Check for updates

DATA NOTE

The genome sequence of the segmented worm, *Sthenelais limicola* (Ehlers, 1864) [version 1; peer review: 2 approved, 1 approved with reservations]

Teresa Darbyshire [iD][1], Mitchell Brennan[2], Sean McTierney[2],
Marine Biological Association Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Amgueddfa Cymru, Cardiff, UK
[2]Marine Biological Association, Plymouth, Devon, UK

Abstract
We present a genome assembly from an individual *Sthenelais limicola* (the segmented worm; Annelida; Polychaeta; Phyllodocida; Sigalionidae). The genome sequence is 1,131 megabases in span. Most of the assembly is scaffolded into nine chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 16.7 kilobases in length.

Keywords
Sthenelais limicola, segmented worm, genome sequence, chromosomal, Phyllodocida

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✔ ? ✔

	1	2	3
version 1 19 Jan 2023	✔ view	? view	✔ view

1. **Joseph Ryan**, University of Florida, St Augustine, USA

2. **Torsten H Struck** [iD], University of Oslo, Oslo, Norway

3. **Erich Schwarz** [iD], Cornell University, Ithaca, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Darbyshire T: Investigation, Resources, Writing – Original Draft Preparation; **Brennan M**: Investigation, Resources; **McTierney S**: Investigation, Resources;

How to cite this article: Darbyshire T, Brennan M, McTierney S *et al.* **The genome sequence of the segmented worm, *Sthenelais limicola* (Ehlers, 1864) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2023, **8**:31 https://doi.org/10.12688/wellcomeopenres.18856.1

First published: 19 Jan 2023, **8**:31 https://doi.org/10.12688/wellcomeopenres.18856.1

Species taxonomy

Eukaryota; Metazoa; Spiralia; Lophotrochozoa; Annelida; Polychaeta; Errantia; Phyllodocida; Sigalionidae; *Sthenelais*; *Sthenelais limicola* (Ehlers, 1864) (NCBI:txid1210413).

Background

One of three *Sthenelais* species considered valid for UK and Irish waters, *Sthenelais limicola* is widely distributed throughout the northeast Atlantic, including the Mediterranean, and is also recorded from the northwest and southeast Atlantic (Barnich & Fiege, 2003; Chambers & Muir, 1997). Found from littoral (Parapar *et al.*, 2015) to bathyal depths (1,550 m) (Hartmann-Schröder, 1996), it inhabits sand and muddy substrates (Parapar *et al.*, 2015).

Sthenelais limicola can be distinguished from other European species by the elytra having a smooth margin, bifurcate or notched, and a smooth surface with only a few microtubercles close to the point of attachment. Animals are generally colourless or white, with transparent elytra, often with a brownish patch on the posterior half, forming a V-shape or horseshoe appearance when combined with the appearance of the opposite elytron, and can reach up to 100 mm in size (Hartmann-Schröder, 1996).

(Jumars *et al.*, 2015) classified all Sigalionidae as carnivores, although the method of hunting and prey are unknown for *S. limicola*. Little data is available on reproduction in Sigalionidae, but they are gonochoric as far as it is known (Rouse *et al.*, 2022). It is neither under threat nor considered as a non-native species anywhere in the world.

The genome of the segmented worm, *Sthenelais limicola*, was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland.

Genome sequence report

The genome was sequenced from one *S. limicola* (Figure 1) collected from East Breakwater, Plymouth Sound, UK (50.34, –4.14). A total of 45-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 444 missing or mis-joins and removed 210 haplotypic duplications, reducing the assembly length by 3.71% and the scaffold number by 72.31%, and decreasing the scaffold N50 by 2.18%.

The final assembly has a total length of 1,131.1 Mb in 201 sequence scaffolds with a scaffold N50 of 137.0 Mb (Table 1). Most (99.59%) of the assembly sequence was assigned to nine chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). The scaffold order and orientation is uncertain on chromosome 6 (0.39–8.86 Mb). Heterozygous inversion was observed on chromosome 7 (76.84–95.76 Mb). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype

have also been deposited. The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 95.4% (single 94.3%, duplicated 1.0%) using the metazoa_odb10 reference set.

Methods

Sample acquisition and nucleic acid extraction

An individual *S. limicola* (wpSthLimi1; Figure 1) was collected from East Breakwater, Plymouth Sound, UK (latitude 50.34, longitude –4.14) by Teresa Darbyshire (National Museum Wales) and Mitchell Brennan and Sean McTierney (Marine Biological Association) and identified by Teresa Darbyshire. The sample was collected from muddy substrate using a grab sampler (MV Sepia) and preserved in liquid nitrogen.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The wpSthLimi1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Mid-body tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. High molecular weight (HMW)
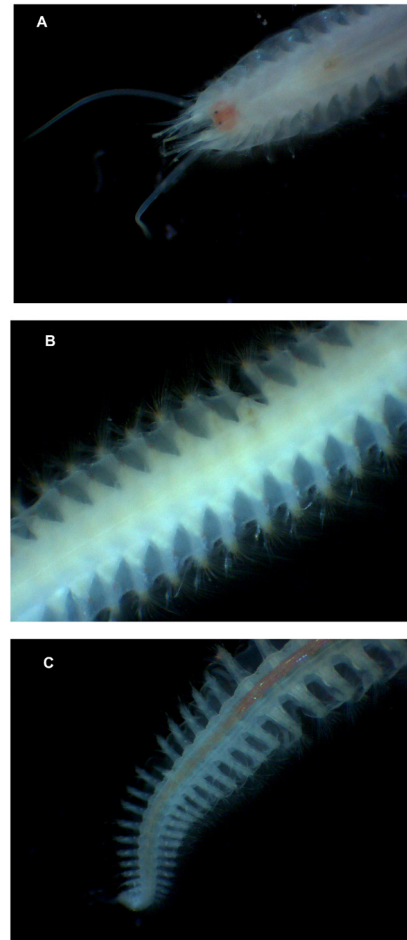

Figure 1. Photographs of the Sthenelais limicola (wpSthLimi1) specimen used for genome sequencing showing the dorsum of the head (**A**), mid body (**B**) and tail (**C**).

Table 1. Genome data for *Sthenelais limicola*, wpSthLimi1.1.

Project accession data		
Assembly identifier	wpSthLimi1.1	
Species	*Sthenelais limicola*	
Specimen	wpSthLimi1	
NCBI taxonomy ID	1210413	
BioProject	PRJEB51037	
BioSample ID	SAMEA8724794	
Isolate information	Mid-body tissue (DNA, RNA, Hi-C sequencing)	
Assembly metrics*		*Benchmark*
Consensus quality (QV)	59.9	*≥ 50*
k-mer completeness	100	*≥ 95%*
BUSCO**	C:95.4%[S:94.3%,D:1.0%], F:2.8%,M:1.8%,n:954	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.59%	*≥ 95%*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8978461–ERR8978463	
Hi-C Illumina	ERR8702825	
PolyA RNA-Seq Illumina	ERR10123685	
Genome assembly		
Assembly accession	GCA_942159475.1	
Accession of alternate haplotype	GCA_942183725.1	
Span (Mb)	1,131.1	
Number of contigs	1,310	
Contig N50 length (Mb)	10.2	
Number of scaffolds	201	
Scaffold N50 length (Mb)	137.0	
Longest scaffold (Mb)	161.9	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from (Rhie *et al.,* 2021).

** BUSCO scores based on the metazoa_odb10 BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/wpSthLimi1.1/dataset/CALNXF01/busco.

DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.
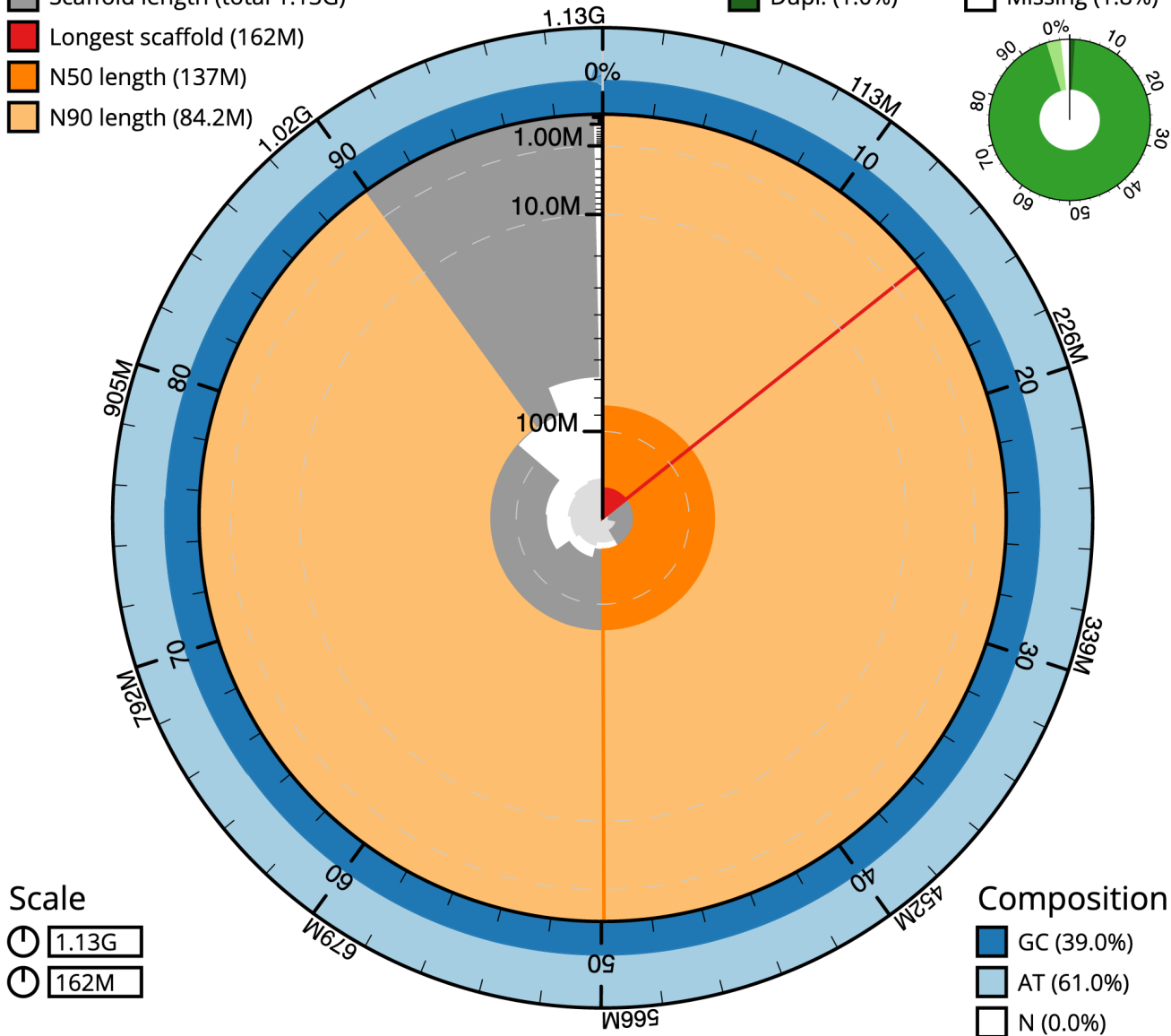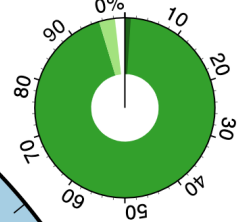
RNA was extracted from the mid-body tissue of wpSthLimi1 in the Tree of Life Laboratory at the WSI using TRIzol,

Figure 2. Genome assembly of *Sthenelais limicola*, wpSthLimi1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,131,150,536 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (161,876,864 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (136,979,486 and 84,233,480 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the metazoa_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/wpSthLimi1.1/dataset/CALNXF01/snail.

according to the manufacturer's instructions. RNA was then eluted in 50 µl RNAse-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.
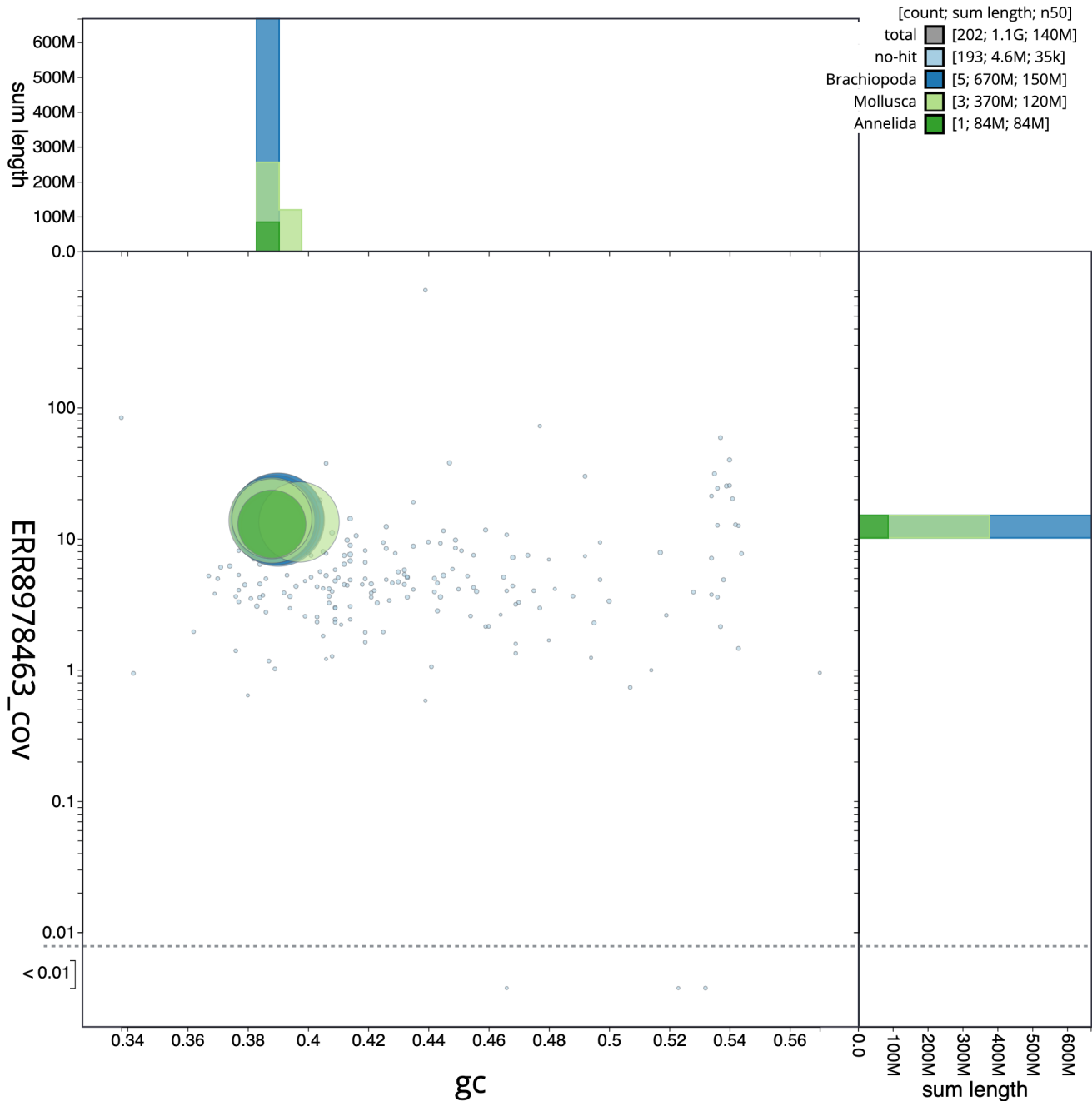
Figure 3. Genome assembly of *Sthenelais limicola*, wpSthLimi1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/wpSthLimi1.1/dataset/CALNXF01/blob.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from wpSthLimi1 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then
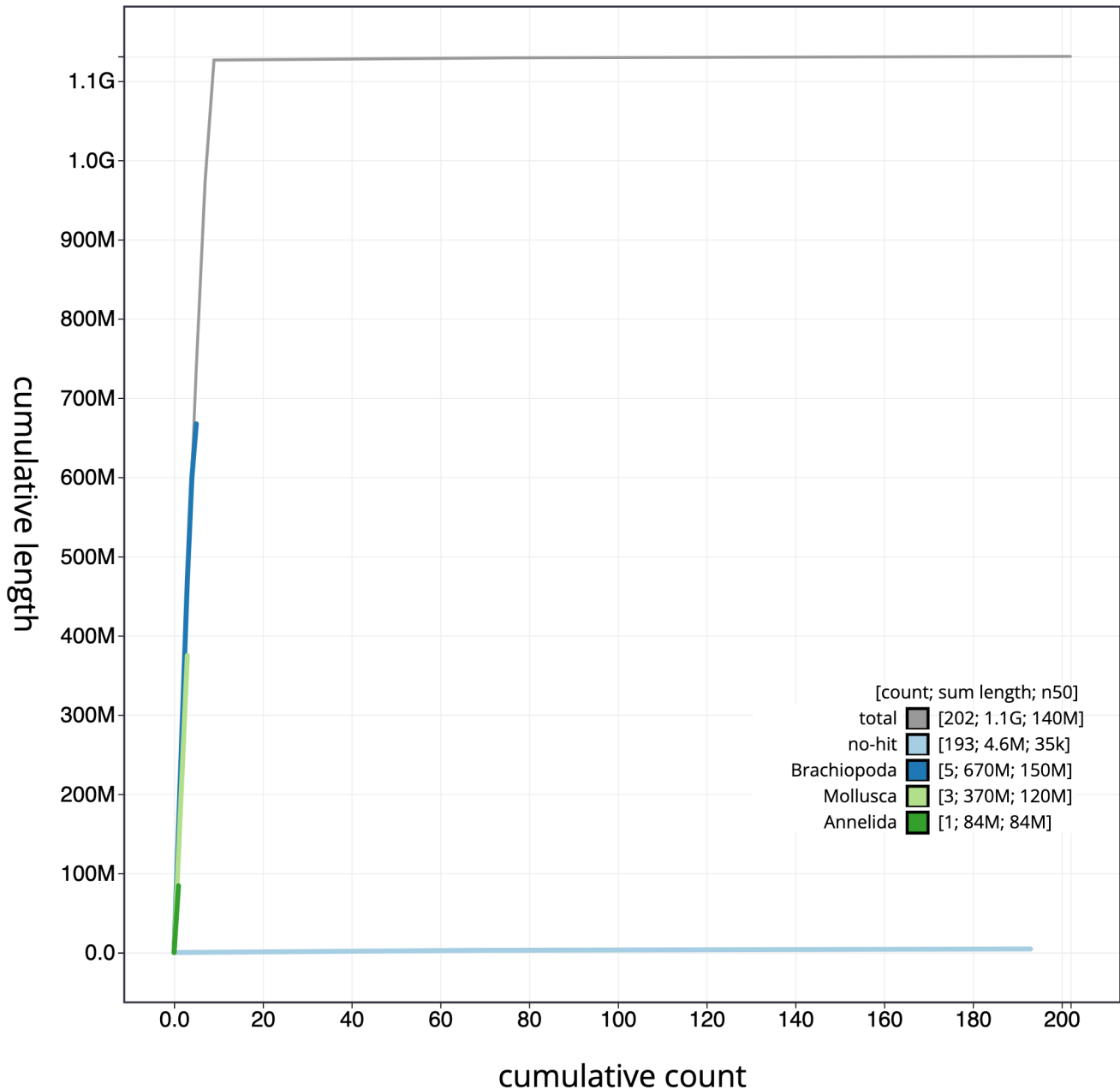
Figure 4. Genome assembly of *Sthenelais limicola*, wpSthLimi1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/wpSthLimi1.1/dataset/CALNXF01/cumulative.

scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2022). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), which performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.
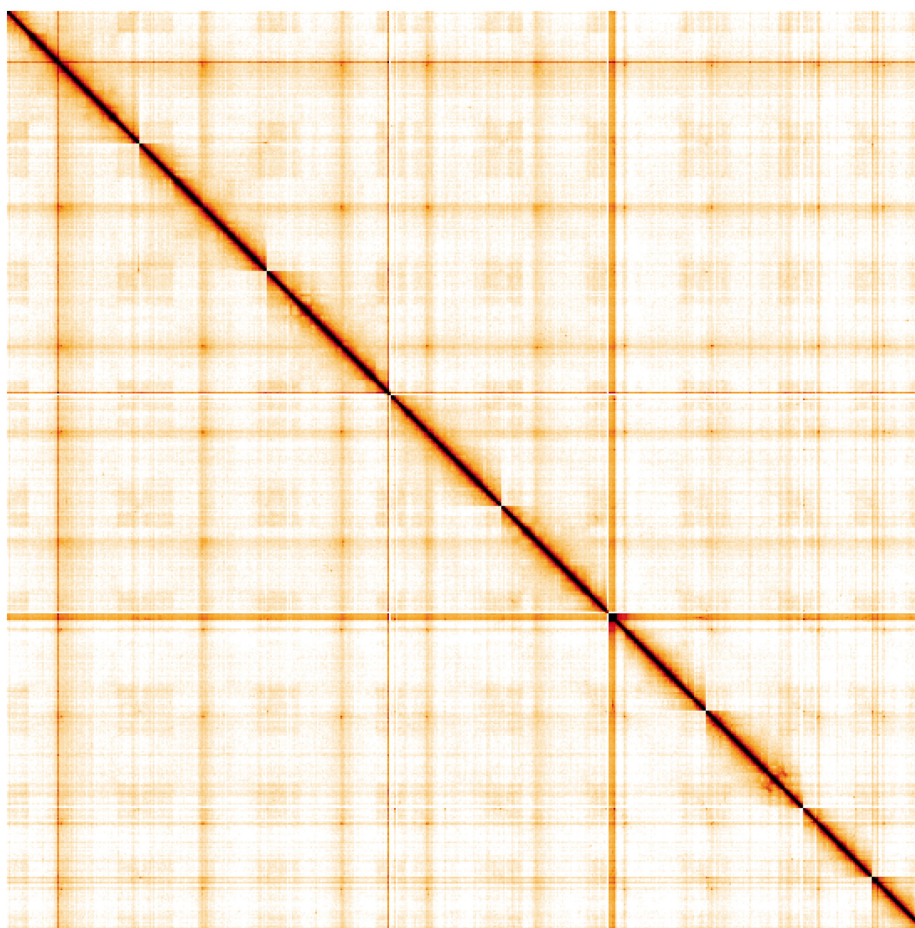
Figure 5. Genome assembly of *Sthenelais limicola*, wpSthLimi1.1: Hi-C contact map. Hi-C contact map of the wpSthLimi1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=X-bn6mmVRHqGemal-zIDZA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Sthenelais limicola*, wpSthLimi1.

INSDC accession	Chromosome	Size (Mb)	GC%
OW804094.1	1	161.88	39
OW804095.1	2	156.73	39
OW804096.1	3	150.5	38.9
OW804097.1	4	136.98	38.8
OW804098.1	5	130.19	38.9
OW804099.1	6	119.44	39.7
OW804100.1	7	118.52	38.8
OW804101.1	8	84.23	38.8
OW804102.1	9	68.09	38.8
OW804103.1	MT	0.02	33.8

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.3.10	Challis *et al.*, 2020
gEVAL	N/A	Chow *et al.*, 2016
Hifiasm	0.16.1	Cheng *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
MitoHiFi	1	Uliano-Silva *et al.*, 2022
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan *et al.*, 2020
YaHS	1.0	Zhou *et al.*, 2022

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission

of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability
Underlying data
European Nucleotide Archive: *Sthenelais limicola* (a segmented worm). Accession number PRJEB51037; https://identifiers.org/ena.embl/PRJEB51037 (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Sthenelais limicola* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data

and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information
Members of the Marine Biological Association Genome Acquisition Lab are listed here:https://doi.org/10.5281/zenodo.4783605.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.4783585.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Barnich R, Fiege D: **The Aphroditoidea (Annelida: Polychaeta) of the Mediterranean Sea.** *Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft,* 2003; **559**: 1–167. **Reference Source**

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Chambers SJ, Muir AI: **Polychaetes: British Chrysopetaloidea, Pisionoidea and Aphroditoidea.** *Synopses of the British Fauna (New Series).* 1997; **54**: 1–202. **Reference Source**

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; (Accessed: 19 October 2022). **Reference Source**

Hartmann-Schröder G: **Annelida, Borstenwürmer, Polychaeta [Annelida, bristleworms, Polychaeta].** 2nd revised edition. The fauna of Germany and adjacent seas with their characteristics and ecology, 58, Jena: Gustav Fischer, 1996. **Reference Source**

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Jumars PA, Dorgan KM, Lindsay SM: **Diet of Worms Emended: An Update of Polychaete Feeding Guilds.** *Ann Rev Mar Sci.* 2015; **7**(1): 497–520. **PubMed Abstract** | **Publisher Full Text**

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Parapar J, Moreira J, Núñez J, *et al.*: **Fauna Ibérica 41: Annelida Polychaeta IV.** Museo Nacional de Ciencias Naturales, 2015. **Reference Source**

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746. **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rouse GW, Pleijel F, Tilic E: **Annelida,** Oxford: Oxford University Press. 2022. **Reference Source**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* 2022; [Preprint]. **Publisher Full Text**

Wellcome Sanger Institute: **The genome sequence of the segmented worm, *Sthenelais limicola.*** (Ehlers, 1864). European Nucleotide Archive, [Dataset]. 2022. **https://identifiers.org/ena.embl/PRJEB51037**

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *bioRxiv.* 2022; [Preprint]. **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ❓ ✔

- -

Version 1

Reviewer Report 17 July 2023

https://doi.org/10.21956/wellcomeopenres.20909.r61558

✔ **Erich Schwarz** (iD)

Cornell University, Ithaca, New York, USA

This report concisely but clearly describes the chromosomally scaffolded genome assembly of a segmented worm first identified in 1864 but not given serious genomic analysis until now. The text is well-written with explicit, well-referenced, and versioned software tools in its methods. The data are publicly available on GenBank for further work.

Notably, the authors did not do predictions of protein-coding genes, but did provide RNA-seq data that would support such prediction by other, interested parties.

Although the authors used automatic defaults for initial assembly of HiFi reads by HiFiAsm and removing excess heterozygous alleles with purge_dup, but put substantial manual effort into making the most accurate Hi-C chromosomal scaffolding that they could. The methods they describe for careful manual correction of Hi-C scaffolding after initial use of YaHS, a recently devised Hi-C scaffolder that may be superior to 3d-dna.

Overall, this report provides the first serious genomic resources for this category of segmented worms, and these resources are well-constructed enough to support future work of high quality.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Assembly, annotation, and biological analysis of nematode genomes: this includes both free-living Caenorhabditis nematodes and parasitic nematodes (e.g., hookworms).

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 10 July 2023

https://doi.org/10.21956/wellcomeopenres.20909.r57955

?   **Torsten H Struck** [ID]
University of Oslo, Oslo, Oslo, Norway

The paper by Darbyshire et al presents the nuclear and mitochondrial genome of an annelid scale worm at the reference level and to my knowledge it is the first for the annelid group Aphroditiformia. It is all in all very well described and has all essential statistics. However, I would like to make a few suggestions, which will help to increase repeatability of the study and all researchers not so familiar with genomics research to use this study as their starting point for conducting their genome projects. Especially, annelid researcher might use such a paper as a starting point. At present, this however not possible as a lot information is lacking describing the bioinformatic parts of this project.

In the description of the genome assembly are the settings for the different programs not given. Moreover, also any custom-made scripts used for the analysis are not mentioned. However, given the different data sources it has to be assumed that these scripts are present. The settings and scripts should be made publicly available to allow repetition of the analyses as the authors did.

Additionally, while the lab methods have been described in sufficient detail, several details on the bioinformatic analyses are missing. Which programs have been used for the mapping, blast/diamond and BUSCO analyses needed for blobtools analysis? BlobTools does not do it by itself. Which programs were used to assess the different quality parameters such as N50, kmer completeness, consensus quality and so forth? Also for the HiC, analyses crucial steps in preparation of the data for YAHS analysis are not described. For some one, new to the field it will very challenging to repeat the analyses easily. If these methods have been published in a different paper the authors should refer to this paper and indicate any deviations from it.

Other points to consider are:
The members of the barcoding team are listed as authors. However, the barcoding approach is not described in the paper at all and also none of its result. Did the barcode confirm the morphological species? This is to be assumed as the policy of DToL is to sequence only species with a matching barcode. Nonetheless a statement should be made in the paper. Moreover,

where is the barcode deposited? No accession number has been provided.

Similarly, the generation of mitochondrial genomes is mentioned in the methods, but only described as "complete single alleles" in Table 1. It is not obvious what this description means? Is the full genome retrieved with different somatic, intraspecific haplotypes or is the genome recovered in several pieces? The accession number is provided for only one entire, non-annotated mitochondrial genome. Hence, what are the other allels?

Additionally, a statement concerning the fulfillment of the EBP standards for chromosome-level assemblies as of today would be nice. This could ne the EBP code of quality, which in this case would be 7.C.Q60.

Finally, "Polychaeta" is a paraphyletic group and hence the name "Polychaeta" should not be used. If the authors want to use the term polychaetes, they should use it as polychaetes and not "Polychaeta". The latter would indicate that it is a proper name and monophyletic group. If the authors want to display the NCBI taxonomy here, the authors should name the section NCBI taxonomy and not Species taxonomy as in this case it is not the latter given all recent phylogenetic studies on Annelida showing the paraphyly of polychaetes.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary Genomics; Phylogenetics; Systematics and Taxonomy of Annelida

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 06 February 2023

https://doi.org/10.21956/wellcomeopenres.20909.r54382

✓   **Joseph Ryan**
    Whitney Laboratory for Marine Bioscience, University of Florida, St Augustine, FL, USA

This article describes the sequencing and assembly of the *Sthenelais limicola* genome. The description of the methods are sound and the analyses are repeatable. The quality metrics are stellar. The manuscript is clearly written and the figures and figure legends are easy to comprehend.

One minor point is the description of the tissue that was sequenced for gDNA and RNA sequencing could be better described. This tissue was described as "*mid-body tissue*." I am guessing that this was a cross section of the animal at one of the middle segments, but this could refer to ectoderm only, gut only, etc. If possible, it would help to clarify (a bonus would be the segment number(s)).

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: evolutionary genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.