ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 12/12/2022

DOI: https://doi.org/10.3897/arphapreprints.e98641

Holotype sequencing of *Silvatares holzenthali* (Trichoptera, Pisuliidae)

Jacqueline Heckenhauer, Ernesto Rázuri-Gonzales, François Ngera, Julio Schneider, Steffen Pauls

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 12/12/2022* DOI: https://doi.org/10.3897/arphapreprints.e98641

Holotype sequencing of *Silvatares holzenthali* (Trichoptera, Pisuliidae)

Jacqueline Heckenhauer[1,2,+,*], Ernesto Razuri-Gonzales[1,+], Francois Ngera Mwangi[3], Julio Schneider[1], Steffen U. Pauls[1,4,]

[1] Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany
[2] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany
[3] Centre de Recherche en Sciences Naturelles, Lwiro, Bukavu, Democratic Republic of the Congo
[4] Institute for Insect Biotechnology, Justus-Liebig-University, Gießen, Germany
***** Correspondence
+ equal contribution

JH: Investigation, Formal analysis, Validation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization
ER: Writing – Review & Editing, Visualization
FNM: Investigation
JS: Investigation, Writing – Review & Editing
SUP: Conception, Writing - Original Draft, Writing - Review & Editing, Funding acquisition

Abstract
While “DNA barcodes” are often provided, the whole mitochondrial and nuclear genome are rarely considered to be included in species descriptions. This is unfortunate because whole genome sequencing of holotypes allows eternal genetic characterization of the most representative specimen for a given species. Thus, *de novo* genomes are invaluable added resources and important additional diagnostic characters in species descriptions, provided the integrity of the holotype specimens remains intact. Here, we used a minimally invasive method to extract DNA of the type specimen of the recently described caddisfly species *Silvatares holzenthali (Trichoptera, Pisuliidae)* from the Democratic Republic of the Congo. A low-cost next generation sequencing strategy was used to generate the complete mitochondrial and draft nuclear genome of the holotype. The data in its current form is an important extension to the morphological species description and valuable for phylogenomic studies.

Keywords
museomics, extended specimen, holotype genomics, taxonomy, Trichoptera

Introduction
In entomology, new species are often not recognized as such in the field due to the minute size of the structures used to differentiate them from already described species. Intensive treatment (e.g., preparation and preservation) and careful examination of the collected specimens are required to determine if they are indeed undescribed. In addition, many new species are discovered in regions of the world where the scientific infrastructure is insufficient to guarantee high-quality, unfragmented DNA in collected specimens. Such was also the case for the holotype of *Silvatares holzenthali* (Trichoptera: Pisuliidae) (Rázuri-Gonzales et al. 2022). This specimen

(SMFTRI00018633) was collected by FNM in the eastern D.R. Congo in 2017 into locally produced 80% ethanol. By the time the specimen was identified as representing a new species, it had been transferred into new ethanol, analyzed multiple times under the stereoscope, and shipped between countries. Without the possibility of cooling the preservative or the specimen in the D.R. Congo, it was clear that the DNA of this specimen would be substandard to what might be extractable from a freshly caught caddisfly specimen preserved in high-quality ethanol and with uninterrupted cooling. However, the described scenario for the holotype of *S. holzenthali* is the norm rather than the exception.

Many initiatives are currently trying to harness recent technological developments to sequence and produce reference genomes for all species on Earth (Lewin et al. 2018; Rhie et al. 2021; Blaxter et al. 2022; Formenti et al. 2022). In this context, a reference genome is a highly contiguous, accurate, and annotated genome assembly, which represents the structure and organization of the genome of a species at a specific point in time (Formenti et al. 2022). These endeavors are crucial for documenting the Earth's biodiversity at its most fundamental organization level (i.e., genomic diversity). Understandably, these initiatives focus first on those species that are relatively easy to sequence (i.e., often larger species where tissue is available without destroying the entire specimen and where targeted sampling of freshly collected tissues, cells, or specimens is possible). Attempts to sequence the genome of even the tiniest individuals with minimal input DNA are becoming possible (Schneider et al. 2021), but they still cannot reach the quality standards required for reference genome assemblies. The same is true for specimens and holotypes collected in scenarios similar to the one described above for *S. holzenthali*.

Another limitation of many genome sequencing initiatives is that they generally do not focus on the holotype of a species. However, in the currently accepted type-based taxonomy, the holotype (or, if necessary, the designated lectotype and neotype) serves as the reference for the species definition. For many species, sequencing a reference genome from the holotype is not a viable option. Many type specimens are old, and naturally, all type specimens are rare and of singular value. Thus, they require special care, and invasive DNA extraction methods for genome sequencing should not be used. Thus, reference genome sequencing initiatives that require ample amounts of high-quality DNA for long-read sequencing technologies are logically and correctly focused on less valuable specimens, at best, from the *locus typicus*. Nevertheless, sequencing the holotype of a species allows genetically characterizing the most representative specimen for a given species as an eternal digital reference. Here we show that using a minimally invasive method to extract low-quality DNA from poorly preserved specimens allows taxonomists to capture and present the entire genetic characterization of the holotype while maintaining its morphological and structural integrity.

Material and methods

DNA extraction, library preparation, whole genome sequencing, and sequence read processing

Genomic DNA was extracted as described in Rázuri-Gonzales et al. (2022). A total of 110 ng gDNA was sheared to a mean fragment size of about 420 bp using a Bioruptor Pico (Diagenode, Seraing, Belgium). Genomic libraries were prepared using the NEBNext Ultra II DNA Library Preparation Kit for Illumina (New England Biolabs, Ipswich, MA, U.S.A.) according to the manufacturer's manual. Adapters were diluted

1:10 as recommended for low input libraries, and size selection was conducted based on the insert size using SPRIselect beads (Beckman, Indianapolis, U.S.A.). A dual indexing PCR was run for 8 cycles on a Mastercycler (Eppendorf, Germany). After cleanup, the library was eluted in 0.1X TE and shipped for 150 bp paired-end sequencing (ordering 20 Gbp output) on a partial lane of an Illumina NovaSeq 6000 platform (San Diego, CA) at Novogene (Cambridge, UK). Raw reads are deposited at the NCBI SRA archive under the accession number SRR22404850. The quality of the raw reads was evaluated using FastQC 0.11.9 (Andrews 2019). FASTQC reports were summarized with MultiQC 0.10 (Ewels et al. 2016). Raw reads were trimmed for low-quality regions, adapter sequences, and overrepresented *k-mers* using autotrim.pl 0.6.1 (Waldvogel et al. 2018) and Trimmomatic 0.39 (Bolger et al. 2014) using the adapter_all.fa of Trimmomatic and the following settings ILLUMINACLIP:2:30:10:8:true, SLIDINGWINDOW:4:20, MINLEN:50, and TOPHRED33. Unpaired reads were discarded. Contaminated reads were filtered using Kraken 2.0.9 (Wood and Salzberg 2014). The quality of trimmed, contamination-free reads was evaluated as described above.

Genome size estimation and genomic characterization

We used two different approaches to estimate the genome size. First, we used a *k-mer* distribution-based method. For this, *k-mers* were counted with JELLYFISH v2.3.0 (Marçais and Kingsford 2011) using jellyfish count -C -s 1000000000 -F 2 and a *k-mer* length of 21 based on the raw sequence reads. A histogram of *k-mer* frequencies was created with jellyfish histo and used for analysis with the online web tool GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) using the following parameters: *k-mer* length = 21, ploidy = 2, max *k-mer* coverage = 10000. In addition, we estimated genome size with a re-mapping-based approach using backmap.pl (Schell et al. 2017; Pfenninger et al. 2022). This wrapper script uses the following dependencies samtools (Li et al. 2009), bwa mem (Li 2013), qualimap (Okonechnikov et al. 2015), MultiQC (Ewels et al. 2016), bedtools (Quinlan and Hall 2010), and RScript (R Core Team 2021) to automatically map the trimmed, contamination-free reads to the assembly (see *de novo* nuclear genome assembly) with bwa mem. Then, it executes qualimap bamqc and finally estimates genome size by dividing the mapped nucleotides by the mode of the coverage distribution (>0).

Mitogenome assembly

The mitochondrial genome was first assembled with the raw reads using NOVOplasty 4.2 (Dierckxsens et al. 2016) using the following parameters: type = mito, genome range = 12000-22000, *k-mer* = 33, max memory = 100, read length = 150, insert size = 300, platform = illumina, paired = PE, insert size auto = yes. The partial sequence of the cytochrome c oxidase subunit I (COX1) gene of *Silvatares ensifera* KX291165.1 was used as seed input. All other parameters were kept as default. The circularized mitogenome was aligned to the complete mitochondrial sequence of *Phryganea cinerea* MG980616.1 with MAFFT in Geneious Prime 2022.1.1 with default settings to set the correct start position. Annotation of tRNAs, rRNAs, and protein-coding genes was done with MitoZ 2.3 (Meng et al. 2019) using the module annotate with genetic_code 5 and clade Arthropoda. Positions of trnL, trnT, and trnS were manually curated based on the alignment to *P. cinerea*. The mitochondrial genome assembly was deposited in GenBank under the accessions OP921089.

De novo nuclear genome assembly

Nuclear genome assembly was conducted in Spades 3.14.1 (Bankevich et al. 2012) with the default settings. After scaffolds smaller than 500 bp and those matching the mitochondrial genome assembly were filtered out, assembly statistics were calculated with Quast 5.0.2 (Gurevich et al. 2013), and quality was assessed in several ways. First, completeness was accessed via screening for single-copy orthologs with BUSCO 4.1.4 (Simão et al. 2015) using the endopterygota_odb10 dataset. Second, the backmapping rate of the trimmed reads to the assembly was calculated with backmap.pl 0.3 as described above (see *Genome size estimation and genomic characterization*). Third, the final genome assemblies were screened for potential contaminations with taxon-annotated GC-coverage (TAGC) plots using BlobTools v1.1.1 (Laetsch and Blaxter 2017). For this purpose, the bam file resulting from the backmapping analysis was converted to a blobtools readable cov file with blobtools map2cov. Taxonomic assignment for BlobTools was done with blastn 2.10.0+ (Camacho et al. 2009) using -task megablast and -e-value 1e-25. The blobDB was created and plotted from the cov file and blast hits. The nuclear draft genome assembly was deposited in GenBank under the accession JAPMAF000000000. All commands used in this study are given in the supplementary material.

Results

Whole genome sequencing and genome characterization

Illumina sequencing resulted in 160,534,832 raw short reads with a data amount of 24.1 G. 3.3% of reads were identified as contaminated (2.7% *Homo sapiens*, 0.6% bacteria, 0.1% viruses, 0.03% other). After trimming and contamination filtering, 173,132,236 reads (~21.95 G) were kept.

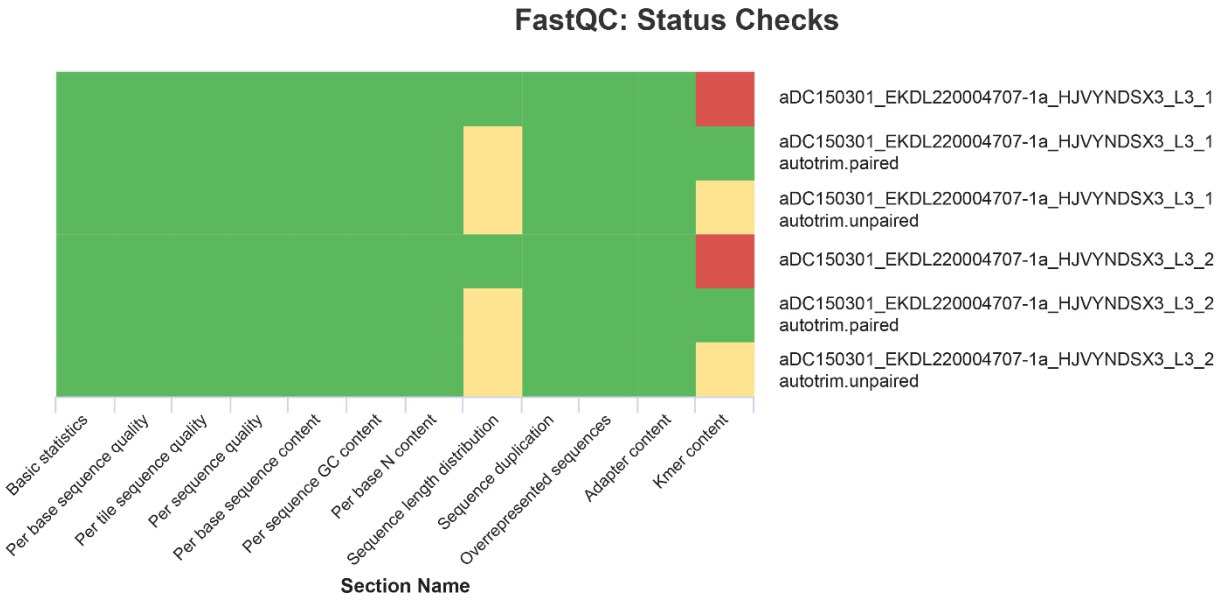

Figure 1. FastQC status checks of raw and trimmed reads (*autotrim), green: good, yellow: ok, red: failed.

K-mer analysis based on raw read data estimated the genome size to be 531,154,828 bp and heterozygosity 37.7%.

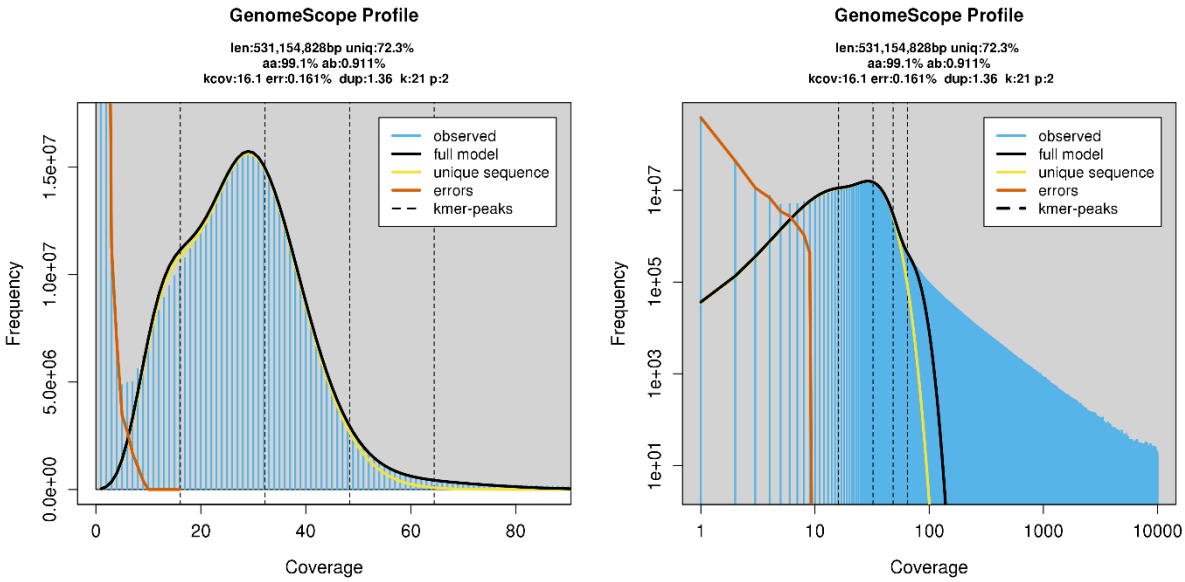

Figure 2. Genomescope2 profiles. Left: linear plot, right: log plot; len: inferred total genome length, uniq: percent of the genome that is unique (not repetitive), het: overall rate of heterozygosity, kcov: mean *k-mer* coverage for heterozygous bases, err: error rate of the reads, dup: average rate of read duplications.

Backmap.pl revealed a genome size of 643.02Mb.

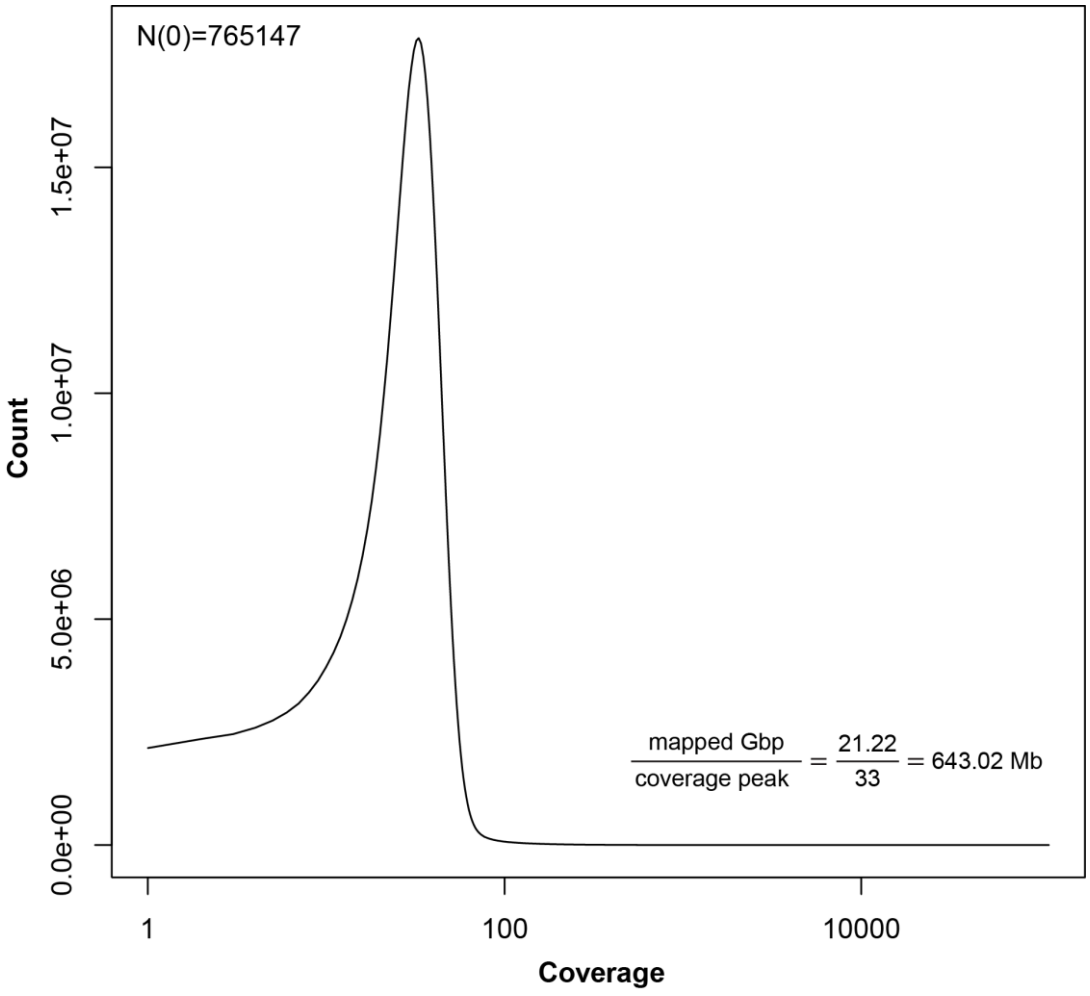

Figure 3. Coverage distribution per position. The x-axis is given in log-scale. Mapped nucleotides: 21.22Gb. The peak 643.02Mb coverage is 33.

Mitochondrial genome

The NOVOplasty assembly resulted in a 17,205 bp long and circularized contig. Its annotation revealed all expected 13 protein-coding genes and both rRNAs and 23 tRNAs. The d-loop was manually curated based on a comparison to the complete mitochondrial sequence of *Limnephilus decipiens* AB971912.1.

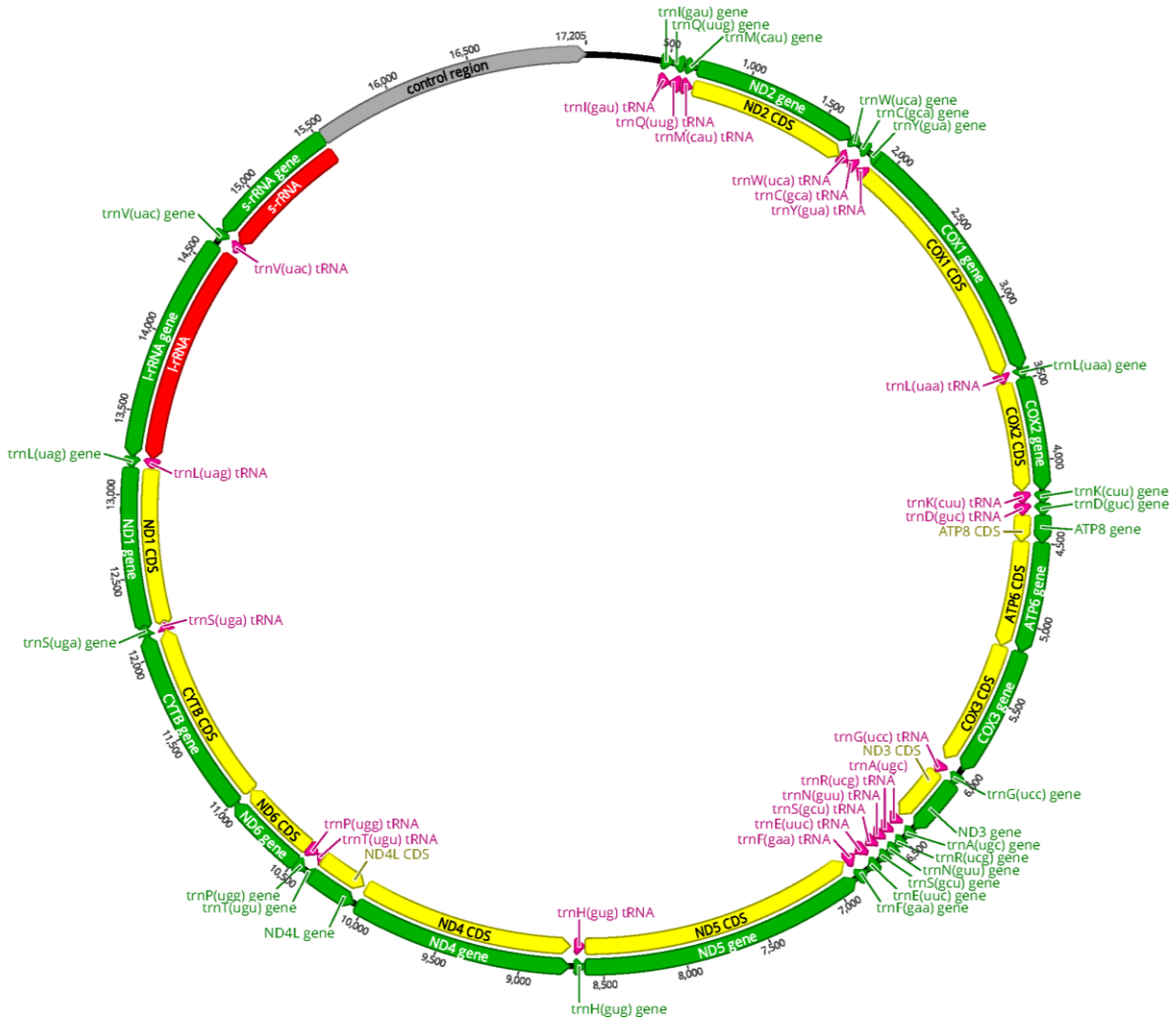

Figure 4. Circular mitochondrial genome of the holotype of *Silvatares holzenthali*.

Nuclear genome

The nuclear genome assembly contains 298,265 scaffolds with a total length of 534,501,986 bp, an N50 of 2,549, and a GC of 35.27%. 99.07% of reads were mapped back to the assembly. The BUSCO search with 2124 Endopterygota orthologs resulted in 74.7% BUSCOs; of these, 44.7% were complete (44.3% single, 0.4% duplicated), and 31% were fragmented. Blobtools detected no contaminations. While uploading the genome to NCBI, NCBI's contamination screening detected and filtered a 29 bp-long contamination (vector, etc.) at the beginning of one scaffold.

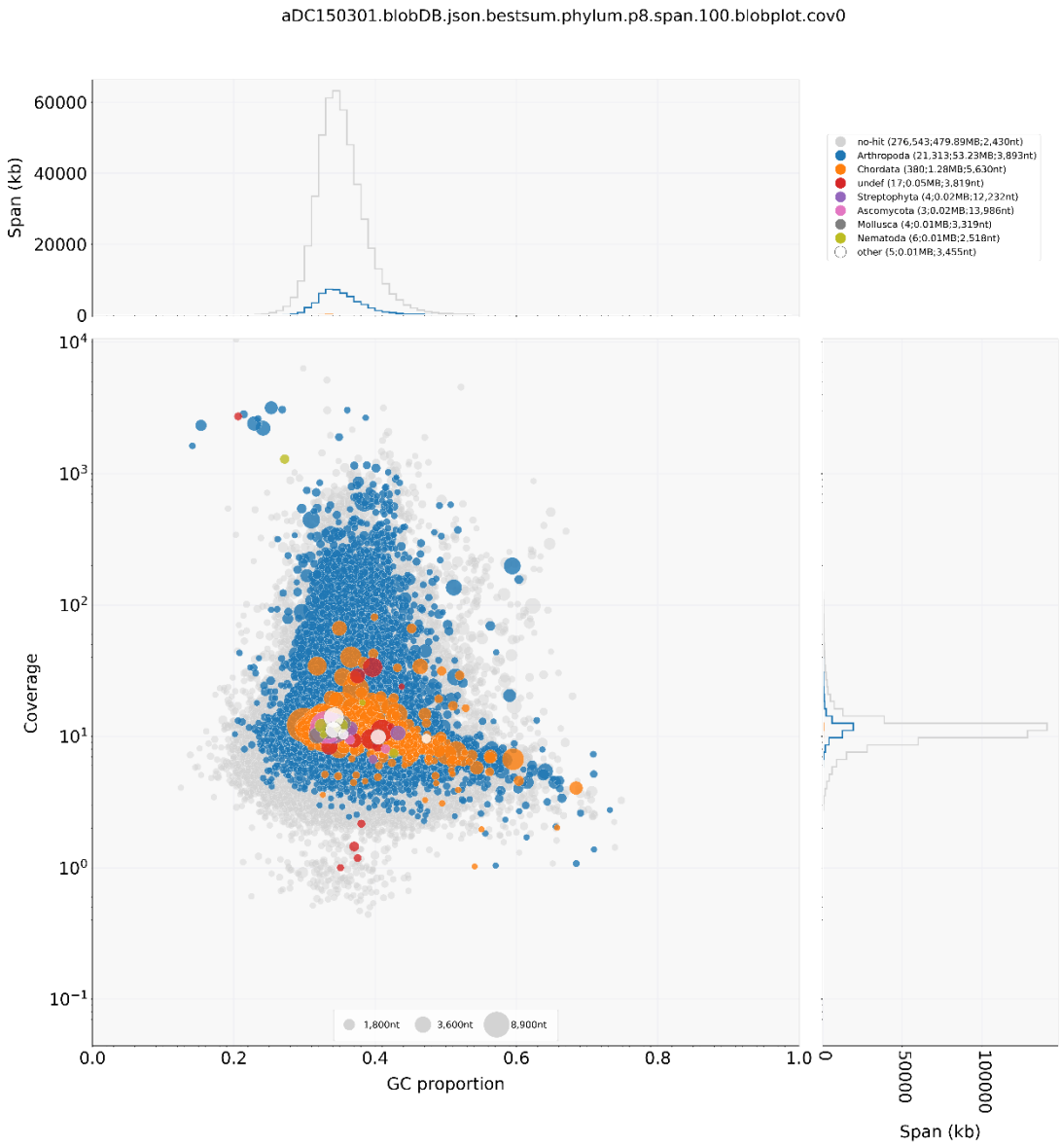

Figure 5. Taxon-annotated GC-coverage (TAGC) plots for the nuclear genome assembly. Scaffolds are represented with circles. Colors indicate the best match to the corresponding taxonomic annotation. The distribution of the total span (kb) of contigs for a given GC proportion or coverage is given in the upper- and right panels, respectively.

Discussion

The caddisfly genera *Silvatares* Navás 1931 and *Pisulia* Marlier 1943 belong to the African endemic family Pisuliidae. The taxonomic history and distribution of *Silvatares* were described by Stoltze (1989), discussed in detail by Prather and Holzenthal (2002), and most recently summarized by Rázuri-Gonzales et al. (2022). Species of *Silvatares* generally inhabit forested streams in Sub-Saharan Africa. The known species are larger than the species of *Pisulia*, the male maxillary palps form a long, flattened spatula, and both sexes have a 2-4-4 tibial spur formula (Stoltze 1989). All Pisuliidae larvae build cases constructed from plant materials and are triangular in cross-section (Stoltze 1989; SUP, FNM personal observation), but again late instar larvae of *Silvatares* are larger than those of *Pisulia*. Unfortunately, too few larvae and adults of

Silvatares have been associated to allow updating the larval species key of Stoltze (1989).

While the morphology of these species has been described extensively, there is little genetic information published so far. Less than a handful of partial genes have been published or uploaded to NCBI GenBank. For example, cadherin (CAD), cytochrome oxidase subunit 1 (cox1), and the 28S and 18S subunit ribosomal RNA are available for *S. ensifer* (Thomas et al. 2020, Zhou et al. 2016); CAD, isocitrate dehydrogenase, and cox 1 for *Silvatares* sp. (Malm et al. 2013); cox1 for *S. collyrifer* (Zhou et al. 2016); and cox1 for *S. thrymmifer* (National Center for Biotechnology Information 2022).

The genome assembly presented herein is an invaluable addition to the characterization of the holotype of *Silvatares holzenthali*. Our genome assembly includes all the partial genes that had been hitherto sequenced for other *Silvatares* species. Additionally, this assembly provides the complete mitogenome, including the barcode markers, highlighting that for marginally higher costs, we can produce much more genomic information on type specimens than the “DNA barcode,” which is often provided in species descriptions. In addition, we present a draft nuclear genome assembly (admittedly of poor quality) through our ~45x sequencing coverage of short-read data. For example, these data allows us to assess the heterozygosity of the type specimen as a proxy for population genetic variation at the time of sampling (Köhler et al. 2021a).

Furthermore, the data in its current form is valuable in a phylogenomic context. For other more sophisticated downstream genomic analyses, the provided data from the holotype can always be mapped to a higher-quality reference genome generated from specimens of lesser value and better DNA quality. Notably, the approach we present here also lends itself to museum specimens, which are usually of older date. Being able to tap into these immense and often irreplaceable resources for genomic study opens a wealth of scientific opportunity. In 2012, Pohl et al. already provided a complete short-read-based genome in the description of a new Strepsiptera species. Köhler et al. (2021a,b,c) also provided *de novo* genomes with the descriptions of a mud snake and three frog species. While we think a *de novo* genome is an invaluable added resource and important additional diagnostic character in any species description, priority should be given to preserving specimen integrity of the type specimens. The methods used to preserve and store specimens may thus not always allow for generating a *de novo* genome from holotypes.

Acknowledgements

Field work and taxonomy were funded by German Science Foundation Grant (DFG PA1617/4-1) to SUP. The genome sequencing 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). Tilman Schell (LOEWE-TBG, Frankfurt) is acknowledged for his advice on bioinformatic methods.

References

Andrews S (2019) FastQC: a quality control tool for high throughput sequence data. 0.11.9 ed, pp.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19: 455-477. doi:https://doi.org/10.1089/cmb.2012.0021.

Blaxter M, Mieszkowska N, Di Palma F, Holland P, Durbin R, Richards T, Berriman M, Kersey P, Hollingsworth P, Wilson W, Twyford A, Gaya E, Lawniczak M, Lewis O, Broad G, Howe K, Hart M, Flicek P, Barnes I (2022) Sequence locally, think globally: The Darwin Tree of Life Project. Proceedings of the National Academy of Sciences 119: e2115642118. doi:https://doi.org/10.1073/pnas.2115642118.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi:https://doi.org/10.1093/bioinformatics/btu170.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10: 421. doi:https://doi.org/10.1186/1471-2105-10-421.

Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Research: gkw955. doi:https://doi.org/10.1093/nar/gkw955.

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047-3048. doi:https://doi.org/10.1093/bioinformatics/btw354.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, Malukiewicz J, Mouton A, Oomen RA, Paez S, Palsbøll PJ, Pampoulie C, Ruiz-López MJ, Svardal H, Theofanopoulou C, De Vries J, Waldvogel A-M, Zhang G, Mazzoni CJ, Jarvis ED, Bálint M, Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Čiampor F, Ciofi C, Crottini A, Godoy JA, Hoglund J, Malukiewicz J, Mouton A, Oomen RA, Paez S, Palsbøll P, Pampoulie C, Ruiz-López MJ, Svardal H, Theofanopoulou C, De Vries J, Waldvogel A-M, Zhang G, Mazzoni CJ, Jarvis E, Bálint M, Aghayan SA, Alioto TS, Almudi I, Alvarez N, Alves PC, Amorim IR, Antunes A, Arribas P, Baldrian P, Berg PR, Bertorelle G, Böhne A, Bonisoli-Alquati A, Boštjančić LL, Boussau B, Breton CM, Buzan E, Campos PF, Carreras C, Castro LF, Chueca LJ, Conti E, Cook-Deegan R, Croll D, Cunha MV, Delsuc F, Dennis AB, Dimitrov D, Faria R, Favre A, Fedrigo OD, Fernández R, Ficetola GF, Flot J-F, Gabaldón T, Galea Agius DR, Gallo GR, Giani AM, Gilbert MTP, Grebenc T, Guschanski K, Guyot R, Hausdorf B, Hawlitschek O, Heintzman PD, Heinze B, Hiller M, Husemann M, Iannucci A, Irisarri I, Jakobsen KS, Jentoft S, Klinga P, Kloch A, Kratochwil CF, Kusche H, Layton KKS, Leonard JA, Lerat E, Liti G, Manousaki T, Marques-Bonet T, Matos-Maraví P, Matschiner M, Maumus F, Mc Cartney AM, Meiri S, Melo-Ferreira J, Mengual X, Monaghan MT, Montagna M, Mysłajek RW, Neiber MT, Nicolas V, Novo M, Ozretić P, Palero F, Pârvulescu L, Pascual M, Paulo OS, Pavlek M, Pegueroles C, Pellissier L, Pesole G, Primmer CR, Riesgo A, Rüber L, Rubolini D, Salvi D, Seehausen O, Seidel M, Secomandi S, Studer B, Theodoridis S, Thines M, Urban L, Vasemägi A, Vella A, Vella N, Vernes SC, Vernesi C, Vieites DR, Waterhouse RM, Wheat CW, Wörheide G, Wurm Y, Zammit G (2022) The era of reference genomes in conservation genomics. Trends in Ecology & Evolution 37: 197-202. doi:https://doi.org/10.1016/j.tree.2021.11.008.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072-1075. doi:https://doi.org/10.1093/bioinformatics/btt086.

Köhler G, Khaing KPP, Than NL, Baranski D, Schell T, Greve C, Janke A, Pauls SU (2021a) A new genus and species of mud snake from Myanmar (Reptilia, Squamata, Homalopsidae). Zootaxa 4915: 301-325. doi:https://doi.org/10.11646/zootaxa.4915.3.1.

Köhler G, Vargas J, Than NL, Schell T, Janke A, Pauls SU, Thammachoti P (2021b) A taxonomic revision of the genus *Phrynoglossus* in Indochina with the description of a new species and comments on the classification within Occidozyginae (Amphibia, Anura, Dicroglossidae). Vertebrate Zoology 71: 1-26. doi:https://doi.org/10.3897/vz.71.e60312.

Köhler G, Zwitzers B, Than NL, Gupta DK, Janke A, Pauls SU, Thammachoti P (2021c) Bioacoustics Reveal Hidden Diversity in Frogs: Two New Species of the Genus *Limnonectes* from Myanmar (Amphibia, Anura, Dicroglossidae). Diversity 13: 399. doi:https://doi.org/10.3390/d13090399.

Laetsch DR, Blaxter ML (2017) BlobTools: Interrogation of genome assemblies. F1000Research 6: 1287. doi:https://doi.org/10.12688/f1000research.12232.1.

Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IV, Hackett KJ, Haussler D, Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, Van Sluys M-A, Soltis PS, Xu X, Yang H, Zhang G (2018) Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences 115: 4325-4333. doi:https://doi.org/10.1073/pnas.1720115115.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. doi:https://doi.org/10.1093/bioinformatics/btp352.

Malm T, Johanson KA, Wahlberg N (2013) The evolutionary history of Trichoptera (Insecta): A case of successful adaptation to life in freshwater. Systematic Entomology 38: 459-473. doi:https://doi.org/10.1111/syen.12016.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764-770. doi:https://doi.org/10.1093/bioinformatics/btr011.

Marlier G (1943) Trichoptera. Exploration du Parc National Albert 1 Mission GF de Witte (1933-1935) Fascicule 44: 3-17.

Meng G, Li Y, Yang C, Liu S (2019) MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research 47: e63-e63. doi:https://doi.org/10.1093/nar/gkz173.

GenBank: Accession No. KC559654. https://www.ncbi.nlm.nih.gov/nuccore/KC559654.1 [accessed 06/12/2022).

Navás L (1931) Insectos del museo de Paris. Brotéria, Serie Zoologica 27: 114-136.

Okonechnikov K, Conesa A, García-Alcalde F (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292-294. doi:https://doi.org/10.1093/bioinformatics/btv566.

Pfenninger M, Schönnenbeck P, Schell T (2022) ModEst: Accurate estimation of genome size from next generation sequencing data. Molecular Ecology Resources 22: 1454-1464. doi:https://doi.org/10.1111/1755-0998.13570.

Pohl H, Niehuis O, Gloyna K, Misof B, Beutel R (2012) A new species of *Mengenilla* (Insecta, Strepsiptera) from Tunisia. ZooKeys 198: 79-102. doi:https://doi.org/10.3897/zookeys.198.2334.

Prather AL, Holzenthal RW (2002) The identity of *Silvatares excelsus* Navás, 1931. Nova Supplementa Entomologica (Proceedings of the 10th International Symposium on Trichoptera) 15: 231-234.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842. doi:https://doi.org/10.1093/bioinformatics/btq033.

R Core Team (2021) R: A language and Environment for Statistical Computing. Vienna, Austria, pp.

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 11: doi:https://doi.org/10.1038/s41467-020-14998-3.

Rázuri-Gonzales E, Ngera MF, Pauls SU (2022) A new species of *Silvatares* (Trichoptera, Pisuliidae) from the Democratic Republic of the Congo. ZooKeys 1111: 371-380. doi:https://doi.org/10.3897/zookeys.1111.85307.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C, Ko BJ, Chaisson M, Gedman GL, Cantin LJ, Thibaud-Nissen F, Haggerty L, Bista I, Smith M, Haase B, Mountcastle J, Winkler S, Paez S, Howard J, Vernes SC, Lama TM, Grutzner F, Warren WC, Balakrishnan CN, Burt D, George JM, Biegler MT, Iorns D, Digby A, Eason D, Robertson B, Edwards T, Wilkinson M, Turner G, Meyer A, Kautt AF,

Franchini P, Detrich HW, Svardal H, Wagner M, Naylor GJP, Pippel M, Malinsky M, Mooney M, Simbirsky M, Hannigan BT, Pesout T, Houck M, Misuraca A, Kingan SB, Hall R, Kronenberg Z, Sović I, Dunn C, Ning Z, Hastie A, Lee J, Selvaraj S, Green RE, Putnam NH, Gut I, Ghurye J, Garrison E, Sims Y, Collins J, Pelan S, Torrance J, Tracey A, Wood J, Dagnew RE, Guan D, London SE, Clayton DF, Mello CV, Friedrich SR, Lovell PV, Osipova E, Al-Ajli FO, Secomandi S, Kim H, Theofanopoulou C, Hiller M, Zhou Y, Harris RS, Makova KD, Medvedev P, Hoffman J, Masterson P, Clark K, Martin F, Howe K, Flicek P, Walenz BP, Kwak W, Clawson H, Diekhans M, Nassar L, Paten B, Kraus RHS, Crawford AJ, Gilbert MTP, Zhang G, Venkatesh B, Murphy RW, Koepfli K-P, Shapiro B, Johnson WE, Di Palma F, Marques-Bonet T, Teeling EC, Warnow T, Graves JM, Ryder OA, Haussler D, O'Brien SJ, Korlach J, Lewin HA, Howe K, Myers EW, Durbin R, Phillippy AM, Jarvis ED (2021) Towards complete and error-free genome assemblies of all vertebrate species. Nature 592: 737-746. doi:https://doi.org/10.1038/s41586-021-03451-0.

Schell T, Feldmeyer B, Schmidt H, Greshake B, Tills O, Truebano M, Rundle SD, Paule J, Ebersberger I, Pfenninger M (2017) An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biology and Evolution 9: 585-592. doi:https://doi.org/10.1093/gbe/evx032.

Schneider C, Woehle C, Greve C, D'Haese CA, Wolf M, Hiller M, Janke A, Bálint M, Huettel B (2021) Two high-quality de novo genomes from single ethanol-preserved specimens of tiny metazoans (Collembola). GigaScience 10: giab035. doi:https://doi.org/10.1093/gigascience/giab035.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210-3212. doi:https://doi.org/10.1093/bioinformatics/btv351.

Stoltze M (1989) The afrotropical caddisfly family Pisuliidae. Systematics, zoogeography, and biology (Trichoptera: Pisuliidae). Steenstrupia (Copenhagen) 15: 1-50.

Thomas JA, Frandsen PB, Prendini E, Zhou X, Holzenthal RW (2020) A multigene phylogeny and timeline for Trichoptera (Insecta). Systematic Entomology 45: 670-686. doi:https://doi.org/10.1111/syen.12422.

Waldvogel AM, Wieser A, Schell T, Patel S, Schmidt H, Hankeln T, Feldmeyer B, Pfenninger M (2018) The genomic footprint of climate adaptation in *Chironomus riparius*. Molecular Ecology 27: 1439-1456. doi:https://doi.org/10.1111/mec.14543.

Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 15: R46. doi:https://doi.org/10.1186/gb-2014-15-3-r46.

Zhou X, Frandsen PB, Holzenthal RW, Beet CR, Bennett KR, Blahnik RJ, Bonada N, Cartwright D, Chuluunbat S, Cocks GV, Collins GE, deWaard J, Dean J, Flint OS, Hausmann A, Hendrich L, Hess M, Hogg ID, Kondratieff BC, Malicky H, Milton MA, Morinière J, Morse JC, Mwangi FN, Pauls SU, Gonzalez MR, Rinne A, Robinson JL, Salokannel J, Shackleton M, Smith B, Stamatakis A, StClair R, Thomas JA, Zamora-Muñoz C, Ziesmann T, Kjer KM (2016) The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20160025. doi:https://doi.org/10.1098/rstb.2016.0025.