DATA NOTE

The genome sequence of the red deer, *Cervus elaphus* Linnaeus 1758 [version 1; peer review: 1 approved, 1 approved with reservations]

Josephine Pemberton¹, Susan E. Johnston¹, T. John Fletcher²,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹University of Edinburgh, Edinburgh, UK
²Reediehill Farm, Auchtermuchty, Fife, UK

Abstract
We present a genome assembly from an individual female *Cervus elaphus* (the red deer; Chordata; Mammalia; Artiodactyla; Cervidae). The genome sequence is 2,887 megabases in span. The majority of the assembly is scaffolded into 34 chromosomal pseudomolecules, with the X sex chromosome assembled.

Keywords
*Cervus elaphus*, red deer, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Reviewer Status

Invited Reviewers

1

2

version 1
13 Dec 2021

Report

Report

1. Klaus-Peter Koepfli, George Mason University, Fairfax, USA
Smithsonian Conservation Biology Institute, Front Royal, USA

2. Kun Wang, Northwestern Polytechnical University, Xi'an, China

Any reports and responses or comments on the article can be found at the end of the article.
Species taxonomy
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Artiodactyla; Ruminantia; Pecora; Cervidae; Cervinae; Cervus; *Cervus elaphus* Linnaeus 1758 (NCBI:txid9860).

Background
Long term studies of individuals living in the wild offer unparalleled opportunities to study ecological and evolutionary processes (Clutton-Brock & Sheldon, 2010). One such study concerns the red deer on the Isle of Rum, Scotland (https://rum-deer.bio.ed.ac.uk/). For 50 years the lives of individuals living in the northeast part of the island have been followed from birth, through all breeding attempts, to death, giving rise to detailed understanding of the impacts of weather and density on individual survival and breeding success (Clutton-Brock et al., 1982) and how these effects on individuals build up into population dynamics, with implications for deer management (Clutton-Brock et al., 2002). An ambition to measure male breeding success accurately and thereby recover the pedigree for this wild population led to the development of microsatellite and then single nucleotide polymorphism (SNP) assays for red deer and inspired two widely-used computer programs for inferring parentage from genetic markers, CERVUS and Sequoia (Huisman, 2017; Marshall et al., 1998). The resulting pedigree has enabled several evolutionary genetic investigations, such as measurement of the heritability of different traits including breeding success and antler size (Kruuk et al., 2000; Kruuk et al., 2002) and tests of the causes of evolutionary stasis or change in traits such as antler size and birth weight (Bonnet et al., 2019; Kruuk et al., 2002). Our most recent studies have used thousands of genome-wide SNP markers to demonstrate the impact of inbreeding depression in the population (Huisman et al., 2016), to resolve a genetic map for red deer (Johnston et al., 2017) and to investigate determinants of individual variation in recombination in the population (Johnston et al., 2018). Increasingly, our studies depend on knowing precisely where each marker is in the genome.

When the Darwin Tree of Life project announced its intention to make a high quality genome for all UK species, we were delighted that it was proposed to sequence a red deer from Rum, so that the information gained would be maximally relevant to the Rum study population. An effort to obtain suitably high molecular weight DNA from animals on the island failed, and so we turned to the one deer from Rum living on the mainland.

Thistle was born on Rum in 1993 and had a bad start: unusually, her mother rejected her after she was weighed and tagged by the research team. In a rare move, she was taken in and hand reared by Fiona Guinness of the deer project and later shipped to Reediehill Deer Farm, Fife, Scotland, UK (latitude 56.304268, longitude -3.274638). A total of 30-fold coverage in Pacific Biosciences single-molecule circular consensus (HiFi) long reads and 26-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 35 missing/misjoins, reducing the assembly length by 8.65% and the scaffold number by 59.66%, and increasing the scaffold N50 by 16.54%.

The final assembly has a total length of 2,887 Mb in 145 sequence scaffolds with a scaffold N50 of 83 Mb (Table 1). The majority, 96.02%, of the assembly sequence was assigned to 34 chromosomal-level scaffolds, representing 33 autosomes (numbered by synteny to the *C. elaphus* genome described by (Bana et al., 2018)), and the X sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Manni et al., 2021)
Table 1. Genome data for *Cervus elaphus*, mCerEla1.1.

| **Project accession data**       |                |
|----------------------------------|----------------|
| Assembly identifier              | mCerEla1.1     |
| Species                          | *Cervus elaphus*|
| Specimen                         | mCerEla1       |
| NCBI taxonomy ID                 | NCBI:txid41139 |
| BioProject                       | PRJEB45171     |
| BioSample ID                     | SAMEA7523520   |
| Isolate information              | Female, blood sample |

| **Raw data accessions**          |                |
|----------------------------------|----------------|
| PacificBiosciences SEQUEL II     | ERR6406211-ERR6406213, ERR6436379 |
| 10X Genomics Illumina            | ERR6054832-ERR6054835 |
| Hi-C Illumina                    | ERR6054836     |

| **Genome assembly**              |                |
|----------------------------------|----------------|
| Assembly accession               | GCA_910594005.1 |
| Accession of alternate haplotype | GCA_910593805.1 |
| Span (Mb)                        | 2,887          |
| Number of contigs                | 185            |
| Contig N50 length (Mb)           | 69             |
| Number of scaffolds              | 145            |
| Scaffold N50 length (Mb)         | 83             |
| Longest scaffold (Mb)            | 99             |
| BUSCO genome score*              | C:96.1%[S:92.8%,D:3.3%], F:1.1%, M:2.8%, n:13335 |

| **Genome annotation**            |                |
|----------------------------------|----------------|
| Number of genes                  | 32,781         |
| Number of protein-coding genes   | 22,941         |
| Average length of gene (bp)      | 37,644         |
| Average number of exons per transcript | 11.26    |
| Average exon size (bp)           | 320            |
| Average intron size (bp)         | 6,705          |
| BUSCO annotation score**         | C:99%[S:95.5%, D:3.5%], F:0.3%, M:0.7%, n:13335 |

C= complete [S= single copy, D= duplicated], F= fragmented, M= missing, n= number of orthologues in comparison.

*BUSCO scores based on the cetartiodactyla_odb10 BUSCO set using v5.1.2, run on the mCerEla1.1 genome assembly using BlobToolKit. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/mCerEla1.1/dataset/CAJVBLO1/busco.***

**BUSCO scores based on the cetartiodactyla_odb10 BUSCO set using v4.1.4, run on the NCBI RefSeq annotation of the mCerEla1.1 genome assembly (NCBI *Cervus elaphus* Annotation Release 100).**

v5.1.2 completeness of 96.1% using the cetartiodactyla_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.
**Figure 2. Genome assembly of *Cervus elaphus*, mCerEla1.1: 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 2,886,619,877 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 (179,953,079 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (83,473,711 and 51,846,305 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 cetartiodactyla_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/mCerEla1.1/dataset/CAJVBL01/snail.

**Genome annotation report**

The mCerEla1.1 genome has been annotated using the NCBI RefSeq annotation pipeline (Table 1; NCBI *Cervus elaphus* Annotation Release 100). The resulting annotation includes 57,591 transcribed mRNAs from 22,941 protein-coding and 9,840 non-coding genes. There are 2.25 transcripts per gene and 11.26 exons per transcript. The annotated genome has a BUSCO v4.1.4 completeness of 99.0% using the cetartiodactyla_odb10 reference set.

**Methods**

Sample acquisition, DNA extraction and sequencing

A blood sample from a single female *Cervus elaphus* was collected from Reediehill Deer Farm, Fife, Scotland, UK (latitude
Figure 3. Genome assembly of *Cervus elaphus*, mCerEla1.1: 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/mCerEla1.1/dataset/CAJVBL01/blob.

DNA was extracted from the blood sample in the Scientific Operations core of the Wellcome Sanger Institute using the Qiagen MagAttract HMW DNA kit, according to the manufacturer’s instructions. Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud sequencing libraries were constructed according to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II and Illumina HiSeq X instruments. Hi-C data
were generated using the Arima v2 Hi-C kit and sequenced on an Illumina NovaSeq 6000 instrument.

Genome assembly
Assembly was carried out with Hiiasm (Cheng et al., 2021); haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao et al., 2014) using SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using HiGlass (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021), which performed 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.
Figure 5. Genome assembly of *Cervus elaphus*, mCerEla1.1: Hi-C contact map. Hi-C contact map of the mCerEla1 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Cervus elaphus*, mCerEla1.1.

| INSDC accession | Chromosome | Size (Mb) | GC%  |
|-----------------|------------|-----------|------|
| OU343078.1      | 1          | 99.37     | 42.9 |
| OU343079.1      | 2          | 50.83     | 44.4 |
| OU343080.1      | 3          | 55.86     | 40.4 |
| OU343081.1      | 4          | 68.74     | 45.4 |
| OU343082.1      | 5          | 134.76    | 44.5 |
| OU343083.1      | 6          | 51.85     | 41.8 |
| OU343084.1      | 7          | 52.32     | 43.1 |
| OU343085.1      | 8          | 55.69     | 42.5 |
| OU343086.1      | 9          | 109.73    | 42.1 |
| OU343087.1      | 10         | 55.77     | 48.3 |
| OU343088.1      | 11         | 103.26    | 42.6 |
| OU343089.1      | 12         | 100.05    | 41.3 |
| OU343090.1      | 13         | 72.18     | 43.6 |
| OU343091.1      | 14         | 81.69     | 42.9 |
| OU343092.1      | 15         | 94.14     | 42.2 |
| OU343093.1      | 16         | 47.76     | 43.3 |
| OU343094.1      | 17         | 65.07     | 38.6 |
| OU343095.1      | 18         | 125.77    | 41.4 |
| OU343096.1      | 19         | 115.14    | 42.9 |
| OU343097.1      | 20         | 140.94    | 43.8 |
| OU343098.1      | 21         | 83.47     | 41.7 |
| OU343099.1      | 22         | 85.85     | 46.5 |
| OU343100.1      | 23         | 82.66     | 43.5 |
| OU343101.1      | 24         | 71.62     | 45   |
| OU343102.1      | 25         | 71.80     | 41.4 |
| OU343103.1      | 26         | 50.00     | 43.8 |
| OU343104.1      | 27         | 64.51     | 42.5 |
| OU343105.1      | 28         | 64.36     | 39.8 |
| OU343106.1      | 29         | 63.76     | 40.3 |
| OU343107.1      | 30         | 86.20     | 40.4 |
| OU343108.1      | 31         | 55.17     | 39.1 |
| OU343109.1      | 32         | 52.23     | 43.4 |
| OU343110.1      | 33         | 79.32     | 39.7 |
| OU343077.1      | X          | 179.95    | 42.4 |
| OU343111.1      | MT         | 0.02      | 37.8 |
| -               | Unplaced   | 114.79    | 52.2 |
Table 3. Software tools used.

| Software tool        | Version | Source                                                                 |
|----------------------|---------|------------------------------------------------------------------------|
| HiFiasm              | 0.12    | Cheng et al., 2021                                                    |
| purge_dups           | 1.2.3   | Guan et al., 2020                                                     |
| SALSA2               | 2.2     | Ghurye et al., 2019                                                   |
| longranger align     | 2.2.2   | https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines |
| freebayes            | 1.3.1-17-gaa2ace8 | Garrison & Marth, 2012                                               |
| MitoHiFi             | 2.2     | Uliano-Silva et al., 2021                                             |
| gEVAL                | N/A     | Chow et al., 2016                                                     |
| HiGlass              | 1.11.6  | Kerpedjiev et al., 2018                                               |
| PretextView           | 0.2.x   | https://github.com/wtsi-hpag/PretextView                               |
| BlobToolKit          | 2.6.2   | Challis et al., 2020                                                  |

Genome annotation
The C. elaphus assembly was annotated by the NCBI Eukaryotic Genome Annotation Pipeline, an automated pipeline that annotates genes, transcripts and proteins on draft and finished genome assemblies. The annotation was generated from transcripts and proteins retrieved from NCBI Entrez by alignment to the genome assembly, as described here (Pruitt et al., 2014).

Ethical/compliance issues
The materials that have contributed to this genome note were supplied by a Tree of Life collaborator. 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 areas of consideration are:
- Ethical review of provenance and sourcing of the material;
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability
European Nucleotide Archive: Cervus elaphus (red deer). Accession number PRJEB45171: https://www.ebi.ac.uk/ena/browser/view/PRJEB45171.

The genome sequence is released openly for reuse. The C. elaphus 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 Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.5744972.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.5744840.

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.5743293.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.5638618.
Guan D, McCarthy SA, Wood 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

Bana NA, Nyiri A, Nagy J, et al.: The Red Deer Cervus Elaphus Genome CerElA1.0: Sequencing, Annotating, Genes, and Chromosomes. Mol Genet Genomics. 2018; 293(3): 665-84.

PubMed Abstract | Publisher Full Text

Blaxter K: Cervus Annus Vivere: An Account of Opinion about the Length of Life of the Deer. Br Vet J. 1979; 135(6): 591-99.

PubMed Abstract | Publisher Full Text

Bonnet T, Morrissey MB, Morris A, et al.: The Role of Selection and Evolution in Changing Parturition Date in a Red Deer Population. PLoS Biol. 2019; 17(11): e3000493.

PubMed Abstract | Publisher Full Text | Free Full Text

Chow W, Brugger K, Caccamo M, et al.: Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm. Nat Methods. 2021; 18(2): 70-75.

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

Clutton-Brock TH, Counsil TN, Milner-Gulland EJ, et al.: Sex Differences in Emigration and Mortality Affect Optimal Management of Deer Populations. Nature. 2002; 415(6872): 633-37.

PubMed Abstract | Publisher Full Text

Clutton-Brock TH, Guinness FE, Albion SD: Red Deer: Behavior and Ecology of Two Sexes. University of Chicago Press, 1982.

Reference Source

Clutton-Brock T, Sheldon BC: Individuals and Populations: The Role of Long-Term, Individual-Based Studies of Animals in Ecology and Evolutionary Biology. Trends Ecol Evol. 2010; 25(10): 562-73.

PubMed Abstract | Publisher Full Text

Garrison E, Marth G: Haplotype-Based Variant Detection from Short-Read Sequencing. arXiv: 1207.3007. 2012.

Reference Source

Ghurye J, Rhee 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

Guo D, McCarthy SA, Wood J, et al.: Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. Bioinformatics. 2020; 36(9): 2896-98.

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): giaa153.

PubMed Abstract | Publisher Full Text | Free Full Text

Huisman J: Pedigree Reconstruction from SNP Data: Parentage Assignment, Sibship Clustering and beyond. Mol Ecol Resour. 2017; 17(5): 1009-24.

PubMed Abstract | Publisher Full Text | Free Full Text

Huisman J, Kruuk LE, Ellis PA, et al.: Inbreeding Depression across the Lifespan in a Wild Mammal Population. Proc Natl Acad Sci U S A. 2016; 113(13): 3585-90.

PubMed Abstract | Publisher Full Text | Free Full Text

Johnson SE, Huisman J, Ellis PA, et al.: A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus Elaphus). G3 (Bethesda). 2017; 7(8): 2859-70.

PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekchos 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

Kruuk LE, Clutton-Brock TH, Slate, et al.: Heritability of Fitness in a Wild Mammal Population. Proc Natl Acad Sci U S A. 2000; 97(2): 698-703.

PubMed Abstract | Publisher Full Text | Free Full Text

Kruuk EB, Slate, J, Pemberton JM, et al.: Antler Size in Red Deer: Heritability and Selection but No Evolution. Evolution. 2002; 56(8): 1683-95.

PubMed Abstract | Publisher Full Text

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

Marshall TC, Slate J, Kruuk LE, et al.: Statistical Confidence for Likelihood-Based Paternity Inference in Natural Populations. Mol Ecol. 1998; 7(3): 639-55.

PubMed Abstract | Publisher Full Text

Pruneti KD, Brown GR, Hiatt SM, et al.: ReSeq: An Update on Mammalian Reference Sequences. Nucleic Acids Res. 2014; 42(Database issue): D756-63.

PubMed Abstract | Publisher Full Text | Free Full Text

Rao SS, Huntley MH, Durand NC, et al.: A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014; 159(7): 1665-80.

PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, et al.: marcelualian/MitoHifi: mitohifi_v2.0. 2021.

Publisher Full Text
Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 19 January 2022

https://doi.org/10.21956/wellcomeopenres.19343.r47815

© 2022 Wang K. 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.

Kun Wang
School of Ecology and Environment, Northwestern Polytechnical University, Xi’an, China

The manuscript by Josephine Pemberton and coauthors provides a genome assembly and genome annotation report of the red deer. We note that the red deer genome has been reported (see: https://link.springer.com/article/10.1007%2Fs00438-017-1412-3), and my suggestions for improvement are:

My concerns:
1. The authors should make a comparison of the different versions of the genome, showing the improvement in genome quality.

2. The repetitive elements should be annotated and provided, which are important for further analysis.

3. The BUSCO assessment should also be performed for the annotated genes.

4. “Table 3 contains a list of all software tool versions used, where appropriate.” has an extra punctuation mark and the sentence seems not finished.

References
1. Bana NÁ, Nyiri A, Nagy J, Frank K, et al.: The red deer Cervus elaphus genome CerEla1.0: sequencing, annotating, genes, and chromosomes. Mol Genet Genomics. 2018; 293 (3): 665-684
PubMed Abstract | 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?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Vertebrate 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.

Reviewer Report 10 January 2022

https://doi.org/10.21956/wellcomeopenres.19343.r47814

© 2022 Koepfli K. 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.

**Klaus-Peter Koepfli**

1 Smithsonian-Mason School of Conservation, George Mason University, Fairfax, VA, USA
2 Center for Species Survival, Smithsonian Conservation Biology Institute, Front Royal, VA, USA

Pemberton and colleagues report the assembly and annotation of a new red deer genome, which was sequenced and assembled using a combination of PacBio HiFi, 10X Genomics short reads, and high-throughput chromosomal conformation capture (Hi-C). This assembly represents a marked improvement over the original red deer genome assembly (CerEla1.0; GCA_002197005.1) reported by Bana et al. 2018 (*Molecular Genetics and Genomics*, 293: 665–684),1 which was generated from a male individual using only Illumina short reads and the ALLPATHS-LG assembler.

**General comments:**
The manuscript is well written. I especially enjoyed reading about the history of the female red deer, Thistle, whose genome was sequenced and assembled for this study. The assembly represents the most contiguous genome assembly for this species to date and will no doubt provide an enduring resource for the ongoing evolutionary, ecological, and genetic studies of the red deer study system on the Isle of Rum and elsewhere. The assembly, annotation and quality assessment analyses appear to have been rigorously conducted. The figures and tables are of high quality and generally interpretable, although I don’t think that Figures 3 and 4 are really all that useful. Nonetheless, through the BlobToolKit package, these figures allow interactive exploration by interested readers, so it’s fine to retain them. Having reviewed many genome report manuscripts, many important details are missing from the Methods section that I think should be added in the interest of reproducibility. I provide recommendations for these and a few other issues for the authors to address, which I think would improve the manuscript.
Specific comments:

Genome sequence report
- “The assembly has a BUSCO (Manni et al., 2021) v5.1.2 completeness of 96.1% using the cetartiodactyla_odb10 reference set.” Here, it would be useful to specify the number of BUSCOs in the cetartiodactyla_odb10 gene set (n = 13,335), which is also included in Figure 2.

- In conjunction with the genome assembly metrics reported in Table 1, it would be useful to compare the metrics of this new red deer assembly with the previous assembly reported by Bana et al. 2018 to highlight the significant improvement in contiguity of the newer assembly. For example, number of contigs = 185 versus 406,637; contig N50 = 69 Mb versus 7.9 Kb). This increased contiguity has undoubtedly resulted in an improved annotation that will facilitate more accurate inferences of genetic processes and phenomena in this species (per the discussion in Rhie et al 2021 Nature 592, pages 737–746).

Methods:
- Sample acquisition: The authors state that a blood sample was collected from a female red deer (presumably Thistle). Was the blood collected into an EDTA vacutainer tube? How and how long was the sample stored before genomic DNA was extracted?

- DNA extraction: Following DNA extraction, how was DNA quality assessed? How were DNA fragment lengths assayed prior to HiFi sequencing?

- Sequencing: Prior to HiFi circular consensus sequencing, what library kit was used to prepare the templates (e.g., SMRTbell Express Template Prep Kit 2.0)? How many SMRT cells were used to generate the 30x coverage long reads?

- Sequencing: 10X Genomics and Arima Hi-C library preparation and sequencing: the authors should provide details about the laboratory protocols used for these methods (i.e., genomic DNA or nuclei input, library QC, number of reads generated, etc.).

- Genome assembly: Please specify the commands/flags/parameter settings (or default) implemented for assembly or scaffolding using Hifiasm, SALSA2 and MitoHiFi.

- Genome assembly: “Manual curation was performed using HiGlass and Pretext.” Assembly curation is a highly idiosyncratic process. What exactly was curated? Were any misjoined contigs or scaffolds encountered and resolved? How many gaps were filled? How many and what type of sequence and other assembly errors were corrected?

References
1. Bana NÁ, Nyiri A, Nagy J, Frank K, et al.: The red deer Cervus elaphus genome CerEl a1.0: sequencing, annotating, genes, and chromosomes. *Mol Genet Genomics.* 2018; **293** (3): 665-684
PubMed Abstract | 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?  
No

Are the datasets clearly presented in a useable and accessible format? 
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Conservation and evolutionary 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.