DATA NOTE

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

Chris Newman¹, Ming-shan Tsai¹, Christina D. Buesching¹−³, Peter W. H. Holland¹, 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

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.
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 predominately 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 (*Rhee 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.](image-url)
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...
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 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.
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 extragroup 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...
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 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.
**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=OELEHKFTiaSzLQOV68HQ](https://genome-note-higlass.tol.sanger.ac.uk/l/?d=OELEHKFTiaSzLQOV68HQ).

**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.

<table>
<thead>
<tr>
<th>INSDC accession</th>
<th>Chromosome</th>
<th>Size (Mb)</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OV277453.1</td>
<td>12</td>
<td>96.43</td>
<td>43.2</td>
</tr>
<tr>
<td>OV277454.1</td>
<td>13</td>
<td>87.85</td>
<td>43.4</td>
</tr>
<tr>
<td>OV277455.1</td>
<td>14</td>
<td>86.54</td>
<td>39.6</td>
</tr>
<tr>
<td>OV277456.1</td>
<td>15</td>
<td>84.98</td>
<td>42.3</td>
</tr>
<tr>
<td>OV277457.1</td>
<td>16</td>
<td>82.95</td>
<td>45</td>
</tr>
<tr>
<td>OV277458.1</td>
<td>17</td>
<td>69.24</td>
<td>43.3</td>
</tr>
<tr>
<td>OV277459.1</td>
<td>18</td>
<td>64.52</td>
<td>47.5</td>
</tr>
<tr>
<td>OV277460.1</td>
<td>19</td>
<td>62.17</td>
<td>46.7</td>
</tr>
<tr>
<td>OV277461.1</td>
<td>20</td>
<td>61.49</td>
<td>45.6</td>
</tr>
<tr>
<td>OV277462.1</td>
<td>21</td>
<td>45.12</td>
<td>49.1</td>
</tr>
<tr>
<td>OV277448.1</td>
<td>X</td>
<td>132.96</td>
<td>40.7</td>
</tr>
<tr>
<td>OV277463.1</td>
<td>Y</td>
<td>44.75</td>
<td>45.1</td>
</tr>
<tr>
<td>-</td>
<td>MT</td>
<td>0.02</td>
<td>38.9</td>
</tr>
<tr>
<td>-</td>
<td>Unplaced</td>
<td>132.63</td>
<td>51.7</td>
</tr>
</tbody>
</table>
Hi-C data were generated in the Tree of Life laboratory from the blood sample of mMelMel1 using the Arima v2 kit and sequenced on an Illumina NovaSeq 6000 instrument. Standard read sequencing libraries were generated for the paternal (mMelMel1) and maternal (mMelMel2) 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 ‘splitH-aplotype’ 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 Pretext. 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.

<table>
<thead>
<tr>
<th>Software tool</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiFiasm</td>
<td>0.15.2</td>
<td>Cheng et al., 2021</td>
</tr>
<tr>
<td>yak</td>
<td>0.1</td>
<td><a href="https://github.com/lh3/yak">https://github.com/lh3/yak</a></td>
</tr>
<tr>
<td>Canu split haplotype</td>
<td>2.2</td>
<td>Koren et al., 2018</td>
</tr>
<tr>
<td>SALSA2</td>
<td>2.2</td>
<td>Ghurye et al., 2019</td>
</tr>
<tr>
<td>MitoHiFi</td>
<td>2.0</td>
<td>Uliano-Silva et al., 2021</td>
</tr>
<tr>
<td>HiGlass</td>
<td>1.11.6</td>
<td>Kerpedjiev et al., 2018</td>
</tr>
<tr>
<td>PretextView</td>
<td>0.2.x</td>
<td><a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a></td>
</tr>
<tr>
<td>BlobToolKit</td>
<td>3.0.5</td>
<td>Challis et al., 2020</td>
</tr>
</tbody>
</table>

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


References