptera genomes were sequenced using a low-cost sequencing approach. Even so, the genome assembly statistics are comparable to the previously published *S. tienmushanensis* genome, which was sequenced at a much higher depth of coverage, and they exhibited improved continuity (based on N50) when compared with all other Trichoptera genome assemblies (table 1). Of the different de novo assembly methods compared for *P. conspersa* (supplementary table 1, Supplementary Material online), sequence assembly of Nanopore reads (~17.5× and 16.5× coverage for *P. conspersa* and *H. tenuis*, respectively) using wtdbg2 (Ruan and Li 2019), followed by polishing with long Nanopore reads [Racon (Vaser et al. 2017)], and Illumina short reads [Pilon (Walker 2014); 82.9× and 167.6× coverage for *P. conspersa* and *H. tenuis*, respectively] performed best with respect to both genome quality and computing time. The draft *P. conspersa* assembly comprised 1,614 scaffolds (1,613 ≥ 1,000 bp) with a cumulative length of 396,695,105 bp. It exhibited a contig N50 of 868,980, a contig L50 of 141 and a GC content of 30.1%. The genome contained 97.8% (complete: 93.5%, fragmented: 4.3%) of an Endopterygota core gene collection indicating an almost complete coverage of known single copy orthologs in the coding fraction. Remapping the preprocessed Illumina reads showed that 99.5% could be placed within the assembly (supplementary fig. 1, Supplementary Material online). The wtdbg2-Racon-Pilon pipeline also yielded excellent results for the *H. tenuis* data, which had even lower coverage Nanopore data (~16.5× coverage). After contaminated contigs were filtered out using BlobTools (supplementary fig. 2*B*, Supplementary Material online), the assembly consisted of 403 scaffolds of ≥1,000 bp with a cumulative length of 229,663,394 bp. The assembly had a contig N50 of 2,190,134 bp, the largest known contig N50 for a caddisfly, a contig L50 of 33 and a GC content of 33%. In total, 97.7% (complete: 94.2%, fragmented: 3.5%) of the Endopterygota core gene collection was detected by BUSCO. Mapping the preprocessed Illumina reads showed a mapping rate of 99.3% (supplementary fig. 1, Supplementary Material online). FCM-based estimation revealed a GC content of 30.39% for *Hydropsyche siltalai* and 29.77% for *P. conspersa*, which shows that GC rich regions are not underrepresented in the genome assemblies and thus adds toward quality of the assemblies.

Genome Size Estimations

Regarding *P. conspersa* (supplementary note 2, Supplementary Material online; table 1), the mean haploid genome size estimate resulting from the mapping-based genome size estimation with backmap.pl was ~365 Mb, FCM yielded an estimate of 455.20 ± 8.39 Mb, and GenomeScope predicted a size of ~316 Mb. A high level of repeats, that fail to assemble correctly may account for this discrepancy. Differences between k-mer-based methods and FCM have been observed before and can be explained by the presence of repetitive elements, which can affect k-mer estimates (Austin et al. 2017; Sánchez-Herrero et al. 2019). Generally, our results indicate that researchers should not rely on a single method for estimating genome size. For *H. tenuis*, the genome size estimates based on sequencing data ranged from ~222 to ~229 Mb (supplementary note 2, Supplementary Material online). This was in line with FCM-based estimates of closely related species of *Hydropsyche* (*H. siltalai*: 257.72 ± 7.24 Mb, *H. pellucidula* 283.94 ± 4.16 Mb, *H. saxonica* 241.65 ± 4.41 Mb), and with the assembly length. To our knowledge, we present the first genome size estimates for Trichoptera based on FCM. Varying genome sizes (based on sequence data) have been observed between suborders, that is, between retreat making Annulipalpia [*Hydropsyche*: 241.65–283.94 Mb; *Stenopsyche*: 407–453 Mb (Luo et al. 2018), *Plectrocnemia*: ~ 455 Mb, *Glossosoma conforme*: 1.5 Gb (Weigand et al. 2018)] and case-building Integripalpia [*Sericostoma* sp.: 616 Mb (Weigand et al. 2017), assembly length of *Limnephillus lunatus*: ~1.4 Gb

(i5k Consortium), table 1], but also among families within suborders (table 1). Further studies are needed to investigate genome size evolution in Trichoptera.

Genome Annotation and Functional Annotation of Protein-Coding Genes

A total of 34.36% of the *P. conspersa* genome assembly was masked as repeats. A 32.39% of the annotated repeats were interspersed repeats. Almost half of the interspersed repeats, 17.58%, could not be classified by comparison with known repeat databases, and therefore may be specific for Trichoptera. Only 19.12% of the *H. tenuis* genome assembly was masked as repeats. A 17.16% of the repeats were classified as interspersed repeats. A large portion, 10.89%, of the interspersed repeats remained unclassified. In comparison, repeat content in the previously published *S. tienmushanensis* is 36.76 (10.56% unclassified; Luo et al. 2018). Details on the repeat classes are given in supplementary note 6, Supplementary Material online. The annotation of the genomes resulted in the prediction of 19,104 proteins and 18,586 genes and 15,844 proteins and 15,873 genes for *P. conspersa* and *H. tenuis*, respectively. In total, BUSCO detected 95.6% of the Endopterygota core gene collection in the predicted proteins of *P. conspersa* (complete: 88.1%, fragmented: 7.5%) and *H. tenuis* (complete: 90.5%, fragmented: 5.1%), respectively. In comparison, 15,658 proteins and 14,672 genes were annotated in *S. tienmushanensis* (Luo et al. 2018). More details can be found in supplementary table 2, Supplementary Material online. Of the annotated proteins 70% and 62% showed significant sequence similarity to entries in the NCBI nr database for *P. conspersa* and *H. tenuis*, respectively. When searched against InterPro databases, 14,476 (*P. conspersa*: 75.7%) and 12,949 (*H. tenuis*: 81.7%) annotated proteins include motifs/domains were identified by InterProScan of which 7,192 (*P. conspersa*) and 6,889 (*H. tenuis*) genes were assigned to Gene Ontology IDs with a corresponding InterPro entry (supplementary figs. 3 and 4, Supplementary Material online). The major biological process found in all three caddisfly genomes was a cellular process (*P. conspersa:* 1,639 genes, *H. tenuis*: 1,460 genes, *S. tienmushanensis:* 1,523 genes). Although binding was the largest subcategory in molecular function in *H. tenuis* (1,338 genes) and *S. tienmushanensis* 1,506 genes), catalytic activity was the largest subcategory in *P. conspersa* (1,415 genes). Regarding the cellular component category, most genes were assigned to the cell subcategory in *H. tenuis* (1,017 genes) and *S. tienmushanensis* (1,008 genes) and to the membrane subcategory in *P. conspersa* (1,123 genes). InterProScan repeat distribution revealed that leucine-rich repeats ranked highest among all repeats for *P. conspersa* and *H. tenuis*, whereas ankyrin-rich repeats ranked highest in *S. tienmushanensis*. A detailed distribution of InterProScan repeats can be found in supplementary figure 5, Supplementary Material online.

Identification of the L-fibroin Gene and Evolutionary Implications

L-fibroin, a protein component of larval caddisfly silk was successfully assembled, annotated, and compared with previously published L-fibroin sequences. Translation of the L-fibroin sequence yielded a protein of 280 and 276 amino acid residues for *Hydropsyche tenuis* and *Plectrocnemia conspersa*, respectively. Amino acid sequences of the L-fibroin proteins exhibit 56.5% pairwise identity and 26.7% identical sites among the eight caddisfly species (fig. 1*C*). The L-fibroins of Trichoptera exhibit a similar organization to the L-fibroin of Lepidoptera (19.2% pairwise identity; 18.7% identical sites; fig. 1: *Bombyx mori*) suggesting that the ancestral L-fibroin gene evolved prior to the separation of the two orders and was retained in both (Yonemura et al. 2009). Moreover, distribution of residues with characteristic properties is similar in both orders, for example, that of Cystein (polar, hydrophobic, and neutral) which has proved to play an important role for L-fibroin function in *B. mori* (Tanaka et al. 1999). Within Trichoptera, the conservation of regions with alternating groups of hydrophobic and hydrophilic residues can be observed. The two reconstructed trees of the L-fibroin exon sequences agree closely with previously reported phylogenetic trees (Kjer et al. 2016; Frandsen et al., in preparation). The distance tree is separated into two large clades that correspond to the suborders Annulipalpia (fig. 1I) and Integripalpia (fig. 1II). Within Annulipalpia, *Hydropsyche* (Hydropsychidae) and *Stenopsyche* (Stenopsychidae) form a clade with *Plectrocnemia* (Polycentropodidaea) as sister. The maximum likelihood tree revealed a similar topology (fig. 1*B*) and high bootstrap support (>90) for most of the nodes. However, the position of *Plectrocnemia* remains unclear. Although Hydropsychidae are typically characterized by a relative flat filter net built adjacent to their actual retreat (Clarke 1883; Schuhmacher 1970), larvae of *Plectrocnemia* are "sit-and-wait" predators with silken retreats attached to aquatic plants with many irregularly radiating threads forming capture nets adapted for predation of organisms in the benthos which become entangled in its nets (Wesenberg-Lund 1911; Townsend and Hildrew 1976, Lillpopp et al. 1998). Retreats/capture nets of Stenopsychidae are loosely woven (Morse et al. 2019). The capture-net portion consists of three different parts which serve as a cover, feeding mesh, and collecting seston (Tanida 2000). Current phylogenetic evidence suggests that retreat-making behavior followed evolutionary transitions from 1) organisms with a silken retreat that housed the organism with a separated filter net outside of the larval retreat to 2) larvae with retreats that also served as filter nets to 3) more advanced retreats (such as trumpet nets) that

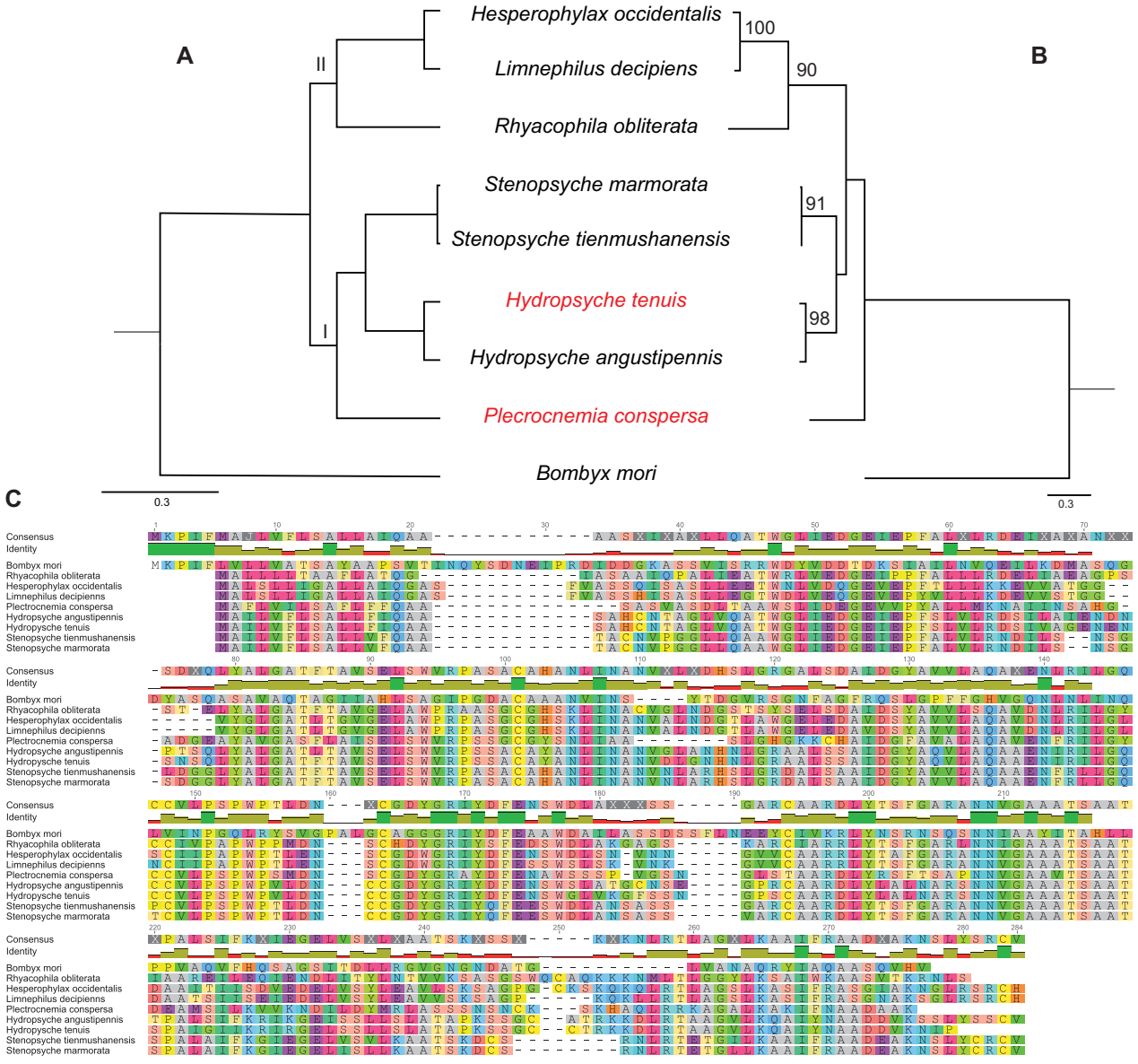

FIG. 1.—(*A*) Jukes–Cantor distance tree. I: Annulipalpia; II: Integripalpia. (*B*) Maximum likelihood tree. Bootstrap values ≥90% are given for each node. L-fibroins of the two sequenced genomes are indicated in red. Genbank accessions: *Bombyx mori:* NM_001044023.1 (Suetsugu et al. 2013), *Rhyacophila obliterata*: AB354690.1 (Yonemura et al. 2009), *Limnephilus decipienns*: AB214510.1 (Yonemura et al. 2006), *Hesperophylax occidentalis*: KM384738.1 (Wang et al. 2014), *Hydropsyche tenuis*: this study, *Hydropsyche angustipennis*: AB214508.1 (Yonemura et al. 2006), *Plectrocnemia conspersa*: this study, *Stenopsyche marmorata*: LC057252.1 (Bai et al. 2015), and *Stenopsyche tienmushanensis*: Luo et al. (2018). (*C*) Aligned amino acid residues of L-fibroin. Each color in the alignment represents a different amino acid. Mean pairwise identity over all pairs in the column: green, 100% identity; greeny-brown, at least 30% and under 100% identity; red, below 30% identity.

enable behaviors such as collecting fine particulate matter, gardening, scraping perphyton, and predation (Morse et al. 2019).

The new genomic resources presented in this paper, which includes the full assembly of the L-fibroin are a further step to disentangling strands of silk, that is, to better understand the evolution of silk biosynthesis, silk production, and associated behavior in caddisflies. Further studies will include this data to assess the phylogeny of caddisflies and evolution of silk usage, and to investigating structural differences in genes involved in biosynthesis of silk and silk usage regulation.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work is a result of the LOEWE-Centre for Translational Biodiversity Genomics funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK).

Authors' Contributions

J.H., S.U.P., and P.B.F. designed the study. S.U.P. collected the specimens. J.S. performed the laboratory work for genome sequencing. J.P. estimated genome size and GC content using flow cytometry. D.K.G., T.S., J.H., and P.B.F. designed and performed the bioinformatic analyses. J.H. and P.B.F. performed analyses of the silk genes. R.S. advised on annotation of L-fibroins. S.P. advised on de novo genome assembly. J.H. and S.U.P. drafted the manuscript. All authors edited and approved the final version of the manuscript.

Literature Cited

Adler PH, Foottit RG. 2017. Introduction. In: Footitt RG, Adler PH, editors. Insect biodiversity: science and society, 2nd ed. Hoboken (NJ): John Wiley Sons. p. 1–7.

Ashton NN, Roe DR, Weiss RB, Cheatham TE, Stewart RJ. 2013. Self-tensioning aquatic caddisfly silk: Ca^{2+}-dependent structure, strength, and load cycle hysteresis. Biomacromolecules 14(10):3668–3681.

Ashton NN, Pan H, Stewart RJ. 2016. Connecting caddisworm silk structure and mechanical properties: combined infrared spectroscopy and mechanical analysis. Open Biol. 6(6):160067.

Austin CM, et al. 2017. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. GigaScience 6(8):1–6.

Bai X, et al. 2015. Molecular cloning, gene expression analysis, and recombinant protein expression of novel silk proteins from larvae of a retreat-maker caddisfly, *Stenopsyche marmorata*. Biochem Biophys Res Commun. 464(3):814–819.

Brooks AE. 2015. The potential of silk and silk-like proteins as natural mucoadhesive biopolymers for controlled drug delivery. Front Chem. 3:65.

Camacho C, et al. 2009. BLAST plus: architecture and applications. BMC Bioinformatics 10(1):421.

Clarke C. 1883. Description of two interesting houses made by native caddis-fly. Proc Boston Soc Nat Hist. 22:67–71.

Doležel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol Plant. 85:625–631.

Ferguson L, et al. 2014. Ancient expansion of the *Hox* cluster in Lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation. PLoS Genet. 10(10):e1004698.

Frandsen PB, et al. Resolving the deep splits within the caddisfly (Insecta: Trichoptera) tree of life using transcriptomes and targeted enrichment. Unpublished data, manuscript in preparation.

Götz S, et al. 2008. High-throughputfunctional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10):3420–3435.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.

i5K Consortium. 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. J Hered. 104:595–600.

Kent WJ. 2002. BLAT-the BLAST-like alignment tool. Genome Res. 12(4):656–664.

Kjer KM, et al. 2016. Progress on the phylogeny of caddisflies. In: Vshivkova TS, Morse JC, editors. Proceedings of the 14th international symposium on Trichoptera, Vladivostok, Russia, 2–7 July 2012; Zoosymposia 2016, Vol. 10, p. 248–256.

Lane DD, Kaur S, Weerasakare GM, Stewart RJ. 2015. Toughened hydrogels inspired by aquatic caddisworm silk. SoftMatter 11:6981–6990.

Lillpopp S, Bohle HW, Fiedler A. 1998. Beiträge zur Biologie von *Plectrocnemia conspersa* (CURTIS) (Polycentropodidae, Trichoptera). Lauterbornia 34:227–238.

Luo S, Tang M, Frandsen PB, Stewart RJ, Zhou X. 2018. The genome of an underwater architect, the caddisfly *Stenopsyche tienmushanensis* Hwang (Insecta: trichoptera). Gigascience 7(12):143. giy

Macadam CR, Stockan JA. 2015. More than just fish food: ecosystem services provided by freshwater insects. Ecol Entomol. 40:113–123.

Malm T, Johanson KA, Wahlberg N. 2013. The evolutionary history of Trichoptera (Insecta): a case of successful adaptation to life in freshwater. Syst Entomol. 38(3):459–473.

Morse JC. Trichoptera World Checklist. Available from: http://entweb.sites.clemson.edu/database/trichopt/; last accessed December 4, 2019.

Morse JC. 2017. Biodiversity of aquatic insects. In: Footitt RG, Adler PH, editors. Insect biodiversity: science and society, 2nd ed. Hoboken (NJ): John Wiley Sons. p. 205–227.

Morse JC, Frandsen PB, Graf W, Thomas JA. 2019. Diversity and ecosystem services of Trichoptera. Insects 10(5):125.

Otto F. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. Methods Cell Biol. 33:105–110.

Quevillon E, et al. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33 (2): W116–W120.

Ranallo-Benavidez TH, Jaron KS, Schatz MC. 2019. GenomeScope 2.0 and Smudgeplots: reference-free profiling of polyploid genomes. Preprint at bioRxiv; doi.org/10.1101/747568

Ruan J, Li H. 2019. Fast and accurate long-read assembly with wtdbg2. Preprint at bioRxiv 530972.

Sánchez-Herrero JF, et al. 2019. The draft genome sequence of the spider *Dysdera silvatica* (Araneae, Dysderidae): a valuable resource for functional and evolutionary genomic studies in chelicerates. GigaScience 8(8):1–9.

Schell T, et al. 2017. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biol Evol. 9(3):585–592.

Schuhmacher H. 1970. Untersuchungen zur Taxonomie, Biologie und Ökologie einiger Köcherfliegenarten der Gattung *Hydropsyche* POT. (Insecta, Trichoptera). Int Revue Ges Hydrobiol Hydrogr. 55(4):511–557.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212.

Šmarda P, Bureš P, Horová L, Foggi B, Rossi G. 2008. Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. Ann Bot. 101(3):421–433.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5):637–644.

Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. Available from: http://www.repeatmasker.org; last accessed December 4, 2019.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org; last accessed December 4, 2019.

Stewart RJ, Wang CS. 2010. Adaptation of caddisfly larval silks to aquatic habitats by phosphorylation of H-fibroin serines. Biomacromolecules 11(4):969–974.

Suetsugu Y, et al. 2013. Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect. Bombyx mori. G3 (Bethesda) 3:1481–1492.

Tanaka K, Inoue S, Mizuno S. 1999. Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori*. Insect Biochem Mol Biol. 29(3):269–276.

Tanida K. 2000. *Stenopsyche* (Trichoptera: Stenopsychidae): ecology and biology of a prominent Asian genus. In: Mey W, editor. Proceedings of the 10th International Symposium on Trichoptera, Potsdam, Germany, 30 July–5 August 2000. Goecke & Evers: Keltern, Germany, Deutsches Entomologisches Institut: Müncheberg, Germany, 2002.

Townsend CR, Hildrew AG. 1976. Field experiments on the drifting, colonization and continuous redistribution of stream benthos. J Anim Ecol. 45(3):759–772.

Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746.

Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33 (14): 2202–2204.

Walker B, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

Wang Y, et al. 2010. Characterization of unique heavy chain fibroin filaments spun underwater by the caddisfly *Stenopsyche marmorata* (Trichoptera; Stenopsychidae). Mol Biol Rep. 37(6):2885–2892.

Wang CS, Ashton NN, Weiss RB, Stewart RJ. 2014. Peroxinectin catalyzed dityrosine crosslinking in the adhesive underwater silk of a casemaker caddisfly larvae, *Hysperophylax occidentalis*. Insect Biochem Mol Biol. 54:69–79.

Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35(3):543–548.

Weigand H, et al. 2018. Fishing in troubled waters: revealing genomic signatures of local adaptation in response to freshwater pollutants in two macroinvertebrates. Sci Total Environ. 633:875–891.

Weigand H, et al. 2017. Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. Mol Ecol. 26(20):5705–5715.

Wesenberg-Lund C. 1911. Biologische Studien über Netzspinnende Trichopterenlarven; plates I-IV. Mitt. Biol. Süsswasserlab. Fred. Lyngby, Denmark 11:1–54.

Yonemura N, Sehnal F, Mita K, Tamura T. 2006. Protein composition of silk filaments spun under water by caddisfly larvae. Biomacromolecules 7(12):3370–3378.

Yonemura N, et al. 2009. Conservation of silk genes in Trichoptera and Lepidoptera. J Mol Evol. 68(6):641–653.

Zdobnov EM, Apweiler R. 2001. InterProScan-an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9):847–848.

Associate editor: Esther Betran