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

Dan Mead¹, Frank Hailer², Elisabeth Chadwick², Roberto Portela Miguez³, Michelle Smith¹, Craig Corton¹, Karen Oliver¹, Jason Skelton¹, Emma Betteridge¹, Jale Doulcan Doulcan¹, Olga Dudchenko⁴, Arina Omer⁴, David Weisz⁴, Erez Lieberman Aiden⁴, Shane McCarthy¹, Kerstin Howe¹, Ying Sims¹, James Torrance¹, Alan Tracey¹, Richard Challis¹, Richard Durbin¹, Mark Blaxter¹

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

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
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;
Doukan 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;
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.

Copyright: © 2020 Mead D et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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: awaiting peer review] 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)

© 2020 Mead D et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
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, and in Britain river restoration and active intervention have resulted in increased populations, and recolonisation of watersheds from which otters had been eliminated. 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.

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 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.

---

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

<table>
<thead>
<tr>
<th><strong>Project accession data</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly identifier</td>
<td>mLutLut1</td>
</tr>
<tr>
<td>Species</td>
<td><em>Lutra lutra</em></td>
</tr>
<tr>
<td>Specimen</td>
<td>NHMUK ZD 2019.215</td>
</tr>
<tr>
<td>NCBI taxonomy ID</td>
<td>9657</td>
</tr>
<tr>
<td>BioProject</td>
<td>PRJEB35340</td>
</tr>
<tr>
<td>Biosample ID</td>
<td>SAMEA994731</td>
</tr>
<tr>
<td>Isolate information</td>
<td>Wild casualty; male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Raw data accessions</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PacificBiosciences SEQUEL I ERR3313238, ERR3313239-ERR3313241, ERR3313246, ERR3313327, ERR3313330, ERR3313333-ERR3313341</td>
<td></td>
</tr>
<tr>
<td>10X Genomics Illumina ERR3316145-ERR3316148, ERR3316169-ERR3316171</td>
<td></td>
</tr>
<tr>
<td>Hi-C Illumina SRR10119468</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Genome assembly</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly accession</td>
<td>GCA_902655055.1</td>
</tr>
<tr>
<td>Accession of alternate haplotype</td>
<td>GCA_902653095.1</td>
</tr>
<tr>
<td>Span (Mb)</td>
<td>2,438.00</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>228</td>
</tr>
<tr>
<td>Contig N50 length (Mb)</td>
<td>30.40</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>43</td>
</tr>
<tr>
<td>Scaffolding N50 length (Mb)</td>
<td>149.00</td>
</tr>
<tr>
<td>Longest scaffold (Mb)</td>
<td>223.45</td>
</tr>
<tr>
<td>BUSCO* genome score</td>
<td>C:95.8%, S:94.3%, D:1.5%, F:1.9%, M:2.3%, n:4104</td>
</tr>
</tbody>
</table>

* 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.genomeshubs.org/view/mLutLut1_1/dataset/mLutLut1_1/busc](https://blobtoolkit.genomeshubs.org/view/mLutLut1_1/dataset/mLutLut1_1/busc).

---

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

National Biodiversity network Atlas [https://species.nbnatlas.org/species/NBNSYS0000005133#overview](https://species.nbnatlas.org/species/NBNSYS0000005133#overview)
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 3. Software tools used.

<table>
<thead>
<tr>
<th>Software tool</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falcon-unzip</td>
<td>falcon-kit 1.2.2</td>
<td>(Chin et al., 2016)</td>
</tr>
<tr>
<td>purge_dups</td>
<td>1.0.0</td>
<td>(Guan et al., 2020)</td>
</tr>
<tr>
<td>3D-DNA</td>
<td>180419</td>
<td>(Dudchenko et al., 2018)</td>
</tr>
<tr>
<td>scaffold10x</td>
<td>4.2</td>
<td><a href="https://github.com/wtsi-hpag/Scaff10X">https://github.com/wtsi-hpag/Scaff10X</a></td>
</tr>
<tr>
<td>arrow</td>
<td>GenomicConsensus 2.3.3</td>
<td><a href="https://github.com/PacificBiosciences/GenomicConsensus">https://github.com/PacificBiosciences/GenomicConsensus</a></td>
</tr>
<tr>
<td>longranger align</td>
<td>2.2.2</td>
<td><a href="https://support.10xgenomics.com/genome-exome/software/pipelines/latest/other-pipelines">https://support.10xgenomics.com/genome-exome/software/pipelines/latest/other-pipelines</a></td>
</tr>
<tr>
<td>freebayes</td>
<td>v1.1.0-3-g961e5f3</td>
<td>(Garrison &amp; Marth, 2012)</td>
</tr>
<tr>
<td>bcftools consensus</td>
<td>1.9</td>
<td><a href="http://samtools.github.io/bcftools/bcftools.html">http://samtools.github.io/bcftools/bcftools.html</a></td>
</tr>
<tr>
<td>gEVAL</td>
<td>2016</td>
<td>(Chow et al., 2016)</td>
</tr>
<tr>
<td>BlobToolKit</td>
<td>1</td>
<td>(Challis et al., 2019)</td>
</tr>
</tbody>
</table>

### 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 scaffold10x (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
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. It is also part of the Vertebrate Genome Project (VGP) ordinal references programme, the DNA Zoo Project and the Darwin Tree of Life (DToL) project. 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

---

**References**


