DATA NOTE

The genome sequence of the Eurasian river otter, Lutra lutra Linnaeus 1758 [version 1; peer review: 2 approved]

Dan Mead[1], Frank Hailer[2], Elisabeth Chadwick[2], Roberto Portela Miguez[3], Michelle Smith[1], Craig Corton[1], Karen Oliver[1], Jason Skelton[1], Emma Betteridge[1], Jale Doulcan Doulcan[1], Olga Dudchenko[4], Arina Omer[4], David Weisz[4], Erez Lieberman Aiden[4], Shane McCarthy[1], Kerstin Howe[1], Ying Sims[1], James Torrance[1], Alan Tracey[1], Richard Challis[1], Richard Durbin[1], Mark Blaxter[1]

[1]Wellcome Genome Campus, Wellcome Sanger Institute,, Hinxton, CB10 1SA, UK
[2]Cardiff Otter Project, Cardiff University School of Biosciences, Cardiff, CF10 3AX, UK
[3]Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK
[4]Baylor College of Medicine, Houston, Texas, TX 77030 USA, USA

v1 **First published:** 19 Feb 2020, **5**:33 (https://doi.org/10.12688/wellcomeopenres.15722.1)
Latest published: 19 Feb 2020, **5**:33 (https://doi.org/10.12688/wellcomeopenres.15722.1)

Abstract

We present a genome assembly from an individual male *Lutra lutra* (the Eurasian river otter; Vertebrata; Mammalia; Eutheria; Carnivora; Mustelidae). The genome sequence is 2.44 gigabases in span. The majority of the assembly is scaffolded into 20 chromosomal pseudomolecules, with both X and Y sex chromosomes assembled.

Keywords

Lutra lutra river otter genome sequence chromosomal

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 1 19 Feb 2020	✓ report	✓ report

1 **Yibo Hu**, Chinese Academy of Sciences, Beijing, China

2 **Frank Panitz**, Aarhus University, Aarhus, Denmark

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

Corresponding author: Mark Blaxter (mark.blaxter@sanger.ac.uk)

Author roles: Mead D: Conceptualization, Investigation, Writing – Review & Editing; **Hailer F**: Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Chadwick E**: Writing – Review & Editing; **Portela Miguez R**: Data Curation, Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Smith M**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Corton C**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Oliver K**: Formal Analysis, Investigation, Methodology, Supervision, Writing – Review & Editing; **Skelton J**: Writing – Review & Editing; **Betteridge E**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Doulcan JD**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Dudchenko O**: Formal Analysis, Investigation, Methodology, Visualization, Writing – Review & Editing; **Omer A**: Formal Analysis, Investigation, Methodology, Software, Writing – Review & Editing; **Weisz D**: Writing – Review & Editing; **Lieberman Aiden E**: Software, Supervision, Writing – Review & Editing; **McCarthy S**: Data Curation, Formal Analysis, Investigation, Software, Writing – Review & Editing; **Howe K**: Data Curation, Formal Analysis, Investigation, Software, Supervision, Validation, Writing – Review & Editing; **Sims Y**: Data Curation, Formal Analysis, Investigation, Software, Writing – Review & Editing; **Torrance J**: Data Curation, Formal Analysis, Investigation, Software, Supervision, Validation, Writing – Review & Editing; **Tracey A**: Data Curation, Formal Analysis, Investigation, Software, Validation, Writing – Review & Editing; **Challis R**: Data Curation, Formal Analysis, Investigation, Software, Validation, Writing – Review & Editing; **Durbin R**: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – Review & Editing; **Blaxter M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through core funding to the Wellcome Sanger Institute (WT206194). SMcC and RD were supported by Wellcome grant WT207492. MB was supported by Wellcome grant WT218328. ELA was supported by an NSF Physics Frontiers Center Award (PHY1427654), the Welch Foundation (Q-1866), a USDA Agriculture and Food Research Initiative Grant (2017-05741), and an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Mead D, Hailer F, Chadwick E *et al.* **The genome sequence of the Eurasian river otter, Lutra lutra Linnaeus 1758 [version 1; peer review: 2 approved]** Wellcome Open Research 2020, **5**:33 (https://doi.org/10.12688/wellcomeopenres.15722.1)

First published: 19 Feb 2020, **5**:33 (https://doi.org/10.12688/wellcomeopenres.15722.1)

Species taxonomy

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Caniformia; Mustelidae; Lutrinae; Lutra; *Lutra lutra* Linnaeus 1758 (NCBI txid 9657).

Background

The Eurasian river otter, *Lutra lutra*, is found along the coasts and inland waters of Europe, Asia, China, Japan, Java, Sri lanka, the Middle East and North Africa. Eurasia. Throughout Europe, populations of *L. lutra* declined precipitously through the latter half of the 20th century, and the species is of active conservation concern. In Ireland, *L. lutra* populations have remained relatively stable[1], and in Britain river restoration and active intervention have resulted in increased populations, and recolonisation of watersheds from which otters had been eliminated[2]. There is active research of the continuing impacts of pollutants on otters (Pountney *et al.,* 2015), and on the population genetic patterns that have resulted from their near-extinction and subsequent recovery in Britain (Stanton *et al.,* 2014). Here we present a chromosomally assembled genome sequence for *L. lutra*, based on a male specimen from Britain.

[1] Vincent Wildlife Trust https://www.vincentwildlife.ie/species/otter

[2] National Biodiversity network Atlas https://species.nbnatlas.org/species/NBNSYS0000005133#overview

Genome sequence report

The genome was sequenced from a naturally deceased single male *L. lutra* collected by the Cardiff Otter Project from Wincanton, Somerset. A total of 63-fold coverage in Pacific Biosciences single-molecule long reads (N50 24 kb) and 58-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 57 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation HiC data (17-fold coverage). The final assembly has a total length of 2.44 Gb in 43 sequence scaffolds with a scaffold N50 of 149.0 Mb (Table 1). The majority, 92.7%, of the assembly sequence was assigned to 20 chromosomal-level scaffolds representing 18 autosomes (numbered by sequence length), and the X and Y sex chromosomes (Figure 1–Figure 4; Table 2). The assembly has a BUSCO (Simão *et al.,* 2015) completeness

Table 1. Genome data for *Lutra lutra* mLutLut1.

Project accession data	
Assembly identifier	mLutLut1
Species	*Lutra lutra*
Specimen	NHMUK ZD 2019.215
NCBI taxonomy ID	9657
BioProject	PRJEB35340
Biosample ID	SAMEA994731
Isolate information	Wild casualty; male
Raw data accessions	
PacificBiosciences SEQUEL I	ERR3313238, ERR3313239-ERR3313241, ERR3313246, ERR3313327, ERR3313330, ERR3313333-ERR3313341
10X Genomics Illumina	ERR3316145-ERR3316148, ERR3316169-ERR3316171
Hi-C Illumina	SRR10119468
Genome assembly	
Assembly accession	GCA_902655055.1
Accession of alternate haplotype	GCA_902653095.1
Span (Mb)	2,438.00
Number of contigs	228
Contig N50 length (Mb)	30.40
Number of scaffolds	43
Scaffold N50 length (Mb)	149.00
Longest scaffold (Mb)	223.45
BUSCO* genome score	C:95.8%[S:94.3%,D:1.5%],F:1.9%,M:2.3%,n:4104

*BUSCO scores based on the mammalia_odb9 BUSCO set using v3.0.2. 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/mLutLut1_1/dataset/mLutLut1_1/busco.

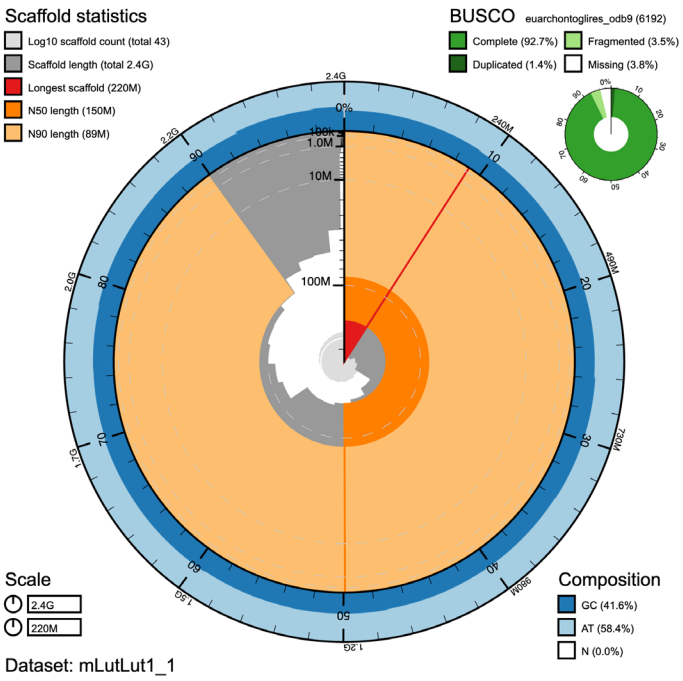

Figure 1. Genome assembly of *Lutra lutra* mLutLut1: BlobToolKit Snailplot. The plot shows N50 metrics for *L. lutra* assembly mLutLut1 and BUSCO scores for the Euarchontoglires set of orthologues. The interactive version of this figure is hosted here.

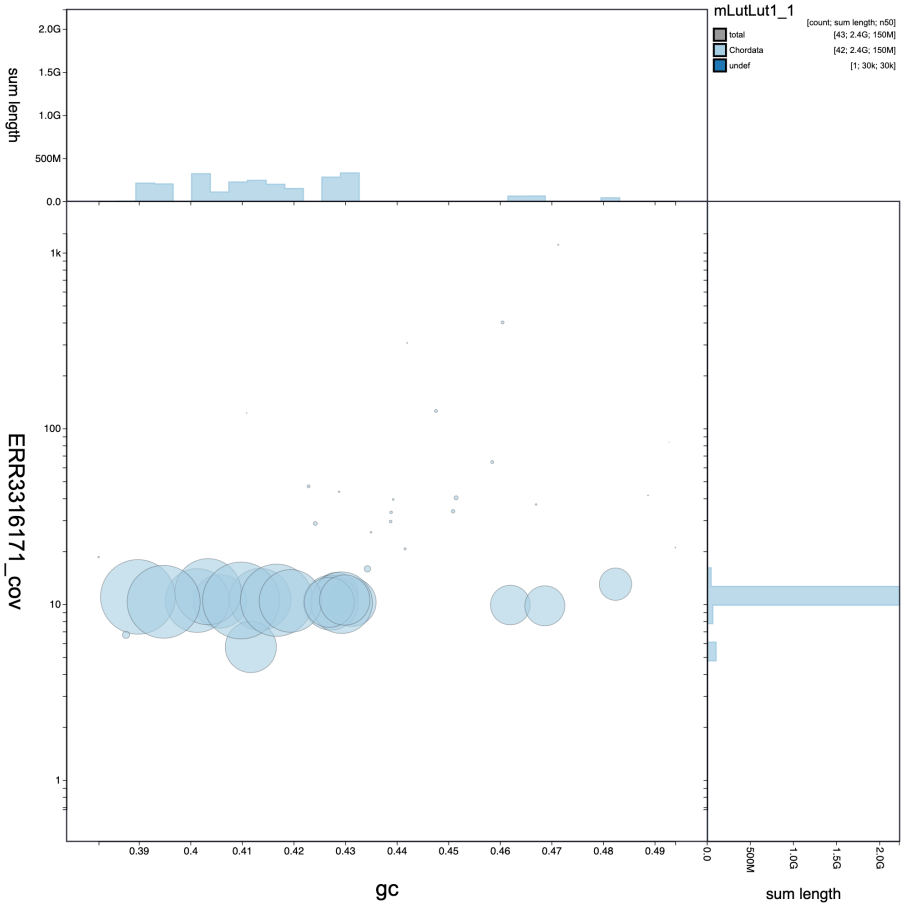

Figure 2. Genome assembly of *Lutra lutra* mLutLut1: BlobToolKit GC-coverage plot. The interactive version of this figure is hosted here.

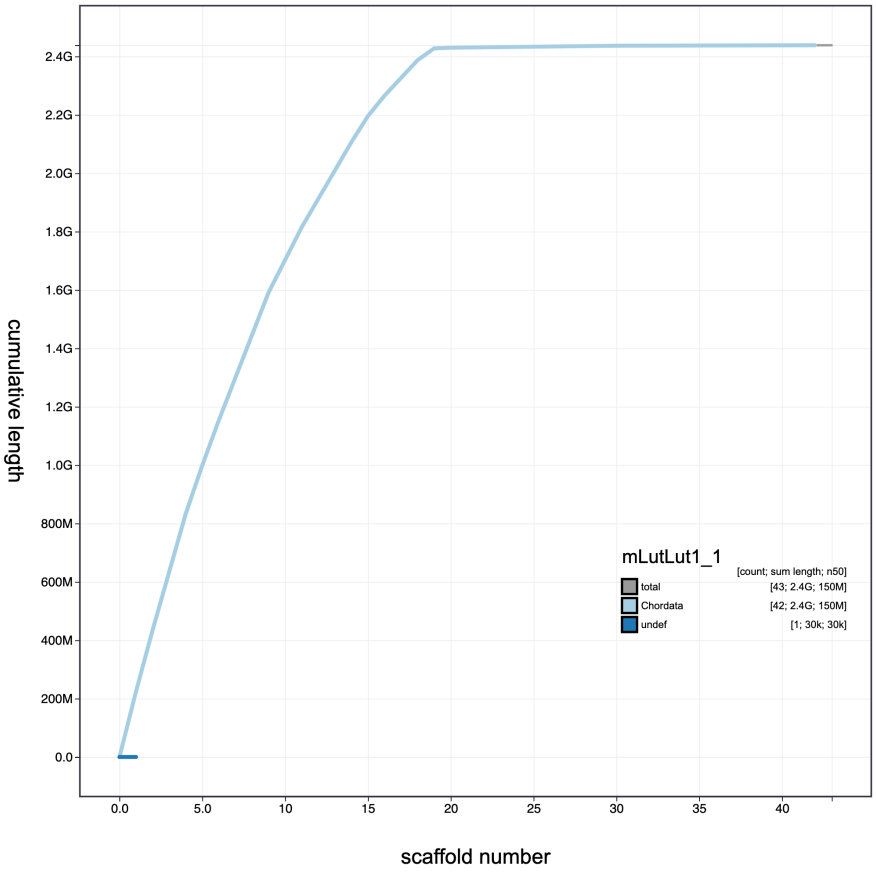

Figure 3. Genome assembly of *Lutra lutra* mLutLut1: BlobToolKit Cumulative sequence plot. The interactive version of this figure is hosted here.

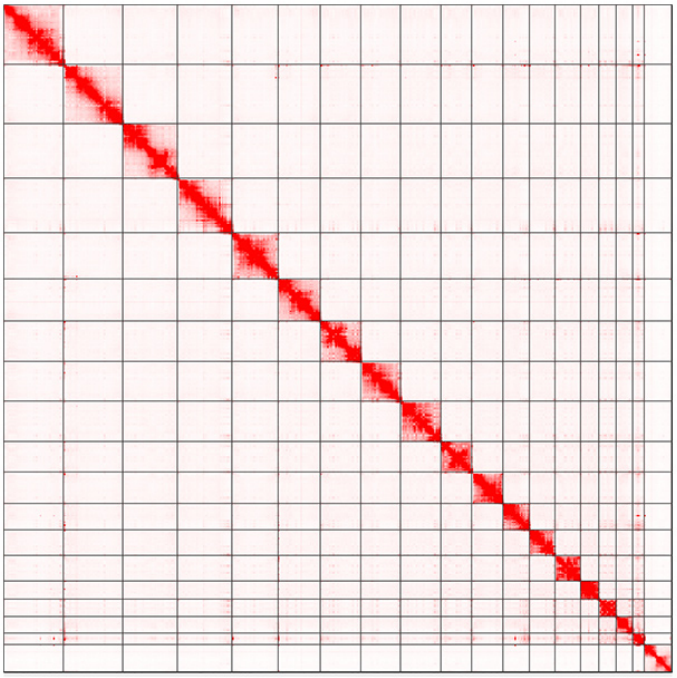

Figure 4. Genome assembly of *Lutra lutra* mLutLut1: Hi-C contact map. Hi-C contact map of the L. lutra mLutLut1 assembly, visualized in Juicebox (Durand *et al.*, 2016). An interactive version of the map hosted here, powered by Juicebox.js (Robinson *et al.*, 2018).

of 95.8% using the mammalia_odb9 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Lutra lutra* mLutLut1.

ENA accession	*Chromosome*	*Size (Mb)*	*GC%*
LR738403.1	1	223.45	41
LR738404.1	2	210.65	39
LR738405.1	3	201.32	39.5
LR738406.1	4	197.71	41.7
LR738407.1	5	165.81	40.3
LR738408.1	6	154.43	40.1
LR738409.1	7	149.01	41.9
LR738410.1	8	144.75	41.3
LR738411.1	9	144.09	42.9
LR738412.1	10	114.66	42.7
LR738413.1	11	108.79	40.6
LR738414.1	12	96.45	43
LR738415.1	13	95.73	42.7
LR738416.1	14	89.08	43.1
LR738417.1	15	69.99	42.8
LR738418.1	16	61.48	46.9
LR738419.1	17	60.35	46.2
LR738420.1	18	40.43	48.2
LR738421.1	X	99.69	41.2
LR738422.1	Y	2.25	38.8

Methods

The river otter specimen was collected from Wincanton, Somerset by the Cardiff Otter Project. A full tissue dissection and preservation in 80% ethanol was undertaken and the specimen accessioned by the Natural History Museum, London.

DNA was extracted using an agarose plug extraction from spleen tissue following the Bionano Prep Animal Tissue DNA Isolation Soft Tissue Protocol. Pacific Biosciences CLR long read and 10X Genomics read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL I and Illumina HiSeq X instruments. Hi-C data were generated by the Aiden lab using an optimised version of their protocols (Dudchenko *et al.*, 2017).

Assembly was carried out using Falcon-unzip (Chin *et al.*, 2016), haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020) and a first round of scaffolding carried out with 10X Genomics read clouds using scaff10x (https://github.com/wtsi-hpag/Scaff10X). Scaffolding with Hi-C data (Rao *et al.*, 2014) was carried out with 3D-DNA (Dudchenko *et al.*, 2017), followed by manual curation with Juicebox Assembly Tools (Dudchenko *et al.*, 2018; Durand *et al.*, 2016; Robinson *et al.*, 2018) and visualisation in HiGlass (Kerpedjiev *et al.*, 2018). The Hi-C scaffolded assembly was polished with arrow using the PacBio data, then polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012) and applying homozygous non-reference edits using bcftools consensus (https://github.com/VGP/vgp-assembly/tree/master/pipeline/freebayes-polish). Two rounds of the Illumina polishing were applied. The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016). We removed two

Table 3. Software tools used.

Software tool	Version	Source
Falcon-unzip	falcon-kit 1.2.2	(Chin *et al.*, 2016)
purge_dups	1.0.0	(Guan *et al.*, 2020)
3D-DNA	180419	(Dudchenko *et al.*, 2018)
scaff10x	4.2	https://github.com/wtsi-hpag/Scaff10X
arrow	GenomicConsensus 2.3.3	https://github.com/PacificBiosciences/GenomicConsensus
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	v1.1.0-3-g961e5f3	(Garrison & Marth, 2012)
bcftools consensus	1.9	http://samtools.github.io/bcftools/bcftools.html
gEVAL	2016	(Chow *et al.*, 2016)
BlobToolKit	1	(Challis *et al.*, 2019)

low-coverage scaffolds that were likely to have derived from the ribosomal DNA cistron of a *Sarcocystis* species (most similar to *Sarcocystis lutrae*). The genome was analysed within the BlobToolKit environment (Challis *et al.,* 2019).

Data availability

European Nucleotide Archive: Lutra lutra (Eurasian otter) genome assembly, mLutLut1. BioProject accession number PRJEB35340; https://www.ebi.ac.uk/ena/data/view/PRJEB35340.

The genome sequence is released openly for reuse. The *L. lutra* genome sequencing initiative is part of the Wellcome Sanger Institute's "25 genomes for 25 years" project[3]. It is also part of the Vertebrate Genome Project (VGP)[4] ordinal references programme, the DNA Zoo Project[5] and the Darwin Tree of Life (DToL) project[6]. The specimen has been preserved in ethanol and deposited with the Natural History Museum, London under registration number NHMUK ZD 2019.215 where it will remain accessible to the research community for posterity. All raw data and the assembly have been deposited in the ENA. The genome will be annotated and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Acknowledgements

We thank Mike Stratton and Julia Wilson for their continuing support for the 25 genomes for 25 years project.

[3] https://www.sanger.ac.uk/science/collaboration/25-genomes-25-years

[4] https://vertebrategenomesproject.org/

[5] https://www.dnazoo.org/

[6] https://www.darwintreeoflife.org/

References

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – Interactive Quality Assessment of Genome Assemblies.** *bioRxiv.* 2019.
Publisher Full Text

Chin CS, Peluso P, Sedlazeck FJ, *et al.*: **Phased diploid genome assembly with single-molecule real-time sequencing.** *Nat Methods.* 2016; **13**(12): 1050–54.
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): 2508–10.
PubMed Abstract | Publisher Full Text | Free Full Text

Dudchenko O, Batra SS, Omer AD, *et al.*: ***De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.** *Science.* 2017; **356**(6333): 92–95.
PubMed Abstract | Publisher Full Text | Free Full Text

Dudchenko O, Shamim MS, Batra SS, *et al.*: **The Juicebox Assembly Tools Module Facilitates *de Novo* Assembly of Mammalian Genomes with Chromosome-Length Scaffolds for under $1000.** *bioRxiv.* 2018.
Publisher Full Text

Durand NC, Robinson JT, Shamim MS, *et al.*: **Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.** *Cell Syst.* 2016; **3**(1): 99–101.
PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** *arXiv [q-bio.GN].* arXiv. 2012.
Reference Source

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; pii: btaa025.
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

Pountney A, Filby AL, Thomas GO, *et al.*: **High liver content of polybrominated diphenyl ether (PBDE) in otters (Lutra lutra) from England and Wales.** *Chemosphere.* 2015; **118**: 81–86.
PubMed Abstract | Publisher Full Text

Rao SS, 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): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Robinson JT, Turner D, Durand NC, *et al.*: **Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data.** *Cell Syst.* 2018; **6**(2): 256–58.e1.
PubMed Abstract | Publisher Full Text | Free Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

Stanton DWG, Hobbs GI, McCafferty DJ, *et al.*: **Contrasting Genetic Structure of the Eurasian Otter (*Lutra Lutra*) across a Latitudinal Divide.** *J Mammal.* 2014; **95**(4): 814–823.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 17 March 2020

https://doi.org/10.21956/wellcomeopenres.17235.r37966

✔ **Frank Panitz**
Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

The manuscript describes the *de novo* genome assembly of the Eurasian river otter (*Lutra lutra*). The authors combine the latest state-of-the art sequencing methods, PacificBioscience and 10X Genomics, to generate primary contigs which are then assembled into 43 scaffolds, having a high N50 value of 149 Mb, using Hi-C data. In addition, 92.7% of the total assembly are assigned to chromosomal pseudomolecules, representing 18 autosomes as well as X and Y sex chromosomes. Together with a high level of completeness as analysed by BUSCO quality assessment, this *de novo* assembly provides a high-quality draft genome sequence.

The genome resources generated in this project will be instrumental to investigate the genetic structure, genetic diversity and phylogeny of Eurasian otters. As monitoring tools are urgently needed for development and evaluation of conservation efforts the genome sequence will provide genetic markers to be used in conservation genetic studies.

The manuscript is concise, data is presented comprehensively and the figures are linked to interactive versions

Recommendation:
The Busco software used for genome quality assessment should to be included in Table 3.

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: Genome and transcriptome assembly, NGS technologies, molecular genetics.

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 05 March 2020

https://doi.org/10.21956/wellcomeopenres.17235.r37967

Yibo Hu
CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

This manuscript described the detailed information about the chromosome-level *de novo* genome assembly of the Eurasian river otter. They used PacBio SMRT and 10X Genomics sequencing techniques to construct the scaffolds of the genome. These two techniques are widely popular sequencing methods that are good at constructing long scaffolds. They assembled as low as 43 scaffolds. Furthermore, they used Hi-C technique to assign the scaffolds on chromosomes. As a result, 92.7% of the assembly sequences were assigned to 20 chromosomal pseudomolecules. The levels of contig N50 and scaffold N50, and BUSCO genome assembly assessment showed that the genome assembly is of high-quality.

The Eurasian river otter is listed as Near Threatened under the IUCN red list, and some populations have been decreasing. So, it is important to know the genetic diversity, genetic structure and adaptive evolution mechanisms of Eurasian river otters. The chromosome-level genome assembly will help answer the above questions, which has important conservation implications for this species.

The report about the genome assembly was detailed, and the manuscript written well. I just have two comments.

1. I suggest to add a photo of Eurasian river otter to show this species to readers.

2. Table 2: it is better to add the number of scaffolds assigned for each chromosome.

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: Conservation genetics and genomics of endangered mammals.

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.