





DATA NOTE

The genome sequence of the European badger, *Meles meles* (Linnaeus, 1758) [version 1; peer review: 1 approved with reservations]

Chris Newman ^{1,2}, Ming-shan Tsai¹, Christina D. Buesching¹⁻³, Peter W. H. Holland ¹, David W. Macdonald¹, Darwin Tree of Life Consortium, University of Oxford and Wytham Woods Genome Acquisition Lab, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective

¹Department of Biology, University of Oxford, Oxford, UK

²Cook's Lake Farming Forestry and Wildlife Inc (Ecological Consultancy), Queens County, Nova Scotia, Canada

³Department of Biology, Irving K. Barber School of Arts and Sciences Unit 2, University of British Columbia, Kelowna, British Columbia, Canada

V1 First published: 23 Sep 2022, 7:239
<https://doi.org/10.12688/wellcomeopenres.18230.1>
Latest published: 23 Sep 2022, 7:239
<https://doi.org/10.12688/wellcomeopenres.18230.1>

Abstract

We present a haplotype resolved, diploid genome assembly from a male *Meles meles* (European badger; Chordata; Mammalia; Carnivora; Mustelidae) using the trio binning approach. The genome sequence is 2,739 megabases in span. The majority of the assembly (95.16%) is scaffolded into 23 chromosomal pseudomolecules with the X and Y sex chromosomes assembled. The complete mitochondrial genome was also assembled and is 16.4 kilobases in length.

Keywords



Meles meles, European badger, genome sequence, chromosomal, Chordata



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status ?

1	
<hr/>	
version 1	 ?
23 Sep 2022	view
<hr/>	
1. Richard Edwards  , University of Western Australia, Perth, Australia	
University of New South Wales, Sydney, Australia	

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: **Newman C:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Tsai Ms:** Investigation, Resources, Writing – Review & Editing; **Buesching CD:** Investigation, Resources, Writing – Review & Editing; **Holland PWH:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Macdonald DW:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Newman C *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: Newman C, Tsai Ms, Buesching CD *et al.* **The genome sequence of the European badger, *Meles meles* (Linnaeus, 1758) [version 1; peer review: 1 approved with reservations]** Wellcome Open Research 2022, 7:239 <https://doi.org/10.12688/wellcomeopenres.18230.1>

First published: 23 Sep 2022, 7:239 <https://doi.org/10.12688/wellcomeopenres.18230.1>

Species taxonomy

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Caniformia; Mustelidae; Melinae; *Meles*; *Meles meles* (Linnaeus, 1758) (NCBI:txid9662).

Background

The European badger *Meles meles* is a stocky and powerfully built mammal in the family Mustelidae, which also includes weasels, stoats, minks, otters and martens. Historically, the genus *Meles* was considered monotypic, but several recent authors have argued for division into at least four species, with *Meles meles* representing badgers found across Europe including the UK and Ireland (Koepfli *et al.*, 2018). Badgers live in complex social groups and are omnivores with a diet dominated by earthworms supplemented by slugs, snails, fruit and occasionally small mammals. Despite their predominantly nocturnal lifestyle, badgers are amongst the most widely recognized of all European mammals and have featured extensively in literature, appearing in books by Kenneth Grahame, C.S. Lewis, Henry Williamson, Roald Dahl, Denys Watkins-Pitchford, Beatrix Potter and others. After centuries of persecution, badgers now have extensive protection in law in the UK (UK Public General Acts, “Protection of Badgers Act 1992”); however, badgers still face threats from road collisions, sett disturbance due to land use change and housing development, and a controversial programme of legal culling designed to limit the spread of bovine tuberculosis of which badgers are a vector (Donnelly *et al.*, 2006; Ham *et al.*, 2019; Jenkins *et al.*, 2010; Woolhouse & Wood, 2013).

The badger population in Wytham Woods, Oxfordshire, UK, has been studied in unprecedented detail since 1987, with over 70% of the extant population trapped and released each year (Bright Ross *et al.*, 2020; Noonan *et al.*, 2015). On capture, a range of biometric measurements and samples are taken routinely, providing a large and long-term data set which has given insights into demography, behaviour, climate change responses, energetics, reproductive biology, epidemiology, immunology and disease (Macdonald *et al.*, 2015a; Macdonald & Newman, 2022). By its conclusion in 2019, the survey had generated a database comprising 11,488 capture records with associated samples and data, from 1,823 individuals.

A male European badger from Wytham Woods, individual 1581, was selected for genome sequencing as part of the Darwin Tree of Life Project, a collaborative effort to sequence all of the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for this individual *M. meles*. This individual was chosen to facilitate use of the trio-based assembly method, whereby lower coverage genome sequence of each parent is used to help distinguish haplotypes in the higher quality genome sequence determined from the target individual (Rhie *et al.*, 2021). Individual 1581 (mMelMel3) was born in spring 2015 in a large social group called Pasticks. After initial recording, he was caught and recorded on 23 further occasions, with this blood sample taken for DNA extraction and

sequencing in November 2019. The parents (mother 1365, mMelMel2; father 999, mMelMel1) were also born in the Pasticks group: 1365 was born in 2010 and was re-trapped on 27 occasions, 999 was born in 2005 and was caught on 26 occasions prior to his last recorded data entry in November 2016. Individual 1581 has no siblings or half-siblings recorded in the Wytham Woods pedigree; however, it is possible for cubs to die prior to it being safe to catch and mark them.

A high-quality badger genome sequence will facilitate further genetic studies into ecology, population biology and evolution of badgers and related mammals. Analysis of the genome sequence is also likely to give insights into unusual features of badger biology such as delayed implantation and embryonic diapause (Sugianto *et al.*, 2021; Yamaguchi *et al.*, 2006). Most pressingly, the badger genome sequence will be useful in efforts to understand the role that badgers, and their immune system, play in the epizootiology of bovine tuberculosis (Bilham *et al.*, 2017; Macdonald *et al.*, 2015b).

Genome sequence report

The genome was sequenced as a trio assembly, using blood samples collected from three *M. meles* individuals; an adult male (mMelMel1, 999), an adult female (mMelMel2, 1365), and their male F1 offspring (mMelMel3, 1581). The samples were collected from Wytham Woods, Oxford, UK. An artistic impression of *M. meles*, generated using data from this genome sequence, can be seen in Figure 1.

A total of 39-fold coverage in Pacific Biosciences single-molecule HiFi long reads were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation of the paternal haplotype corrected 100 missing/misjoins and removed 2 haplotypic duplications, reducing the assembly size by 2.89% and the scaffold number by 12.82%, and increasing the scaffold N50 by 20.89%.

The final assembly has a total length of 2,739 Mb in 537 sequence scaffolds with a scaffold N50 of 132.9 Mb (Table 1).

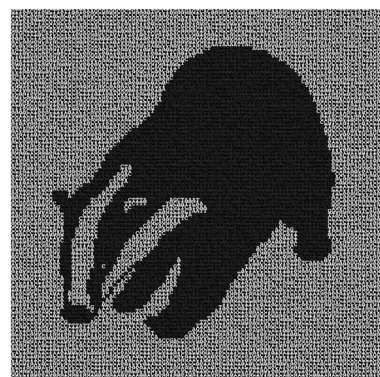


Figure 1. Artistic impression of a European badger, *Meles meles*. This image was generated by Mark Blaxter using genome sequence data obtained from this assembly.

Table 1. Genome data for *Meles meles*, mMelMel3.1.

Project accession data	
Assembly identifier	mMelMel3.2
Species	<i>Meles meles</i>
Specimen	mMelMel3 (genome assembly, Hi-C); mMelMel1 and mMelMel2 (genomic data)
NCBI taxonomy ID	9662
BioProject	PRJEB46333
BioSample ID	SAMEA7524400
Isolate information	Blood samples (mMelMel1, 999; mMelMel2, 1365; mMelMel3, 1581)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6808015-ERR6808016; ERR6939248-ERR6939250
10X Genomics Illumina	ERR6688595-ERR6688598; ERR6688601-ERR6688604
Hi-C Illumina	ERR6688402
Illumina maternal and paternal libraries	ERR6688599 (mMelMel1) ERR6688600 (mMelMel2)
Genome assembly	
Assembly accession (paternal)	GCA_922984935.2
Accession of alternate haplotype (maternal)	GCA_922990625.1
Span (Mb)	2,739
Number of contigs	638
Contig N50 length (Mb)	75.2
Number of scaffolds	537
Scaffold N50 length (Mb)	132.9
Longest scaffold (Mb)	216.97
BUSCO* genome score (paternal haplotype)	C:95.1%[S:93.2%,D:2.0%],F:0.8%,M:4.0%,n:14,502

*BUSCO scores based on the *carnivora_odb10* BUSCO set using v5.1.2. 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/mMelMel3.1%20paternal%20haplotype/dataset/CAKLPM01/busco>.

The majority, 95.16%, of the assembly sequence was assigned to 23 chromosomal-level scaffolds, representing 21 autosomes (numbered by sequence length) and the X and Y sex chromosomes (Figure 2–Figure 5; Table 2).

The paternal assembly has a BUSCO v5.1.2 (Manni *et al.*, 2021) completeness of 95.1% (single 93.2%, duplicated 2.0%) using the *carnivora_odb10* reference set (n=14,502). The assembly deposited is of the paternal haplotype with the X chromosome from the maternal haplotype (mMelMel3.2).

Contigs corresponding to the maternal haplotype have also been deposited (mMelMel3.1).

Methods

Sample acquisition and nucleic acid extraction

Blood samples were collected from three *M. meles* individuals; an adult male (mMelMel1, 999), an adult female (mMelMel2, 1365), and their male offspring (mMelMel3, 1581). The samples were collected from Wytham Woods, Oxford, UK (latitude 51.77, longitude -1.34) by Chris Newman, Ming-shan

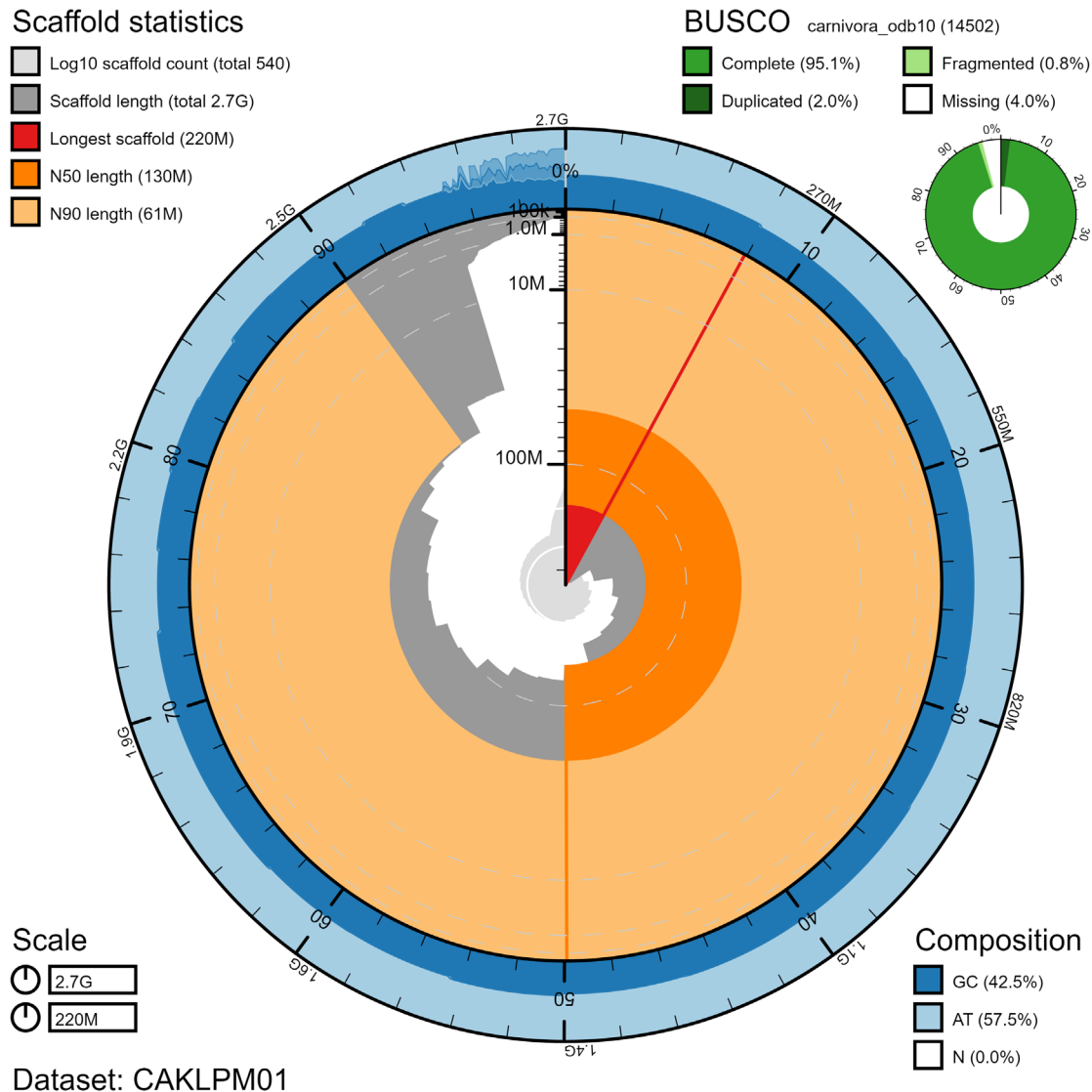


Figure 2. Genome assembly of *Meles meles* (paternal haplotype), mMelMel3.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 2,738,678,004 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (216,965,501 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (132,955,750 and 61,486,999 bp), respectively. The pale grey spiral shows the cumulative chromosome 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 carnivora_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/mMelMel3.1%20paternal%20haplotype/dataset/CAKLPM01/snail>.

Tsai, David Macdonald and Peter Holland (all University of Oxford) and Christina Buesching (University of British Columbia, Canada). The specimens were identified by Chris Newman and Ming-shan Tsai.

The badgers were trapped using 80x40x40 cm cage traps with a string trigger under licence from Natural England (under the Protection of Badgers Act 1992); research was conducted under the Animals (Scientific Procedures) Act, 1986. Upon capture, badgers were sedated with ketamine hydrochloride

(McLaren *et al.*, 2005; Sugianto *et al.*, 2019) allowing a regime of biometric measurements and sampling to be undertaken, including blood sampling by jugular venipuncture, before release. Blood for genome sequencing was collected into EDTA-treated tubes and stored at -80 C before shipping on dry ice to the Wellcome Sanger Institute.

Blood and hair follicle samples also provided the basis of a genetic pedigree for the population (Annavi *et al.*, 2014a; Dugdale *et al.*, 2007), based on polymorphic microsatellite

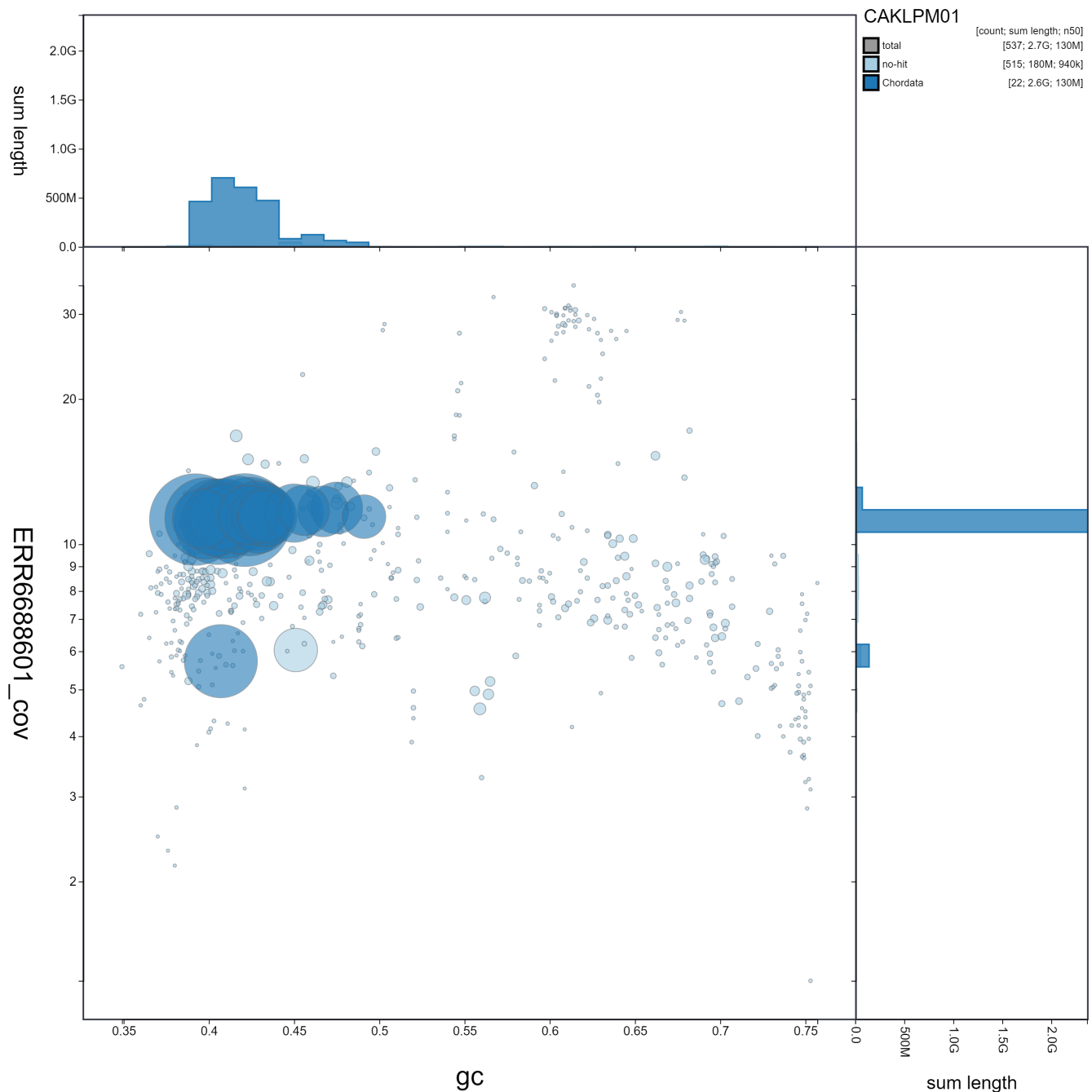


Figure 3. Genome assembly of *Meles meles* (paternal haplotype), mMelMel3.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/mMelMel3.1%20paternal%20haplotype/dataset/CAKLPM01/blob>.

markers (Annavi *et al.*, 2011), which revealed individual 1581 (mMelMel3) to be the offspring of individuals 1365 (mMelMel2) and 999 (mMelMel1). Parentage assignment is challenging with badgers, confounded by high rates of extra-group paternity, half-siblings derived from multiple litter paternity and inter-annual variation in mating pairs. Parentage

was determined sequentially using MasterBayes and Colony analyses (Annavi *et al.*, 2014b).

High molecular weight (HMW) DNA was extracted at the WSI Scientific Operations Core from frozen whole blood samples of mMelMel1, mMelMel2 and mMelMel3 (genome

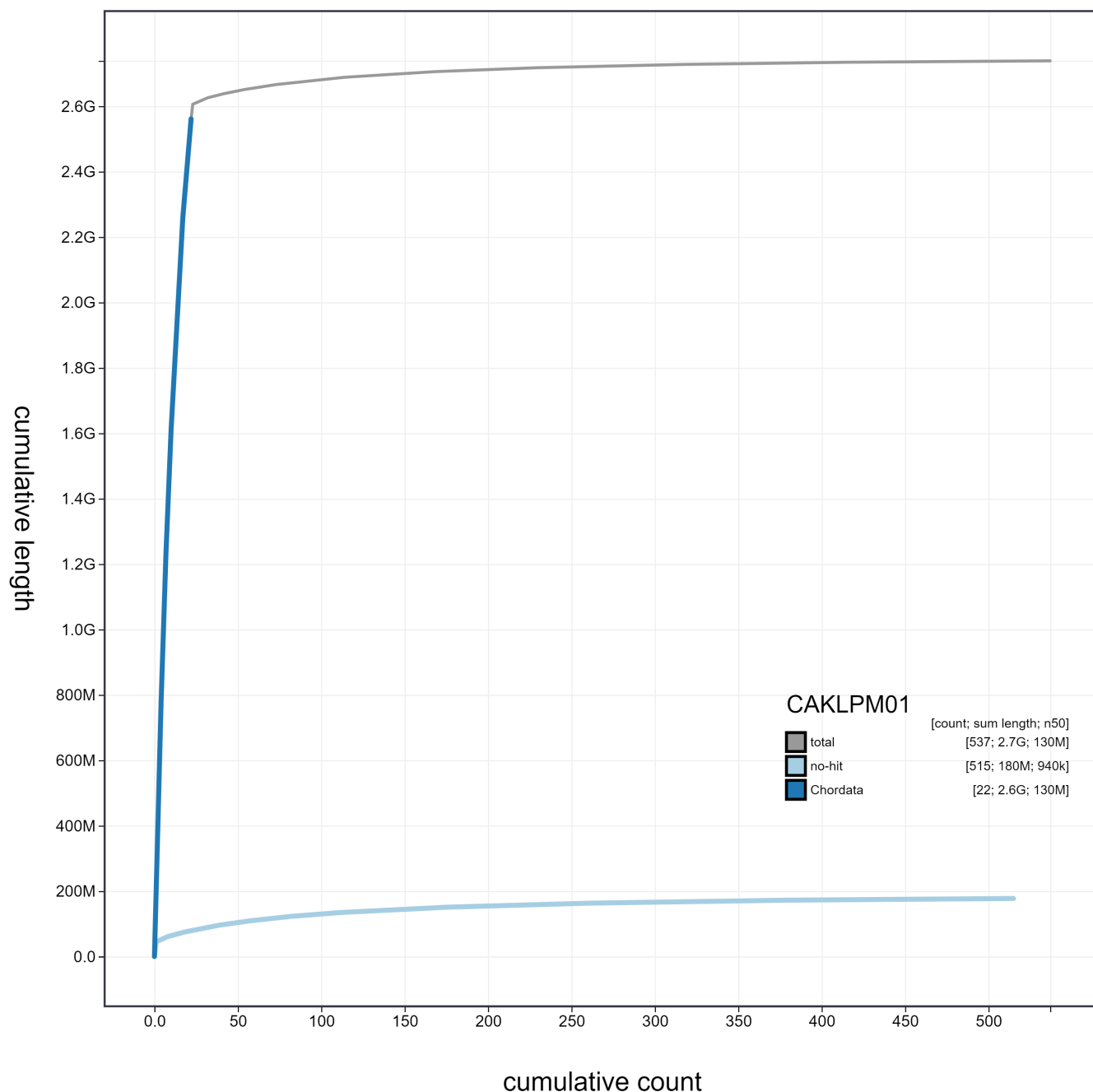


Figure 4. Genome assembly of *Meles meles* (paternal haplotype), mMelMel3.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/mMelMel3.1%20paternal%20haplotype/dataset/CAKLPM01/cumulative>.

assembly, Hi-C). The blood was warmed and agitated and the red blood cells lysed using Qiagen's RBC lysis solution followed by Qiagen's MagAttract HMW DNA extraction kit, according to the manufacturer's instructions. DNA Fragment size distribution was evaluated by running the sample on the Femto Pulse system.

Sequencing

Pacific Biosciences HiFi circular consensus 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 II (HiFi) and Illumina HiSeq (10X).

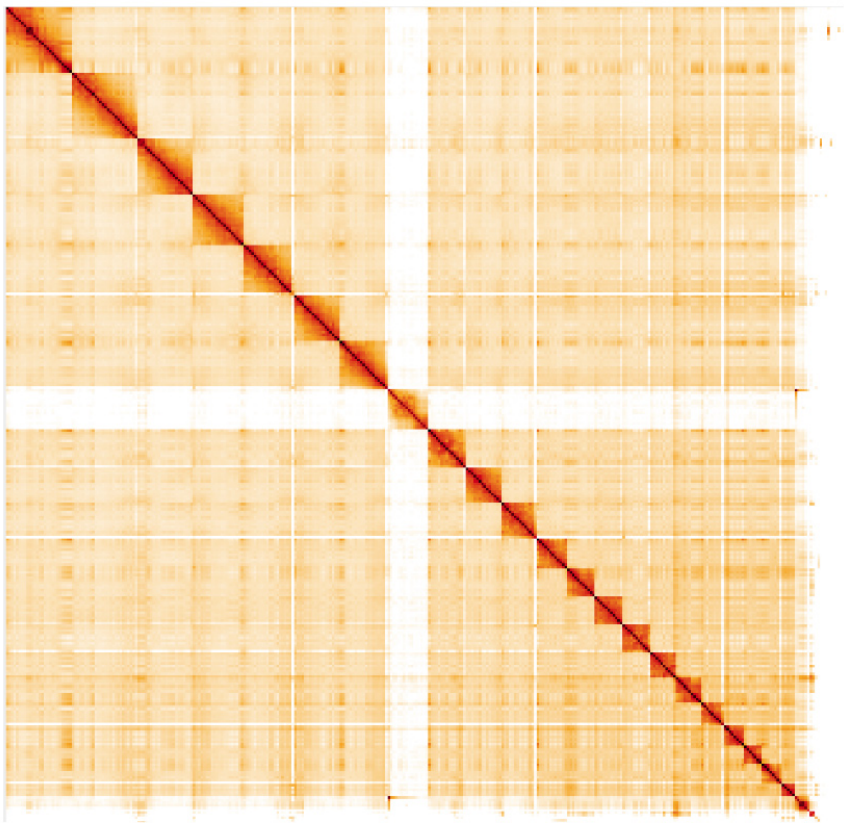


Figure 5. Genome assembly of *Meles meles*, mMelMel3.2: Hi-C contact map. Hi-C contact map of the mMelMel3.2 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=OELEHKHTiaSzLQOV68HQQ>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Meles meles*, mMelMel3.2. The assembly used the paternal haplotype with the X chromosome from maternal haplotype.

INSDC accession	Chromosome	Size (Mb)	GC%
OV277441.1	1	216.97	42.1
OV277442.1	2	210.73	39.2
OV277443.1	3	184.3	40.5
OV277444.1	4	164.77	39.8
OV277445.1	5	158.03	40.9
OV277446.1	6	153.77	42.4
OV277447.1	7	151.14	41.6
OV277449.1	8	120.13	42.9
OV277450.1	9	118.71	40.4
OV277451.1	10	109.97	40.8
OV277452.1	11	98.55	42.9

INSDC accession	Chromosome	Size (Mb)	GC%
OV277453.1	12	96.43	43.2
OV277454.1	13	87.85	43.4
OV277455.1	14	86.54	39.6
OV277456.1	15	84.98	42.3
OV277457.1	16	82.95	45
OV277458.1	17	69.24	43.3
OV277459.1	18	64.52	47.5
OV277460.1	19	62.17	46.7
OV277461.1	20	61.49	45.6
OV277462.1	21	45.12	49.1
OV277448.1	X	132.96	40.7
OV277463.1	Y	44.75	45.1
-	MT	0.02	38.9
-	Unplaced	132.63	51.7

Hi-C data were generated in the Tree of Life laboratory from the blood sample of mMe1Me13 using the Arima v2 kit and sequenced on an Illumina NovaSeq 6000 instrument. Standard read sequencing libraries were generated for the paternal (mMe1Me11) and maternal (mMe1Me12) specimen samples using the Illumina HiSeq (10X) instrument, as per the manufacturer's instructions.

Genome assembly

K-mer profiles for the parental illumina read data were generated using yak (<https://github.com/lh3/yak>). Assembly was carried out with Hifiasm in "trio" mode with the generated parental k-mers (Cheng *et al.*, 2021). The k-mers from the parents were used to partition the Hi-C reads (Rao *et al.*, 2014) into three parts: paternal and maternal specific and unclassified. The process was done with Canu's 'splitHaplotype' script (version 2.2; Koren *et al.*, 2018). Each haplotype assembly was then scaffolded using the combined haplotype-specific and unclassified Hi-C reads using SALSA2 (Ghurye *et al.*, 2019). 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 (Howe *et al.*, 2021) was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and PretextView. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performs 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.

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.

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.2	Cheng <i>et al.</i>, 2021
yak	0.1	https://github.com/lh3/yak
Canu split haplotype	2.2	Koren <i>et al.</i>, 2018
SALSA2	2.2	Ghurye <i>et al.</i>, 2019
MitoHiFi	2.0	Uliano-Silva <i>et al.</i>, 2021
HiGlass	1.11.6	Kerpedjiev <i>et al.</i>, 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis <i>et al.</i>, 2020

Data availability

European Nucleotide Archive: *Meles meles* (European badger). Accession number [PRJEB46333](#); <https://identifiers.org/ena.embl/PRJEB46333> (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *M. meles* 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 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 University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.6418202>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.6866293>.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.5746904>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.6125046>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.6418363>.

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](#)
- Annavi G, Dawson DA, Horsburgh GJ, *et al.*: **Characterisation of Twenty-One European Badger (*Meles Meles*) Microsatellite Loci Facilitates the Discrimination of Second-Order Relatives.** *Conservation Genet Resour.*

2011; 3(3): 515–18.

[Publisher Full Text](#)

Annavi G, Newman C, Buesching CD, *et al.*: **Heterozygosity-Fitness Correlations in a Wild Mammal Population: Accounting for Parental and Environmental Effects.** *Ecol Evol.* 2014b; 4(12): 2594–2609.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Annavi G, Newman C, Dugdale HL, *et al.*: **Neighbouring-Grown Composition**

and within-Group Relatedness Drive Extra-Group Paternity Rate in the European Badger (*Meles Meles*). *J Evol Biol*. 2014a; **27**(10): 2191–2203.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Bilham K, Boyd AC, Preston SG, *et al.*: Badger Macrophages Fail to Produce Nitric Oxide, a Key Anti-Mycobacterial Effector Molecule. *Sci Rep*. 2017; **7**: 45470.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Bright Ross JG, Newman C, Buesching CD, *et al.*: What Lies beneath? Population Dynamics Conceal Pace-of-Life and Sex Ratio Variation, with Implications for Resilience to Environmental Change. *Glob Chang Biol*. 2020; **26**(6): 3307–24.

[PubMed Abstract](#) | [Publisher Full Text](#)

Challis R, Richards E, Rajan J, *et al.*: BlobToolKit - Interactive Quality Assessment of Genome Assemblies. *G3 (Bethesda)*. 2020; **10**(4): 1361–74.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Cheng H, Concepcion GT, Feng X, *et al.*: Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm. *Nat Methods*. 2021; **18**(2): 170–75.

[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](#)

Donnelly CA, Woodroffe R, Cox DR, *et al.*: Positive and Negative Effects of Widespread Badger Culling on Tuberculosis in Cattle. *Nature*. 2006; **439**(7078): 843–46.

[PubMed Abstract](#) | [Publisher Full Text](#)

Dugdale HL, Macdonald DW, Pope LC, *et al.*: Polygynandry, Extra-Group Paternity and Multiple-Paternity Litters in European Badger (*Meles Meles*) Social Groups. *Mol Ecol*. 2007; **16**(24): 5294–5306.

[PubMed Abstract](#) | [Publisher Full Text](#)

Ghurye J, Rhie A, Walenz BP, *et al.*: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. *PLoS Comput Biol*. 2019; **15**(8): e1007273.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ham C, Donnelly CA, Astley KL, *et al.*: Effect of Culling on Individual Badger *Meles Meles* Behaviour: Potential Implications for Bovine Tuberculosis Transmission. *J Appl Ecol*. 2019; **56**(11): 2390–99.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Howe K, Chow W, Collins J, *et al.*: Significantly Improving the Quality of Genome Assemblies through Curation. *GigaScience*. 2021; **10**(1): g1aa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jenkins HE, Woodroffe R, Donnelly CA: The Duration of the Effects of Repeated Widespread Badger Culling on Cattle Tuberculosis Following the Cessation of Culling. *PLoS One*. 2010; **5**(2): e9090.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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](#)

Koepfli KP, Dragoo JW, Wang X: The Evolutionary History and Molecular Systematics of the Musteloidea. In: *Biology and Conservation of Musteloids*. Oxford University Press. 2018.

[Reference Source](#)

Koren S, Rhie A, Walenz BP, *et al.*: De Novo Assembly of Haplotype-Resolved Genomes with Trio Binning. *Nat Biotechnol*. 2018.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Macdonald DW, Newman C, Buesching CD: Badgers in the Rural Landscape—conservation Paragon or Farmland Pariah? Lessons from the Wytham Badger Project. In: *Wildlife Conservation on Farmland Volume 2*. Oxford

University Press. 2015a; Accessed July 14, 2022.

[Publisher Full Text](#)

Macdonald DW, Woodroffe R, Riordan P: Badgers and Bovine Tuberculosis: Beyond Perturbation to Life Cycle Analysis. In: *Wildlife Conservation on Farmland Volume 2: Conflict in the Countryside*, Oxford University Press. 2015b; **2**: 96–125.

[Reference Source](#)

Macdonald DW, Newman C: The Badgers of Wytham Woods: A Model for Behaviour, Ecology, and Evolution. Oxford University Press, Oxford. 2022.

[Reference Source](#)

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): 4647–54.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

McLaren GW, Thornton PD, Newman C, *et al.*: The Use and Assessment of Ketamine-Medetomidine-Butorphanol Combinations for Field Anaesthesia in Wild European Badgers (*Meles Meles*). *Vet Anaesth Analg*. 2005; **32**(6): 367–72.

[PubMed Abstract](#) | [Publisher Full Text](#)

Noonan MJ, Abidur Rahman M, Newman C, *et al.*: Avoiding Verisimilitude When Modelling Ecological Responses to Climate Change: The Influence of Weather Conditions on Trapping Efficiency in European Badgers (*Meles Meles*). *Glob Chang Biol*. 2015; **21**(10): 3575–85.

[PubMed Abstract](#) | [Publisher Full Text](#)

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): 1665–80.

[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): 737–46.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sugianto NA, Buesching CD, Macdonald DW, *et al.*: The Importance of Refining Anaesthetic Regimes to Mitigate Adverse Effects in Very Young and Very Old Wild Animals: The European Badger (*Meles Meles*). 2019; **3**(1): 10–17.

[Reference Source](#)

Sugianto NA, Heistermann M, Newman C, *et al.*: Alternative Reproductive Strategies Provide a Flexible Mechanism for Assuring Mating Success in the European Badgers (*Meles Meles*): An Investigation from Hormonal Measures. *Gen Comp Endocrinol*. 2021; **310**: 113823.

[PubMed Abstract](#) | [Publisher Full Text](#)

UK General Public Acts: Protection of Badgers Act 1992. 1992.

[Reference Source](#)

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: marcelauliano/MitoHiFi: mitohifi_v2.0. 2021.

[Publisher Full Text](#)

Wellcome Sanger Institute: The genome sequence of the European badger, *Meles meles* (Linnaeus, 1758), European Nucleotide Archive. [dataset]. 2022. <https://identifiers.org/ena.embl/PRJEB46333>

Woolhouse M, Wood J: Tuberculosis: society should decide on UK badger cull. *Nature*. Nature Publishing Group UK; 2013; **498**(7455): 434.

[PubMed Abstract](#) | [Publisher Full Text](#)

Yamaguchi N, Dugdale HL, Macdonald DW: Female Receptivity, Embryonic Diapause, and Superfetation in the European Badger (*Meles Meles*): Implications for the Reproductive Tactics of Males and Females. *Q Rev Biol*. 2006; **81**(1): 33–48.

[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 16 December 2022

<https://doi.org/10.21956/wellcomeopenres.20209.r53558>

© 2022 Edwards R. This is an open access peer review report 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.



Richard Edwards

¹ UWA Oceans Institute, University of Western Australia, Perth, WA, Australia

² School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

This manuscript reports what appears to be a high-quality genome of an important and charismatic species, the European badger. The genome assembly makes use of appropriate data and technically sound protocols. Overall, I am convinced that this represents a useful addition to the increasing number of reference genomes available, particularly once it has undergone annotation by Ensembl.

Whilst the overall methodology is clear, I thought there were a few areas where details were a little scant and/or it would benefit the community to know a little more. I also feel that some additional QC statistics/benchmarks would be helpful for understanding the assembly quality and completeness.

A. ESSENTIAL REVISIONS

A1. As well as BUSCO completeness, it would be useful to have kmer completeness and quality estimates (e.g. Merqury) and/or variant-calling based error estimates. Personally, I would be interested to see these for both the HiFi and illumina data. Ideally, VGP-style quality metrics would be reported for the estimated error rate in the genome. Similarly, it would be useful to know how the overall assembly size relates to the predicted genome size, and to what extent the repeats in the genome are collapsed. If the Fig 3 y-axis is not constrained, then it does not appear to be a problem, although it is possible that some high-copy scaffolds are off the scale.

B. MINOR COMMENTS/SUGGESTIONS

B1. Figure 1 features "An artistic impression of *M. meles*, generated using data from this genome sequence". This is quite cute and adds a splash of personality to the genome report. It would be nice to know how this was made, particularly if any software was used. (Should something be cited

for this?) If it was a manual process, I suggest that "created" rather than "generated" would be more appropriate.

B2. Why was only the paternal haplotype manually curated? Was the maternal haplotype without detectable errors, or was the goal to produce one high quality haploid reference along with an alternative assembly? The assembly is reported as a "haplotype resolved, diploid genome assembly", which implies that both haplotypes have equal quality/importance.

B3. The algorithm used for BUSCO v5.1.2 should be reported, as Augustus typically generates lower completeness scores than Metaeuk. It would also be useful to know how 95.1% compares to other chromosome-level carnivore genomes (e.g. dog). Are the missing 4% likely to represent limitations with the Carnivora_odb10 dataset, or are they genuinely assembly errors/omissions? (With so few gaps, I would have naively expected higher.)

B4. I note that manual curation and correction was performed using gEVAL. It would be useful for the assembly community to have some information regarding the number and nature of these corrections. Were they mainly edits to the scaffolding, or were there harder mis-assemblies to deal with?

B5. The y-axis legends on Fig 3 need fixing to be the correct orientation and more informative. Was only one quarter of the 10X data used for this plot (based on the SRA code)? Would it not make more sense to generate this plot using (all of) the HiFi data?

B6. It would be useful to have telomere predictions to know which chromosome arms are full-length.

B7. How effective was the haplotype phasing? Are there any QC metrics/checks that could be presented to show how consistent the paternal and maternal haplotypes are with the trio data?

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?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, bioinformatics and molecular evolution.

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.
