



DATA NOTE

REVISED The genome sequence of the American mink, *Neogale vison* (Schreber, 1777)

[version 2; peer review: 1 approved, 3 approved with reservations]

Michelle F. O'Brien¹, Rosa Lopez Colom¹,
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory
team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹Wildfowl and Wetlands Trust, Slimbridge, England, UK

v2 First published: 24 Feb 2025, 10:103
<https://doi.org/10.12688/wellcomeopenres.23780.1>
Latest published: 26 Sep 2025, 10:103
<https://doi.org/10.12688/wellcomeopenres.23780.2>

Abstract

We present a genome assembly from a female specimen of the American mink, *Neogale vison* (Chordata; Mammalia; Carnivora; Mustelidae). The assembly comprises two haplotypes, with total lengths of 2,697.13 megabases and 2,662.78 megabases, respectively. Most of haplotype 1 (98.72%) and haplotype 2 (98.73%) is scaffolded into 15 chromosomal pseudomolecules, including the X chromosome. Gene annotation of this assembly on Ensembl identified 19,566 protein-coding genes. The mitochondrial genome has also been assembled, with a length of 16.59 kilobases.

Keywords

Neogale vison, American mink, genome sequence, chromosomal, Carnivora



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

Open Peer Review

Approval Status ? ? ? ✓

	1	2	3	4
version 2				
(revision)	?			✓
26 Sep 2025	view			view
	↑			
version 1	?	?	?	
24 Feb 2025	view	view	view	

1. **Richard Edwards** , University of Western Australia, Perth, Australia
2. **Gabriela Pozo** , Universidad San Francisco de Quito, Quito, Ecuador
3. **Ellie E Armstrong** , University of California Riverside (Ringgold ID: 8790), Riverside, USA
4. **Manuel Ruedi** , Université de Genève, Genève, Switzerland

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: **O'Brien MF:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Lopez Colom R:** 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 (220540) and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>].

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

Copyright: © 2025 O'Brien MF *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: O'Brien MF, Lopez Colom R, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the American mink, *Neogale vison* (Schreber, 1777) [version 2; peer review: 1 approved, 3 approved with reservations]** Wellcome Open Research 2025, 10:103 <https://doi.org/10.12688/wellcomeopenres.23780.2>

First published: 24 Feb 2025, 10:103 <https://doi.org/10.12688/wellcomeopenres.23780.1>

REVISED Amendments from Version 1

In Version 2 of this genome note, we have edited the Background section to improve clarity and flow.

We have expanded the method section on RNA library preparation and sequencing.

We have also added information about annotation of this genome by Ensembl at the European Bioinformatics Institute.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Carnivora; Caniformia; Musteloidea; Mustelidae; Mustelinae; *Neogale*; *Neogale vison* (Schreber, 1777) (NCBI:txid452646)

Background

The American mink (*Neogale vison*) is a semi-aquatic mustelid which has a body length of up to 37 cm and a tail length of up to 18 cm. Its fur is thick and dark brown with white patches at chin, throat, chest and groin, although these patches are variable. *N. vison* can live for up to 12 years in the wild. Males weigh up to 1,500 g and females up to 800 g (Mammal Society, 2022).

American mink are generalist predators and are extremely aggressive. Their diet consists of birds, fish, mammals and invertebrates (Mammal Society, 2022). They have been major contributors to the decline of the water vole (*Arvicola terrestris*) in the UK and the European mink (*Mustela lutreola*) in Europe due to predation and intraspecific aggression respectively (Macdonald & Harrington, 2003). They also pose a risk to ground-nesting birds, and even to native crayfish (Smal, 1991).

Female mink have reproductive adaptations that allow both superfecundation (multiple ova from a single ovulation) and superfoetation (multiple ovulations within a single mating season) (Shackelford, 1952; Venge, 1973). This means that one litter may have a number of different paternities if multiple matings have taken place, which may provide them with a variety of advantages (Macdonald & Harrison, 2003). Females usually give birth to litters of up to six young in the spring (Mammal Society, 2022).

The range of *Neogale vison* is extremely large, due to its introduction into multiple countries. It is found in North America, South America, Europe and Asia (Bonesi & Palazon, 2007). *Neogale vison* is an invasive species in the United Kingdom, having been imported from North America to be bred in fur farms for its highly prized coat. However,

escapees and intentionally released animals have formed feral populations in the wild (Macdonald & Harrington, 2003).

Neogale vison has a global IUCN status of Least Concern (Mathews & Harrower, 2020). Studies have shown that releases of domesticated mink from fur farms affects the functional genetic diversity of wild mink, as principal component analyses showed clear distinctions between wild individuals near mink farms and those located in areas without mink farms (Morris *et al.*, 2020). Robust genomic data from UK mink are needed to assess genetic diversity and population structure, including the contribution of former farm stock to feral populations.

The *Neogale vison* genome has been sequenced previously (GCF_020171115.1) from a North American specimen (Karimi *et al.*, 2022). The genome sequence presented here was sequenced from a euthanased animal from Arundel, England, as part of the Darwin Tree of Life project.

Genome sequence report**Sequencing data**

The genome of a specimen of *Neogale vison* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 128.66 Gb from 10.46 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 2,810.59 Mb, with a heterozygosity of 0.10% and repeat content of 24.20%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 44.0x coverage of the genome. Chromosome conformation Hi-C data produced 390.29 Gb from 2,584.69 million reads.

Table 1 summarises the specimen and sequencing information, including the BioProject, study name, BioSample numbers, and sequencing data for each technology.



Figure 1. Photograph of the *Neogale vison* (mNeoVis2) specimen from which samples were taken for genome sequencing.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while Haplotype 2 was assembled to scaffold level and has also been deposited. The assembly was improved by manual curation, which corrected 349 misjoins or missing joins. These interventions decreased the scaffold count by 22.32% and increased the scaffold N50 by 52.35%. The final assembly has a total length of 2,697.13 Mb in 434 scaffolds, with 2,957 gaps, and a scaffold N50 of 221.12 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence for each haplotype (98.72% for haplotype 1 and 98.73% for haplotype 2) was assigned to 15 chromosomal-level scaffolds, representing 14 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, chromosome X was identified by synteny to a previous assembly of *Neogale vison* (GCF_020171115.1).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record in GenBank.

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness indicates the proportion of expected *k*-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

For haplotype 1, the estimated QV is 65.9, and for haplotype 2, the QV is 65.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 65.8. The *k*-mer completeness for haplotype 1 is 96.00%, and for haplotype 2 it is 95.42%. The combined assemblies achieve a *k*-mer completeness of 99.65%. BUSCO 5.5.0 analysis using the *carnivora_odb10* reference set (*n* = 14,502) achieved a completeness score of 94.9% (single = 91.4%, duplicated = 3.6%) for haplotype 1.

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on

Table 1. Specimen and sequencing data for *Neogale vison*.

Project information			
Study title	Neovison vison (American mink)		
Umbrella BioProject	PRJEB74153		
Species	Neogale vison		
BioSpecimen	SAMEA112468126		
NCBI taxonomy ID	452646		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	mNeoVis2	SAMEA112468166	muscle
Hi-C sequencing	mNeoVis2	SAMEA112468165	muscle
RNA sequencing	mNeoVis2	SAMEA112468166	muscle
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq X	ERR12791497	2.58e+09	390.29
PacBio Sequel IIe	ERR12779268	2.64e+06	29.46
PacBio Revio	ERR12779267	7.82e+06	99.2
RNA Illumina NovaSeq 6000	ERR12791496	5.86e+07	8.85

Table 2. Genome assembly data for *Neogale vison*.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	mNeoVis2.hap1.1	mNeoVis2.hap2.1
Assembly accession	GCA_964106545.1	GCA_964106205.1
Assembly level	chromosome	chromosome
Length (Mb)	2,697.13	2,662.78
Number of contigs	3,391	3,200
Number of scaffolds	434	351
Longest scaffold (Mb)	325.66	314.01
Assembly metrics* (benchmark)	Haplotype 1	Haplotype 2
Contig N50 length (≥ 1 Mb)	1.44 Mb	1.46 Mb
Scaffold N50 length (= chromosome N50)	221.12 Mb	217.71 Mb
Consensus quality (QV) (≥ 40)	65.9	65.7
k-mer completeness	96.00%	95.42%
Combined k-mer completeness (≥ 95%)	99.65%	
BUSCO** (S > 90%; D < 5%)	C:94.9%[S:91.4%,D:3.6%], F:0.7%,M:4.3%,n:14,502	-
Percentage of assembly mapped to chromosomes (≥ 90%)	98.72%	98.73%
Sex chromosomes (localised homologous pairs)	X	X
Organelles (one complete allele)	Mitochondrial genome: 16.59 kb	-

* BUSCO scores based on the *carnivora_odb10* BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

Assembly Standards [September 2024](#). The assembly achieves the EBP reference standard of **6.C.Q65**.

Methods

Sample acquisition and DNA barcoding

An adult female *Neogale vison* (specimen ID NHMUK014551518, ToLID mNeoVis2) was collected from Mill Road, Arundel, England, United Kingdom (latitude 50.86, longitude -0.55) on 2021-09-21. The sample for this specimen was collected from a deceased wild individual that was euthanased as part of a pest-control programme. The specimen was collected and identified by Michelle O’Brien (Wildfowl & Wetlands Trust). Muscle tissue samples were taken from areas of the thigh and the back and preserved by dry freezing (−80 °C).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by [Twyford et al. \(2024\)](#). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI). The

tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification ([Crowley et al., 2023](#)). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI ([Twyford et al., 2024](#)). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io ([Beasley et al., 2023](#)).

Metadata collection for samples adhered to the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io ([Denton et al., 2023](#)).

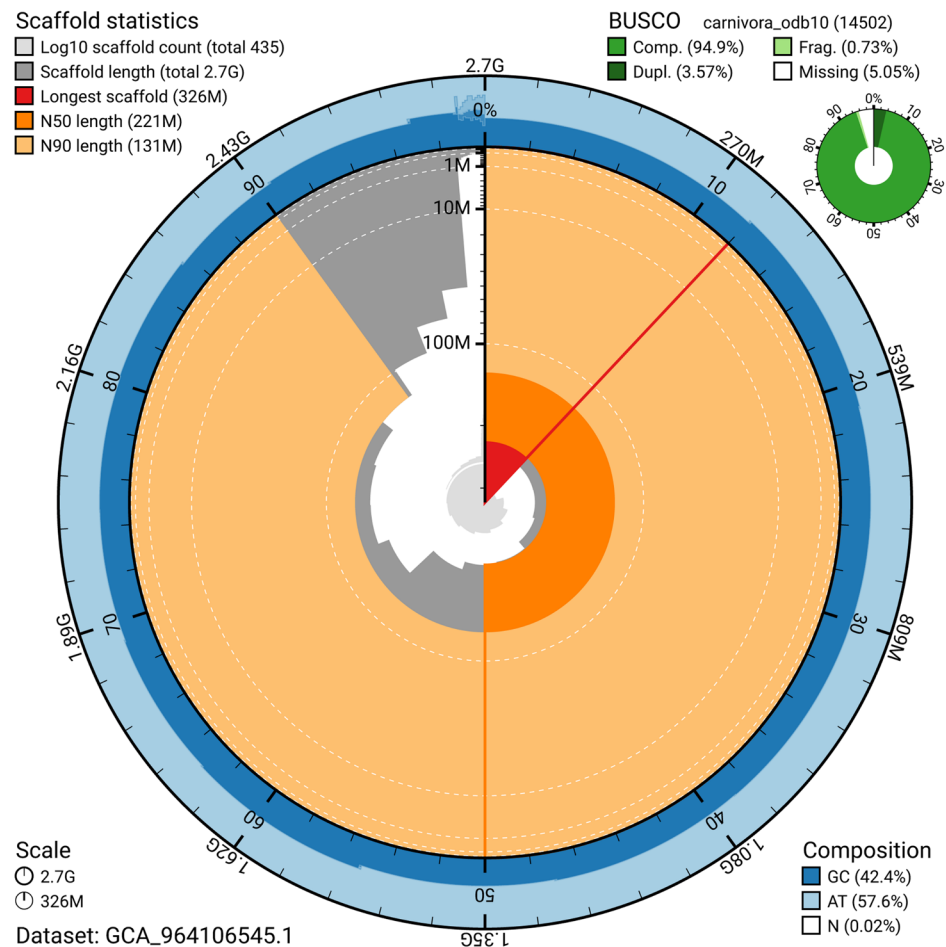


Figure 2. Genome assembly of *Neogale vison*, mNeoVis2.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the carnivora_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964106545.1/dataset/GCA_964106545.1/snail.

The mNeoVis2 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the muscle was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Oatley *et al.*, 2023b). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue of mNeoVis2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Hi-C sample preparation

Tissue from the muscle of the mNeoVis2 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozen tissue (stored at -80°C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. After

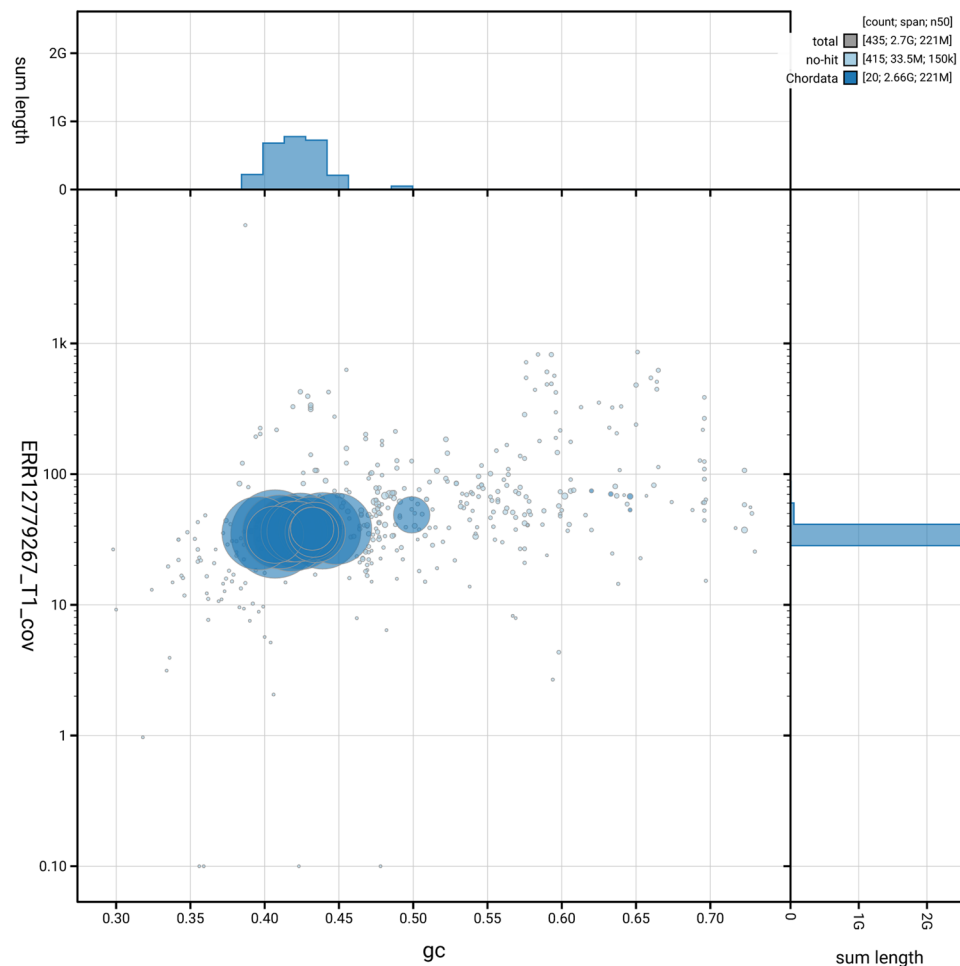


Figure 3. Genome assembly of *Neogale vison*, mNeoVis2.hap1.1: BlobToolKit GC-coverage plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964106545.1/blob.

crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and Arima-HiC v2 QC beads.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

PacBio HiFi

At a minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the

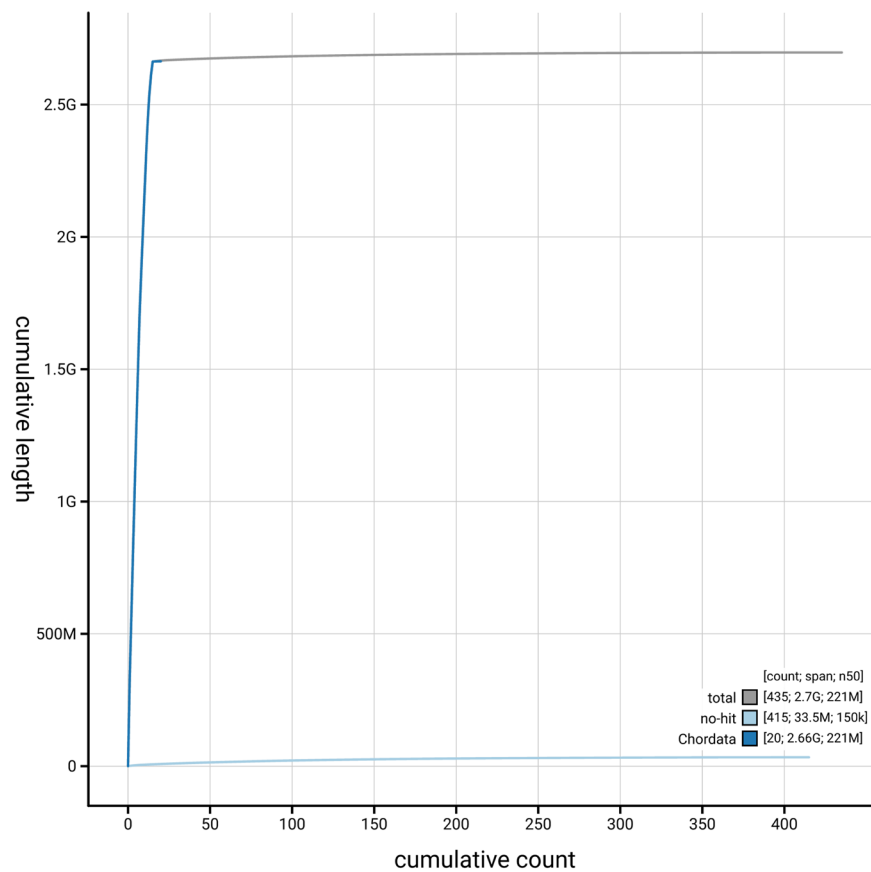


Figure 4. Genome assembly of *Neogale vison* mNeoVis2.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964106545.1/dataset/GCA_964106545.1/cumulative.

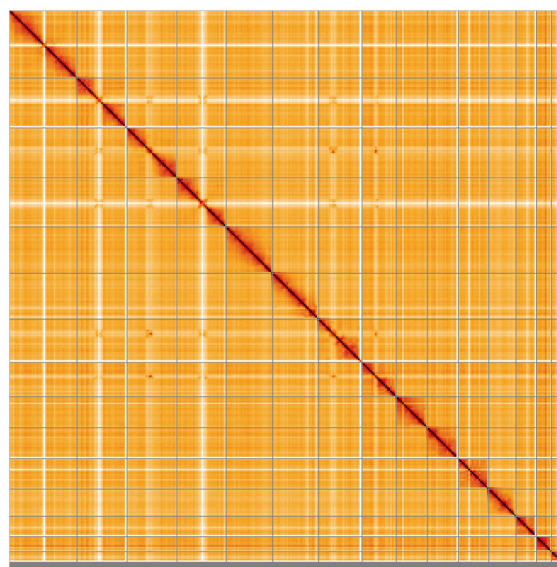


Figure 5. Genome assembly of *Neogale vison* mNeoVis2.hap1.1: Hi-C contact map of the mNeoVis2.hap1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/?d=Wm-v-cR9SCqo9qBbsy2RKA>.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Neogale vison*, mNeoVis2.

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ067083.1	1	325.66	40.5	OZ067339.1	1	314.01	40.5
OZ067084.1	2	241.06	44	OZ067340.1	2	240.85	44
OZ067085.1	3	240.71	42.5	OZ067341.1	3	239.06	42.5
OZ067086.1	4	237.42	42	OZ067342.1	4	229.49	41.5
OZ067087.1	5	224.2	41	OZ067343.1	5	219.87	41
OZ067088.1	6	221.12	39.5	OZ067344.1	6	217.71	39.5
OZ067089.1	7	209.22	45	OZ067345.1	7	208.93	45
OZ067090.1	8	164.45	43	OZ067346.1	8	167.9	43.5
OZ067091.1	9	149.99	42	OZ067348.1	9	146.58	41.5
OZ067092.1	10	149.39	42	OZ067347.1	10	153.18	42.5
OZ067093.1	11	146.76	43.5	OZ067349.1	11	142.98	43
OZ067095.1	12	98.39	43.5	OZ067351.1	12	94.72	43
OZ067096.1	13	72.69	43	OZ067352.1	13	76.75	43.5
OZ067097.1	14	50.33	50	OZ067353.1	14	46.71	49.5
OZ067094.1	X	131.22	40.5	OZ067350.1	X	130.26	40.5
OZ067098.1	MT	0.02	38.5				

Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55kb BAC analysis kit.

Samples were sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Samples were also sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration (in the range 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, a-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument.

RNA library preparation and sequencing

Poly(A) RNA-Seq libraries were constructed using the NEBNext@Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments of sizes 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading.

Sequencing was carried out on the Illumina NovaSeq 6000 to generate 150-bp paired-end reads.

Genome assembly, curation and evaluation

Assembly

Prior to assembly of the PacBio HiFi reads, a database of k -mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the k -mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), resulting in a pair of haplotype-resolved assemblies. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQUERY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. Sex chromosomes were identified by synteny. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation> (article in preparation).

Assembly quality assessment

The Merquery.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate k -mer completeness and assembly quality for the primary and alternate haplotypes using the k -mer databases ($k = 31$) that were computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The BlobToolKit pipeline is a Nextflow port of the previous Snakemake BlobToolKit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoAT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

Genome annotation report

The *Neogale vison* genome assembly (GCA_964106545.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 63,126 transcribed mRNAs from 19,566 protein-coding and 11,613 non-coding genes. The average transcript length is 48,915.76 bp, with an average of 2.01 coding transcripts per gene and 8.70 exons per transcript. For further information about the annotation, please refer to the [annotation page](#) on Ensembl.

Wellcome Sanger Institute – Legal and Governance

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**’, which can be found in full on the Darwin Tree of Life website [here](#). 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.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Goat CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhyip123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.5.1	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

areas of con Vertebrate Genomes Project (PRJNA489243). sideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

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

European Nucleotide Archive: Neovison vison (American mink). Accession number PRJEB74153; <https://identifiers.org/ena.embl/PRJEB74153>. The genome sequence is released openly for reuse. The *Neogale vison* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. (PRJEB40665) and the Vertebrate Genomes Project (PRJNA489243). 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](#) and [Table 2](#).

Production code used in genome assembly at the WSI Tree of Life is available at <https://github.com/sanger-tol>.

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12159242>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

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

Acknowledgements

We would like to thank members of the WWT Living Collections teams and WWT Reserve teams for their help in procuring these samples.

References

- Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Altschul SF, Gish W, Miller W, et al.: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, et al.: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor[®]3 for LI PacBio.** *protocols.io.* 2023.
[Publisher Full Text](#)
- Beasley J, Uhl R, Forrest LL, et al.: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; [Accessed 25 June 2024].
[Publisher Full Text](#)
- Bonesi L, Palazon S: **The American mink in Europe: status, impacts, and control.** *Biol Conserv.* 2007; **134**(4): 470–483.
[Publisher Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Kumar S, Sotero-Caio C, et al.: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- 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](#)
- Cheng H, Jarvis ED, Fedrigo O, et al.: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crowley L, Allen H, Barnes I, et al.: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- da Veiga Leprevost F, Grüning BA, Alves Aflitos S, et al.: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, et al.: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Denton A, Yatsenko H, Jay J, et al.: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023.
[Publisher Full Text](#)
- Diesh C, Stevens GJ, Xie P, et al.: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- do Amaral RJV, Denton A, Yatsenko H, et al.: **Sanger Tree of Life RNA extraction: automated MagMax[™] mirVana.** *protocols.io.* 2023.
[Publisher Full Text](#)
- Ewels P, Magnusson M, Lundin S, et al.: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels PA, Peltzer A, Fillinger S, et al.: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Formenti G, Abueg L, Brajuka A, et al.: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grüning B, Dale R, Sjödin A, et al.: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022.
[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. 2021; **10**(1): gaa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io*. 2023.
[Publisher Full Text](#)

Karimi K, Do DN, Wang J, *et al.*: **A chromosome-level genome assembly reveals genomic characteristics of the American mink (*Neogale vison*).** *Commun Biol*. 2022; **5**(1): 1381.
[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](#)

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One*. 2017; **12**(5): e0177459.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res*. 2022; **7**: 187.
[Publisher Full Text](#)

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics*. 2018; **34**(18): 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Macdonald DW, Harrington LA: **The American mink: the triumph and tragedy of adaptation out of context.** *N Z J Zool*. 2003; **30**(4): 421–441.
[Publisher Full Text](#)

Mammal Society: **American mink fact sheet v2.** 2022.

Manni M, Berkeley MR, Seppely 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](#)

Mathews F, Harrower C: **IUCN - compliant red list for Britain's terrestrial mammals. assessment by the mammal society under contract to natural England, natural resources wales and Scottish natural heritage.** 2020.
[Reference Source](#)

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J*. 2014; **2014**(239): 2, [Accessed 2 April 2024].
[Reference Source](#)

Morris KY, Bowman J, Schulte-Hostedde A, *et al.*: **Functional genetic diversity of domestic and wild American mink (*Neovison vison*).** *Evol Appl*. 2020; **13**(10): 2610–2629.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Narváez-Gómez JP, Mbye H, Oatley G, *et al.*: **Sanger Tree of Life sample homogenisation: Covaris cryoPREP® Automated Dry Pulverizer.** *protocols.io*. 2023.
[Publisher Full Text](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io*. 2023a.
[Publisher Full Text](#)

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI.** *protocols.io*. 2023b.
[Publisher Full Text](#)

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
[Publisher Full Text](#)

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics*. 2010; **26**(6): 841–842.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun*. 2020; **11**(1): 1432.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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): 1665–1680.
[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](#)

Rhie A, Walenz BP, Koren S, *et al.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol*. 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Shackleford RM: **Superfetation in the ranch mink.** *Am Nat*. 1952; **86**: 311–319.
[Publisher Full Text](#)

Smal CM: **Population studies on feral American mink *Mustela vison* in Ireland.** *J Zool*. 1991; **224**(2): 233–249.
[Publisher Full Text](#)

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res*. 2024; **9**: 339.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashennikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics*. 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.
[Publisher Full Text](#)

Venge O: **Reproduction in the mink.** Den Kongelige Veterinære Landbohøjskole, Årsskrift, 1973.

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics*. 2023; **39**(1): btac808.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:



Version 2

Reviewer Report 13 October 2025

<https://doi.org/10.21956/wellcomeopenres.27551.r134680>

© 2025 Ruedi M. 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.



Manuel Ruedi 

Université de Genève, Genève, Switzerland

The current version of the ms is excellent and clear, and I have only the following few changes to propose:

Arvicola terrestris is now considered a junior synonym of *Arvicola amphibius* (refer 1). How many mammal species are there now? Updates and trends in taxonomic, nomenclatural, and geographic knowledge. *Journal of Mammalogy* in press: TBD.).

"and intraspecific aggression" should be replaced by "and interspecific aggression" (*Neovison vison* versus *Mustela lutreola*!)

I would suggest that it is "native to North America but introduced elsewhere".

Perhaps repeat in the legend of the picture of dead mink that this specimen was a female, as this, of course, will immediately explain the absence of Y chromosome in its genome.

References

1. Burgin C, Zijlstra J, Becker M, Handika H, et al.: How many mammal species are there now? Updates and trends in taxonomic, nomenclatural, and geographic knowledge. *bioRxiv*. 2025. [Publisher Full Text](#)

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: Taxonomy mammalogy systematics

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 13 October 2025

<https://doi.org/10.21956/wellcomeopenres.27551.r134425>

© 2025 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

University of Western Australia, Perth, Australia

I don't think the authors have responded to my review. They appear to have addressed my minor comments but have failed to address the question of why the species was sequenced again. I requested a basic comparison with the previous assembly, which still appears to be absent. As a result, my Approval Status of "Approved with Reservations" still stands.

Competing Interests: No competing interests were disclosed.

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.

Version 1

Reviewer Report 23 June 2025

<https://doi.org/10.21956/wellcomeopenres.26228.r119476>

© 2025 Armstrong E. 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.



Ellie E Armstrong

Evolution, Ecology, and Organismal Biology, University of California Riverside (Ringgold ID: 8790), Riverside, California, USA

The report from O'Brien and Lopez Colom presents a novel, highly contiguous American mink genome from a female specimen collected in the U.K. The report is mostly complete and contains a fairly complete description of methods (though some of these are unpublished and have paper in progress placeholders), but the introduction of the manuscript would benefit from a more structured and cohesive presentation. At present, it appears somewhat abrupt and may not effectively guide the reader into the study's context and objectives. Enhancing the flow and clarity of this section could provide a clearer foundation for the research presented. I also find it surprising that the report does not make a direct comparison to the previous American mink genome in the publication, which is fairly standard even for genome reports. I offer some brief, line by line suggestions for improvement below.

'is a semi-aquatic mustelid which' change 'which' to 'that'

'with white patches at chin' change to 'at the chin'

'in the wild, males' change to 'in the wild, with males weighing up to'

The MacDonald in 'MacDonald and Harrington' has incorrect capitalization?

Sentence about decline awkward with in the UK and in Europe, rephrase 'They have been major contributors to the decline of the water vole in the UK and the European mink in Europe, due to predation...'

'The range of the American mink' this seems out of place, perhaps move this up to the first paragraph?

'Study into the genome of individuals in the UK may therefore feed' rephrase, awkward

'functional genetic diversity' what is the functional part of this sentence implying?

The samples for RNA and DNA seem like they are from different the same sample, but the HiC is from a different sample based on Table 1, but in the methods it seems like these are all from the same individual (mNeoVis2)?

Might want to add which BUSCO lineages were used. Was the RNA data used at all or just generated?

In Table 2 I am surprised you are calling this span instead of just total length or something more standard?

I think it would be useful to add in stats comparing the previous assembly from 2022.

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 biology, conservation genomics, population 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, however I have significant reservations, as outlined above.

Author Response 21 Sep 2025

Tree of Life Team Sanger

Thank you for reviewing this data note. Here are some of our responses and points of clarification:

- We have edited the Background section to enhance flow and clarity as you suggest.
- We have changed the capitalisation of “MacDonald and Harrington”, but note the capitalisation in the original reference:
<https://www.tandfonline.com/doi/abs/10.1080/03014223.2003.9518350>.
- Samples for RNA and DNA: all samples were taken from the same individual (mNeoVis2). The BioSample accessions are linked to the submitted tubes, and different BioSample accessions indicate that a different tube of tissue was used for Hi-C sample preparation.
- We clearly state in the footnote to Table 1, in Figure 2 and in the methods section that the BUSCO lineage used for analysis was *carnivora_odb10*.
- The RNA data is generated and released under the same BioProject, but is not used in the assembly steps. The genome has been annotated by the European Bioinformatics Institute since version 1 of the data note, and we have included this information in version 2.
- We commonly use the word “span” rather than “length” in data notes to avoid the implication that the sequence is physically continuous. We have changed it to “Length” in Table 2.
- Thank you for the suggestion to compare the statistics of the previous chromosome-level genome sequence of *Neogale vison* (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_020171115.1/). We have chosen not to include a direct, side-by-side comparison of assembly statistics because the two genomes differ in specimen origin (North America vs UK), biological background (strain/population), and methods/pipelines. Our aim in a data note is to provide a high-quality genomic resource with clear provenance and methods so that researchers can make the comparisons most relevant to their questions. Both assemblies are chromosome-level with high completeness. Interested readers can readily conduct formal comparisons using the deposited data and methods we provide.

Competing Interests: No competing interests were disclosed.

Reviewer Report 05 April 2025

<https://doi.org/10.21956/wellcomeopenres.26228.r119471>

© 2025 Pozo G. 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.



Gabriela Pozo

Universidad San Francisco de Quito, Quito, Ecuador

The article is a very clear description of the high-quality genome assembly of the American mink.

Comments:

1. In the background section, I believe it would be helpful to combine paragraphs so that they are not on or two sentences. For example, I would combine paragraphs 1 and 2, as well as 5 and 6.
2. The last paragraph mentions that the genome has been sequenced previously. There is no justification or explanation as to why it is being re-sequenced. What would this genome be used for. I believe this would strengthen the article.
3. In the methods section I suggest providing more detail regarding:
 - RNA sequencing parameters
 - Assembly - I suggest providing more details regarding the assembly and mapping (what parameters were used). This comment applies to both the nuclear and mitochondrial genome.

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

Partly

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: 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, however I have significant reservations, as outlined above.

Author Response 21 Sep 2025

Tree of Life Team Sanger

Thank you for reviewing this data note. We have submitted a new version of the genome note and we respond to each of your concerns below.

Reviewer comment 1. In the background section, I believe it would be helpful to combine paragraphs so that they are not on or two sentences. For example, I would combine paragraphs 1 and 2, as well as 5 and 6.

Response: We have combined paragraphs and edited the Background section.

Reviewer comment 2. The last paragraph mentions that the genome has been sequenced previously. There is no justification or explanation as to why it is being re-sequenced. What would this genome be used for. I believe this would strengthen the article.

Response: We have added information on the rationale of sequencing a UK mink specimen.

Reviewer comment 3. In the methods section I suggest providing more detail regarding: - RNA sequencing parameters - Assembly - I suggest providing more details regarding the assembly and mapping (what parameters were used). This comment applies to both the nuclear and mitochondrial genome.

Response: We have added more detail to the RNA library preparation and sequencing section in Version 2. The full details of assembly are given in the section "Genome assembly, curation and evaluation". In addition, we have now linked to the pipelines used for genome assembly (<https://github.com/sanger-tol>) in the "Data availability" section.

Competing Interests: No competing interests were disclosed.

Reviewer Report 18 March 2025

<https://doi.org/10.21956/wellcomeopenres.26228.r119469>

© 2025 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 

University of Western Australia, Perth, Australia

This manuscript describes a high quality phased genome assembly for an important UK invasive species, the American mink. The assembly data and methods are generally well described and are of high quality. However, as the authors point out at the end of the Introduction, this species already has a reference genome - itself assembled to chromosome-level. This raises the question as to why the species was sequenced again. Are there important population differences between the Canadian mink and the UK mink? Or were there issues identified with the original genome that this new assembly sought to rectify? A second female was selected for sequencing, so it was not to

add the Y chromosome. Was it to get a phased version? (The original appears to be a primary haploid assembly.)

Whilst I appreciate that this is a genome report with limited analysis, I think it is important to present the statistics of the original genome assembly (both genome stats and chromosome sizes) alongside those of the new assembly. This is important to help guide the reader concerning which assembly they should use for different purposes.

Minor Comments:

- Typo - rogue capitalisation: MacDonald DW, Harrington LA: The American mink: the triumph and tragedy of adaptation out of context. N Z J Zool. 2003; 30(4): 421–441.
- Curation: At two points, details are missing and cited as “(article in preparation).” This needs to be described or a link provided to published details, e.g. a pre-print or similar, or a personal communication from named authors. The rapid curation GitHub citation needs to link to a specific version/release or, if this is not possible (there do not appear to be versions or releases), provide the date of access and the github commit hash.

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

Partly

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: Biodiversity genomics, genome assembly, bioinformatics, 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.
