Draft Genome Assemblies and Annotations of *Agrypnia vestita* Walker, and *Hesperophylax magnus* Banks Reveal Substantial Repetitive Element Expansion in Tube Case-Making Caddisflies (Insecta: Trichoptera)

Lindsey K. Olsen[1,*†], Jacqueline Heckenhauer[2,3,†], John S. Sproul[4,†], Rebecca B. Dikow[5], Vanessa L. Gonzalez[6], Matthew P. Kweskin[7], Adam M. Taylor[1], Seth B. Wilson[1], Russell J. Stewart[8], Xin Zhou[9], Ralph Holzenthal[10], Steffen U. Pauls[2,3,11], and Paul B. Frandsen [1,2,5,*†]

[1]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah, USA

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

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

[4]Department of Biology, University of Rochester, New York, USA

[5]Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, District of Columbia, USA

[6]Global Genome Initiative, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia, USA

[7]Laboratories of Analytical Biology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia, USA

[8]Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah, USA

[9]Department of Entomology, China Agricultural University, Beijing, China

[10]Department of Entomology, University of Minnesota, St. Paul, Minnesota, USA

[11]Institute of Insect Biotechnology, Justus-Liebig University, Gießen, Germany

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: lkolsen9615@gmail.com; paul_frandsen@byu.edu.

Accepted: 16 January 2021

Abstract

Trichoptera (caddisflies) play an essential role in freshwater ecosystems; for instance, larvae process organic material from the water and are food for a variety of predators. Knowledge on the genomic diversity of caddisflies can facilitate comparative and phylogenetic studies thereby allowing scientists to better understand the evolutionary history of caddisflies. Although Trichoptera are the most diverse aquatic insect order, they remain poorly represented in terms of genomic resources. To date, all long-read based genomes have been sequenced from individuals in the retreat-making suborder, Annulipalpia, leaving ~275 Ma of evolution without high-quality genomic resources. Here, we report the first long-read based de novo genome assemblies of two tube case-making Trichoptera from the suborder Integripalpia, *Agrypnia vestita* Walker and *Hesperophylax magnus* Banks. We find that these tube case-making caddisflies have genome sizes that are at least 3-fold larger than those of currently sequenced annulipalpian genomes and that this pattern is at least partly driven by major expansion of repetitive elements. In *H. magnus*, long interspersed nuclear elements alone exceed the entire genome size of some annulipalpian counterparts suggesting that caddisflies have high potential as a model for understanding genome size evolution in diverse insect lineages.

Key words: biodiversity genomics, Trichoptera, repetitive elements, insect genomics, caddisfly, freshwater insects.

Significance

There is a lack of genomic resources for aquatic insects. So far, in the insect order Trichoptera, only three high-quality genomes have been assembled, all from individuals in the retreat-making suborder Annulipalpia. In this article, we report the first high-quality genomes of two case-making species from the suborder Integripalpia, which are essential for studying genomic diversity across this ecologically diverse insect order. Our research reveals larger genome sizes in the tube case-makers (suborder Integripalpia, infraorder Phryganides), accompanied by a disproportionate increase of repetitive DNA. This suggests that genome size is at least partly driven by a major expansion of repetitive elements. Our work shows that caddisflies have high potential as a model for understanding how genomic diversity might be linked to functional diversification and forms the basis for detailed studies on genome size evolution in caddisflies.

Introduction

With 16,544 extant species (Morse 2020), caddisflies (Insecta: Trichoptera) are the most diverse of the primary aquatic insect orders, comprising more species than the other four (Odonata, Ephemeroptera, Plecoptera, and Megaloptera) combined (Dijkstra et al. 2014). This diverse group of insects has successfully colonized all types of freshwater (and even intertidal) habitats across all continents north of Antarctica. Within these freshwater ecosystems caddisflies play important roles, including nutrient cycling and energy flow, and stabilizing the waterbed. They also act as biological indicators of water quality (Morse et al. 2019). Trichoptera is divided into two suborders, Annulipalpia and Integripalpia, both of which produce silk in modified labial glands (Thomas, Frandsen, et al. 2020). Annulipalpians use silk to construct small homes and capture nets that are fixed to the substrate, whereas most integripalpians' use silk to connect material into portable tube cases offering protection, camouflage, and even aiding in respiration (fig. 1) (Wiggins 2004). This innovation in extended phenotype has potentially facilitated their radiation across a multitude of different environments including streams, lakes, ponds, and even marine environments.

Relative to their diversity, most insect groups remain poorly represented in existing genomic resources—a trend which is particularly pronounced in aquatic insects (Hotaling et al. 2020). Yet insects commonly show dynamic genome evolution within groups, including major variation in genome size that is often linked to expansion and loss of repetitive DNA (Lower et al. 2017; Petersen et al. 2019; Pflug et al. 2020). Insect diversity offers a vast supply of potential model systems for understanding how genomes evolve, especially as advancing sequencing technology enables more cost-effective, high-quality genome assemblies in any model system. Currently, there are three long-read based draft Trichoptera genome assemblies (Luo et al. 2018; Heckenhauer et al. 2019; table 1). However, all of these were generated for species within the suborder Annulipalpia, leaving approximately 10,453 integripalpian species (Morse 2020) and 275 Myr of evolutionary history poorly represented by genomic resources (Thomas, Frandsen, et al. 2020). In addition, a lack of genetic resources in the large case-making radiation within caddisflies prevents research into the genomic basis of the fascinating evolutionary history and ecological diversification of this diverse and important group of caddisflies. Here, we report the first long-read based de novo genome assemblies and annotations of two tube-case making intergripalpian caddisflies, *Agrypnia vestita* Walker and *Hesperophylax magnus* Banks. Their estimated genome sizes are more than 3-fold larger than previously sequenced annulipalpian caddisflies. We show that this is, at least partly, due to a large expansion of repeat content in the case-making caddisflies compared with retreat-making caddisflies.

Materials and Methods

Sequencing and Assembly

We collected individuals of both species in the wild, *A. vestita* as an adult and *H. magnus* as a pupa. Following extraction, we sequenced genomic DNA from *A. vestita* on an Illumina HiSeq 2500 lane and on 23 PacBio sequel SMRT cells. We sequenced genomic DNA from *H. magnus* using two Illumina NovaSeq and four Oxford Nanopore FLO-MIN 106 flow cells. Further details are provided in supplementary note 1, Supplementary Material online. For both data sets, we conducted a de novo hybrid assembly using MaSuRCA v.3.1.1 (Zimin et al. 2013, 2017). MaSuRCA aligns high-fidelity short reads to more noisy long reads to generate "megareads," which are then assembled in CABOG (Miller et al. 2008), an overlap-layout-consensus assembler. In the config file for each run, we specified an insert size of 500 bp for the Illumina paired-end reads with a standard deviation of 50 and a Jellyfish hash size of 100,000,000,000. All other parameters were left as defaults. We screened genome assemblies for potential contaminants with BlobTools v1.0 (Laetsch and Blaxter 2017; supplementary note 2 and supplementary figs. 1 and 2, Supplementary Material online). Contigs consisting of contaminant DNA were subsequently removed from the final assemblies. We assessed genome quality and completeness with BUSCO v4.1.1 (Seppey et al. 2019; supplementary note 3 and supplementary table 1,

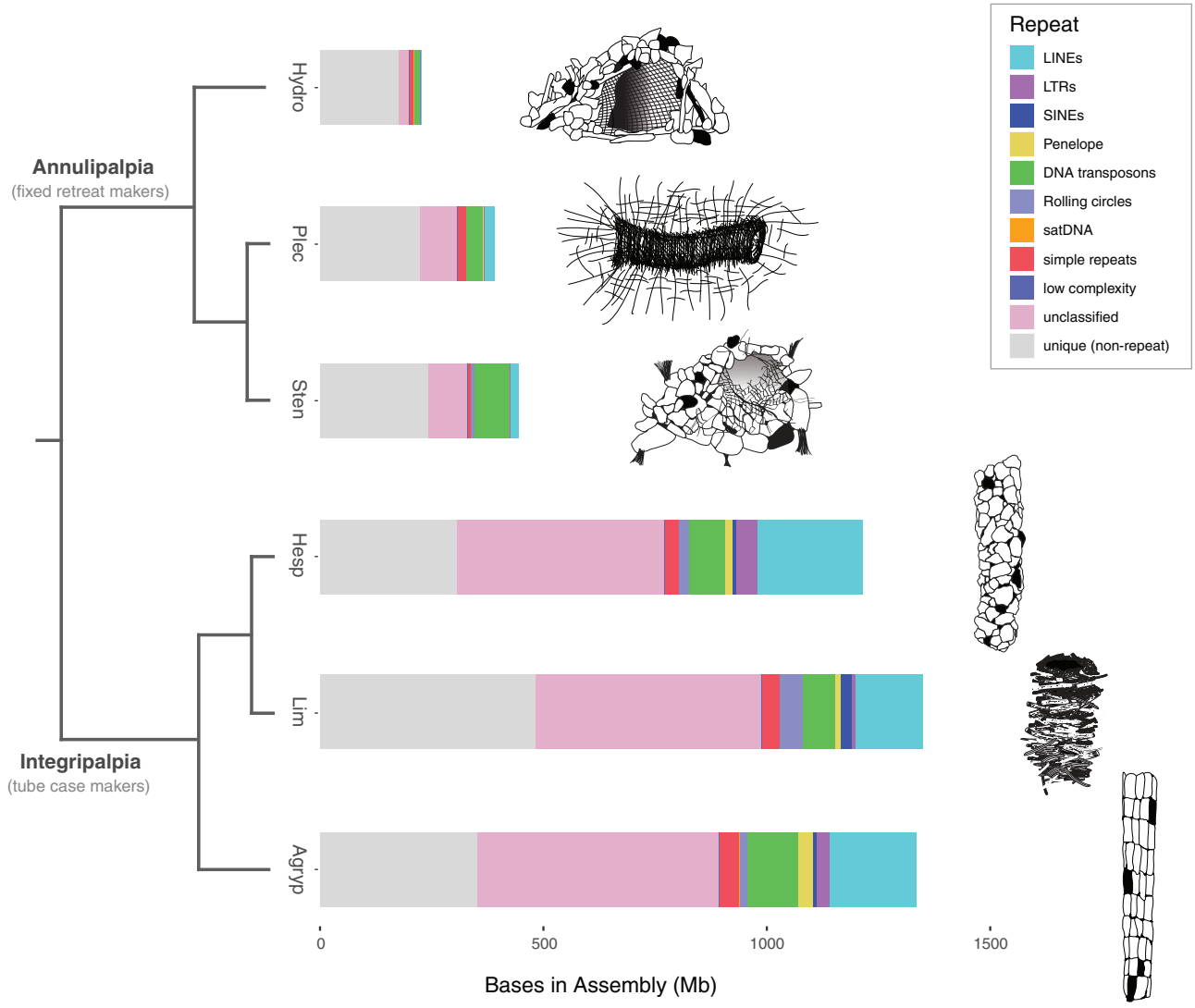

Fig. 1.—Assembly length and repetitive DNA content in Trichoptera suborders. Comparison of assembly length and repetitive DNA content among genome assemblies for three annulipalpan (*Hydropsyche tenuis*, *Plectrocnemia conspersa*, and *Stenopsyche tienmushanensis*) and three integripalpian (*Hesperophylax magnus*, *Limnephilus lunatus*, and *Agrypnia vestita*) species. The total length of bars indicates assembly size and colored segments within bars indicate the fraction of the assembly belonging to major repeat categories identified by RepeatModeler2 and annotated by RepeatMasker. Artwork to the right of plots shows examples of fixed retreats built by annulipalpians compared with integripalpian tube cases. Each Illustration is derived from a member of the same genus as the genome assemblies. The phylogeny is based on Thomas, Frandsen, et al. (2020).

Supplementary Material online) with the OrthoDB v.10 Insecta and Endopterygota gene sets (Kriventseva et al. 2019) and generated genome statistics using the assembly_stats script (Trizna 2020; supplementary table 1, Supplementary Material online, for full output). We conducted genome profiling (estimation of major genome characteristics such as size, heterozygosity, and repetitiveness) on the short-read sequence data with GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) and findGSE (Sun et al. 2018) as described in supplementary note 4, supplementary table 2, and supplementary figures 3 and 4, Supplementary Material online.

Repeat and Gene Annotation

We conducted comparative analysis of repetitive elements for the genomes generated in this study, the three previously available long-read based annulipalpian genomes, and *Limnephilus lunatus*, the integripalpian with the highest quality short-read genome assembly. We identified and classified repetitive elements de novo and generated a library of consensus sequences for each genome using RepeatModeler 2.0 (Flynn et al. 2019). We then annotated and masked repeats in each assembly with RepeatMasker 4.1.0 (Smit et al. 2013–2015) using the custom repeat library for the species generated in the previous step. Finally, we reran RepeatMasker on

Table 1
Comparison of Genome Assemblies against Previously Published Caddisfly Genomes

Species	Accession	Suborder	Sequencing Platform Coverage	Assembly Length (bp)	Contig N50 (kb)	BUSCOs Present (%)[a,b]
Agrpynia vestita (present study)	JADDOH000000000	Integripalpia	PacBio + Illumina (17.86× + 87.96×)	1,353,059,149	111.8	94.4
Hesperophylax magnus (present study)	JADDOG000000000	Integripalpia	Nanopore + Illumina (26.38× + 49.30×)	1,233,588,871	768.2	95.9
Hydropsyche tenuis (Heckenhauer et al. 2019)	GCA_009617725.1	Annulipalpia	Nanopore + Illumina (16.5× + 167.6×)	229,663,394	2,190.1	98.4
Plectrocnemia conspersa (Heckenhauer et al. 2019)	GCA_009617715.1	Annulipalpia	Nanopore + Illumina (17.1× + 82.9×)	396,695,105	869.0	98.7
Stenopsyche tienmushanensis (Luo et al. 2018)	GCA_008973525.1	Annulipalpia	PacBio + Illumina (153× + 150×)	451,494,475	1,296.9	98.1
Limnephilus lunatus (Thomas, Dohmen, et al. 2020)	Llun_2.0	Integripalpia	Illumina (80.1)	1,369,180,260	24.2	93.6
Glossosoma conforme (Weigand et al. 2018)	GCA_003347265.1	Annulipalpia	Illumina (53×)	604,293,666	14.2	94.9
Sericostoma sp. (Weigand et al. 2017)	GCA_003003475.1	Integripalpia	Illumina (43×)	1,015,727,762	2.1	74.0
Glyphotaelius pellucidus (Ferguson et al. 2014)	n.a.	Integripalpia	Illumina (8.12×)	757,289,448	0.656	64.4

[a]Present = complete + fragmented.
[b]$N_{Insecta}$ = 1,367.

the masked genome using the Repbase arthropod repeat library (Bao et al. 2015). We conducted an orthogonal analysis of repeat dynamics using a reference-free approach by normalizing subsampled Illumina data for each sample using RepeatProfiler (Negm et al. 2020) and then analyzing normalized data sets for repeat content in RepeatExplorer2 (Novák et al. 2013), with more details provided in supplementary note 5, Supplementary Material online.

To generate evidence for gene annotation, we aligned previously sequenced transcriptomes to each genome using BLAST-like Alignment Tool v3.6 (BLAT, Kent 2002). We aligned the transcriptome of the closely related *Phryganea grandis* from the 1KITE project (111126_I883_FCD0GUKACXX_L7_INShauTBBRAAPEI-22 http://www.1kite.org/) to *A. vestita*, and we aligned the *Hesperophylax* transcriptome from (Wang et al. 2015) to *H. magnus*. We generated ab initio gene predictions using AUGUSTUS v3.3 (Stanke et al. 2008) with hints generated from RepeatMasker 4.1.0 and BLAT v3.6 and by supplying the retraining parameters obtained from the BUSCO analysis (supplementary note 6, Supplementary Material online). Following annotation, we removed genes from our annotation that did not generate significant BLAST hits or lacked transcript evidence. Lastly, functional annotations were identified using Blast2GO (Götz et al. 2008).

Results and Discussion

Assembly

Here, we generated the first genome assemblies based on long-read sequencing from the species diverse caddisfly suborder, Integripalpia. They provide important genome resources and fill a gap in evolutionary history of more than 275 Myr (Thomas, Frandsen, et al. 2020). The *A. vestita* genome was sequenced using ~88× Illumina sequence coverage and ~18× PacBio read coverage. After contaminated contigs were removed, the resulting assembly contained 25,541 contigs, a contig N50 of 111,757 bp, GC content of 33.77%, and a total length of 1,352,945,503 bp. BUSCO analysis identified 94.4% (91.4% complete, 3.0% fragmented) of the Insecta gene set in the assembly (see supplementary note 3, Supplementary Material online, for further details). The *H. magnus* genome was sequenced with ~49× Illumina sequence coverage and ~26× Oxford Nanopore sequence coverage. The resulting assembly has 6,877 contigs, a contig N50 of 768,217 bp, GC content of 34.36%, and a total length of 1,275,967,528 bp. We identified 95.9% (95.2% complete, 0.7% fragmented) of the Insecta BUSCO gene set in the final assembly. Although the sequencing and assembly techniques were similar to those used in previous efforts to sequence and assemble high-quality reference genomes in Trichoptera (table 1, Luo et al. 2018; Heckenhauer et al. 2019), the contiguity of these genomes was lower. This is likely to have been caused by large genome size and the proliferation of repetitive DNA, which represents one of the primary barriers to genome assembly. However, despite these challenges, both genome assemblies represent a substantial improvement in contiguity to previous assemblies of integripalpian caddisflies generated from short-read data alone. For example, at the time of writing, the highest quality integripalpian genome assembly on GenBank is

Limnephilus lunatus, which was assembled from short-read data and has a contig N50 of 24.2 kb, giving further evidence to the difficulty of assembling large, repetitive caddisfly genomes.

Annotation and Repeat Analysis

We also report the functional annotations of *H. magnus* and *A. vestita*. Of 59,600 proteins predicted by AUGUSTUS for *A. vestita*, 21,637 were verified by BLAST and/or transcript evidence (and maintained in the final annotation), 14,096 were mapped to GO terms, and 5,362 were functionally annotated in Blast2GO. Of 38,490 proteins predicted by AUGUSTUS for *H. magnus*, 16,791 were verified by BLAST and/or transcript evidence (and maintained in the final annotation), 10,605 were mapped to GO terms, and 5,362 were functionally annotated in Blast2GO. Top GO annotations include cellular process (*A. vestita* 3395, *H. magnus* 2366), metabolic process (*A. vestita* 2897, *H. magnus* 2047), binding (*A. vestita* 3425, *H. magnus* 2187), and catalytic activity (*A. vestita* 3243, *H. magnus* 2395) (supplementary figs. 6 and 7, Supplementary Material online).

The results of genome assembly repeat annotation, genome profiling, and de novo repeat assembly with RepeatExplorer2 all showed a disproportionate increase of repetitive DNA in integripalpian genomes compared with annulipalpians for those species sampled (fig. 1, supplementary note 5 and supplementary figs. 5–7, Supplementary Material online). In the integripalpian species, unclassified repeats alone make up an average of >500 million bases, which exceeds the estimated genome sizes of all three annulipalpians analyzed. After unclassified repeats, long interspersed nuclear elements (LINEs) are the most abundant repeat category showing a disproportionate increase. LINEs comprise an average of >200 million bases in integripalpians, and show a ~4-fold average increase in genome proportion (avg. genome proportion = 15.1%) compared with the annulipalpians (avg. genome proportion = 3.8%). *Hesperophylax* has more bases annotated as LINEs (~249 million) than the size of the entire *Hydropsyche* genome assembly. Rolling circles and long terminal repeats (LTRs) also show disproportionate increase in integripalpians (~3.5- and 18-fold increases in genome proportion, respectively), however both categories make up a much smaller fraction of integripalipan genomes (<2.3% on an average). DNA transposons are abundant in all integripalpian genomes we studied (average of 92 million bases annotated), however their genomic proportion decreased relative to annulipalipans in which DNA transposons were the most abundant classifiable repeat category (11.3% avg. genome proportion in annulipalipans vs. 6.9% in integripalpians).

The high abundance of unclassified repeats we observed in the integripalpian genomes is not surprising given that Trichoptera repeats are poorly represented in repeat databases. Unclassified repeats may also represent the remnants of ancient transposable element expansions, which are particularly difficult to annotate (Hoen et al. 2015). This explanation of old repeat expansions accounting for much of the unclassified repeats is consistent with results of clustering analysis in RepeatExplorer2 which shows many unannotated superclusters that make up small fractions of the genome (supplementary fig. 3, Supplementary Material online). We do not observe large unannotated superclusters that would indicate failed annotation of abundant, recently active repeats. Given the apparent suborder-specific increase in unclassified repeats, we hypothesize that ancient transposable element activity in the ancestor of integripalpians contributed to the larger genome sizes we observe; however, denser sampling of genomes across Trichoptera suborders is needed to address this hypothesis.

Given the major variation in genome size and repeat abundance, our findings suggest Trichoptera has high potential as a model for gaining insights into genome evolution in diverse insect lineages. Future investigation on the role of LINEs in genome diversification is of particular interest given our findings. We present preliminary evidence that LINEs show suborder-specific expansions, albeit with very limited taxon sampling. LINEs (especially L1) play major roles in genome stability, cancer, and aging (Van Meter et al. 2014; De Cecco et al. 2019). In many groups LINEs are hypothesized to play important evolutionary roles, including roles in rapid genome evolution though their own movement (Kordiš et al. 2006; Warren et al. 2008; Suh et al. 2015), and by facilitating expansion of other repeat classes (Grandi and An 2013; Sproul et al. 2020). It is possible that these elements have been important drivers in the expansion of integripalpian genomes. The high-quality genome assemblies and repeat libraries we present here provide a starting point for investigating the role of repeats in genome evolution across caddisfly lineages. In addition, we close a large evolutionary gap in genomic resources within a large, ecologically diverse clade in which additional genome sequencing can enable new insights as to the genomic basis of adaptation and diversification within freshwater environments.

Supplementary Material

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

Acknowledgments

This material is based upon work supported by the Global Genome Initiative under Grant No. GGI-Exploratory-2016-049. J.H. was supported by LOEWE-Centre for Translational Biodiversity Genomics funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). J.S.S. was supported by an NSF Postdoctoral Research Fellowship in Biology (DBI-1811930). Part of the computation was

performed on the Smithsonian High-Performance Computing Cluster (SI/HPC: doi.org/10.25572/SIHPC). L.K.O. was supported by a BYU College of Life Sciences Undergraduate Research Award.

Author Contributions

Conceptualization: L.K.O., J.H., V.L.G., M.P.K., R.B.D., S.U.P., J.S.S., and P.B.F. Methodology: L.K.O., J.H., J.S.S., V.L.G., M.P.K., R.B.D., S.U.P., and P.B.F. Software: AMT. Validation: L.K.O., J.S.S., and P.B.F. Formal analysis: J.H., J.S.S., L.K.O., R.B.D., and P.B.F. Investigation: V.L.G., A.M.T., S.B.W., and P.B.F. Resources: V.L.G., R.H., R.J.S., and P.B.F. Data curation: S.B.W., A.M.T., and R.B.D. Writing-original draft preparation: L.K.O., J.H., J.S.S., and P.B.F. Writing-review and editing: L.K.O., J.H., M.P.K., R.B.D., J.S.S., S.U.P., A.M.T., R.J.S., R.H., X.Z., and P.B.F. Visualization: J.S.S. and R.H. Supervision: P.B.F. Project administration: L.K.O. and P.B.F. Funding acquisition: L.K.O., V.L.G., M.P.K., R.B.D., S.U.P., R.J.S., X.Z., and P.B.F.

Data Availability

This project has been deposited at NCBI under the Bioproject ID: PRJNA668166 and the accession JADDOG000000000 and JADDOH000000000; the annotations and predicted peptides are available on FigShare at the link: https://doi.org/10.6084/m9.figshare.c.5266130.v1.

Literature Cited

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6(1):11.

De Cecco M, et al. 2019. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature 566(7742):73–78.

Dijkstra K-DB, Monaghan MT, Pauls SU. 2014. Freshwater biodiversity and aquatic insect diversification. Annu Rev Entomol. 59(1):143–163.

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.

Flynn JM, et al.. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 117(17):9451–9457.

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

Grandi FC, An W. 2013. Non-LTR retrotransposons and microsatellites. Mob Genet Elem. 3(4):e25674.

Heckenhauer J, et al. 2019. Annotated draft genomes of two caddisfly species *Plectrocnemia conspersa* CURTIS and *Hydropsyche tenuis* NAVAS (Insecta: Trichoptera). Genome Biol Evol. 11(12):3445–3451.

Hoen DR, et al. 2015. A call for benchmarking transposable element annotation methods. Mob DNA. 6(1):13.

Hotaling S, Kelley JL, Frandsen PB. 2020. Aquatic insects are dramatically underrepresented in genomic research. Insects 11(9):601.

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

Kordiš D, Lovšin N, Gubenšek F. 2006. Phylogenomic analysis of the L1 retrotransposons in deuterostomia. Syst Biol. 55(6):886–901.

Kriventseva EV, et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47(D1):D807–D811.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Res. 6:1287.

Lower SS, et al. 2017. Genome size in North American fireflies: substantial variation likely driven by neutral processes. Genome Biol Evol. 9(6):1499–1512.

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). doi: 10.1093/gigascience/giy143.

Miller JR, et al. 2008. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics 24(24):2818–2824.

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

Morse JC. 2020. Trichoptera world checklist. Available online: http://entweb.sites.clemson.edu/database/trichopt/.

Negm S, Greenberg A, Larracuente AM, Sproul JS. 2020. RepeatProfiler: a pipeline for visualization and comparative analysis of repetitive DNA profiles. Mol Ecol Resour. doi: 10.1111/1755-0998.13305.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29(6):792–793.

Petersen M, et al. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 19(1):11.

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. 2020. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3 (Bethesda) 10:3047–3060.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. Gene prediction: methods and protocols. New York: Springer. p. 227–245.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0 [cited 2019 Dec 4]. Available from: http://www.repeatmasker.org.

Sproul JS, Barton LM, Maddison DR. 2020. Repetitive DNA profiles reveal evidence of rapid genome evolution and reflect species boundaries in ground beetles. Syst Biol. 69(6):1137–1148.

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.

Suh A, et al. 2015. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. Genome Biol Evol. 7(1):205–217.

Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. Bioinformatics 34(4):550–557.

Thomas GWC, Dohmen E, et al. 2020. Gene content evolution in the arthropods. Genome Biol. 21(1):15.

Thomas JA, Frandsen PB, Prendini E, Zhou X, Holzenthal RW. 2020. A multigene phylogeny and timeline for Trichoptera (Insecta). Syst Entomol. 45(3):670–686.

Trizna M. 2020. assembly_stats 0.1.4. Zenodo. Available from: 10.5281/zenodo.3968775.

Van Meter M, et al. 2014. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat Commun. 5(1):5011.

Wang C-S, Pan H, Weerasekare GM, Stewart RJ. 2015. Peroxidase-catalysed interfacial adhesion of aquatic caddisworm silk. J R Soc Interface. 12(112):20150710.
Warren WC, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453(7192):175–183.
Weigand H, et al. 2017. Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. Mol Ecol. 26(20):5705–5715.
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.
Wiggins GB. 2004. Caddisflies—the underwater architects. Toronto, ON, Canada: University of Toronto Press.
Zimin AV, et al. 2013. The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677.
Zimin AV, et al. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res. 27(5):787–792.

Associate editor: Venkatesh B.