DATA NOTE

The genome sequence of the Norway rat, *Rattus norvegicus* Berkenhout 1769 [version 1; peer review: 2 approved]

Kerstin Howe[1], Melinda Dwinell[2], Mary Shimoyama[2+], Craig Corton[1], Emma Betteridge[1], Alexander Dove[1], Michael A. Quail[1], Michelle Smith[1], Laura Saba[3], Robert W. Williams[4], Hao Chen[5], Anne E. Kwitek[2], Shane A. McCarthy[1,6], Marcela Uliano-Silva[1], William Chow[1], Alan Tracey[1], James Torrance[1], Ying Sims[1], Richard Challis[1], Jonathan Threlfall[1], Mark Blaxter[1]

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[2]Medical College of Wisconsin, Milwaukee, Wisconsin, 53226, USA
[3]Skaggs School of Pharmacy and Pharmaceutical Sciences,, University of Colorado Anschutz Medical Center, Aurora, Colorado, 80045, USA
[4]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee, 38103, USA
[5]Department of Pharmacology, Addiction Science, and Toxicology, University of Tennessee Health Science Center, Memphis, Tennessee, 38103, USA
[6]Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

[+] Deceased author

v1 **First published:** 18 May 2021, **6**:118
https://doi.org/10.12688/wellcomeopenres.16854.1
Latest published: 18 May 2021, **6**:118
https://doi.org/10.12688/wellcomeopenres.16854.1

Abstract

We present a genome assembly from an individual male *Rattus norvegicus* (the Norway rat; Chordata; Mammalia; Rodentia; Muridae). The genome sequence is 2.44 gigabases in span. The majority of the assembly is scaffolded into 20 chromosomal pseudomolecules, with both X and Y sex chromosomes assembled. This genome assembly, mRatBN7.2, represents the new reference genome for *R. norvegicus* and has been adopted by the Genome Reference Consortium.

Keywords

Rattus norvegicus, Norway rat, genome sequence, chromosomal, reference genome

This article is included in the Tree of Life gateway.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 1 18 May 2021	✓ report	✓ report

1. **Zhihua Jiang**, Washington State University, Pullman, USA

2. **Kim Pruitt**, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Kerstin Howe (kj2@sanger.ac.uk)

Author roles: Howe K: Conceptualization, Data Curation, Formal Analysis, Investigation, Project Administration, Software, Supervision, Validation, Writing – Review & Editing; **Dwinell M**: Data Curation, Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Shimoyama M**: Data Curation, Formal Analysis, Investigation, Resources; **Corton C**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Betteridge E**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Dove A**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Quail MA**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Smith M**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Saba L**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Williams RW**: Formal Analysis, Methodology, Validation, Writing – Review & Editing; **Chen H**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Kwitek AE**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **McCarthy SA**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Uliano-Silva M**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Chow W**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Tracey A**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Torrance J**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Sims Y**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Challis R**: Formal Analysis, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Threlfall J**: Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Blaxter M**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: J. Threlfall was an employee of F1000Research up until January 2021.

Grant information: This work was supported by the Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). Maintenance of the BN/NHsdMcwi colony is supported by funding from the National Institutes of Health (NIH grants R24OD024617, DA044223) and the UTHSC Center for Integrative and Translational Genomics. SAM is supported by Wellcome (207492). Genetic marker data are available from the Rat Genome Database (NIH grant R01HL064541).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Howe K, Dwinell M, Shimoyama M *et al.* **The genome sequence of the Norway rat, *Rattus norvegicus* Berkenhout 1769 [version 1; peer review: 2 approved]** Wellcome Open Research 2021, **6**:118
https://doi.org/10.12688/wellcomeopenres.16854.1

First published: 18 May 2021, **6**:118 https://doi.org/10.12688/wellcomeopenres.16854.1

Species taxonomy

Eukaryota; Metazoa; Chordata; Mammalia; Rodentia; Muridae; Rattus; *Rattus norvegicus* Berkenhout 1769 (NCBI:txid10116).

Introduction

Rattus norvegicus is one of the most well-established experimental model organisms, with use of the species dating back to the mid-19th century (Modlinska & Pisula, 2020). The long-standing use of *R. norvegicus* in the laboratory as a model organism has led to a multitude of discoveries, providing insight into human physiology, behaviour and disease. The complexity of *R. norvegicus* relative to many other model organisms, in addition to its well-characterised physiology, means that it is frequently used in cancer research, behavioral neuroscience, and the pharmaceutical industry.

We present the reference genome mRatBN7.2 for the Norway rat, *Rattus norvegicus*. This genome assembly represents a substantial improvement on the previous assemblies, correcting areas of potential mis-assembly in the 2014 reference assembly, Rnor_6.0 (Ramdas *et al.*, 2019). The new reference has a mean genome coverage of ~92x for a single male individual of the BN/NHsdMcwi strain, which was obtained from the same colony as the original "Eve" rat that was sampled 18 years ago for use in previous rat reference genome assemblies (Eve was a female rat of generation F14, the index male described here is generation F61). The new assembly contains no gaps between scaffolds and has a scaffold N50 an order of magnitude higher than the previous reference assembly; with just 756 contigs (N50 >29 Mb), its contiguity is comparable to that of reference assemblies for humans and mice.

The production of a high-quality reference genome assembly for *R. norvegicus* allows researchers using rats for research, as a model organism for human diseases, and for determining drug interactions to have as complete and reliable a genome as possible. The result is a greater depth and certainty in data interpretation and species comparison, which will have numerous benefits for biological understanding and health.

Genome sequence report

The genome was sequenced from the kidney tissue of a single male *R. norvegicus* (strain BN/NHsdMcwi, generation F61) housed at the Medical College of Wisconsin, Milwaukee, Wisconsin, USA. A total of 80-fold coverage in Pacific Biosciences single-molecule long reads (N50, 37 kb) and 31-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 26 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data (29-fold coverage). Manual assembly curation corrected 234 missing/misjoins and removed 34 haplotypic duplications, reducing the scaffold number by 4.8%, increasing the scaffold N50 by 0.04% and decreasing the assembly length by 0.9%. The final assembly has a total length of 2.65 Gb in 219 sequence scaffolds with a scaffold N50 of 135.0 Mb (Table 1). The majority, 99.7%, of the assembly sequence was assigned to 20 chromosomal-level scaffolds representing 20 autosomes and the X and Y sex chromosomes (Figure 1–Figure 4; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) completeness of 96.2% using the mammalia_odb10 reference set. The primary assembly is a large-scale mosaic of both haplotypes (i.e. is not fully phased) and we have therefore also deposited the contigs corresponding to the alternate haplotype.

Table 1. Genome data for *R. norvegicus*.

Project accession data	
Assembly identifier	mRatBN7.2
Species	*Rattus norvegicus*
Specimen	mRatNor1
NCBI taxonomy ID	10116
BioProject	PRJNA662962
BioSample ID	SAMN16261960, SAMEA5928170
Isolate information	Laboratory animal, male, kidney tissue
Raw data accessions	
PacificBiosciences SEQUEL II	ERR5310326-ERR5310327
10X Genomics Illumina	ERR5309015-ERR5309022
Hi-C Illumina	ERR5309023, ERR5309024
BioNano	ERZ1741012
Genome assembly	
Assembly accession	GCA_015227675.2
Accession of alternate haplotype	GCA_015244455.1
Span (Mb)	2,648
Number of contigs	738
Contig N50 length (Mb)	34
Number of scaffolds	219
Scaffold N50 length (Mb)	135
Longest scaffold (Mb)	260
BUSCO* genome score	C:96.2%[S:94.0,D:2.2%],F:0.9%,M:2.8%,n:9226

*BUSCO scores based on the mammalia_odb10 BUSCO set using v5.0.0. C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Rattus%20norvegicus/dataset/JACYVU01/busco.

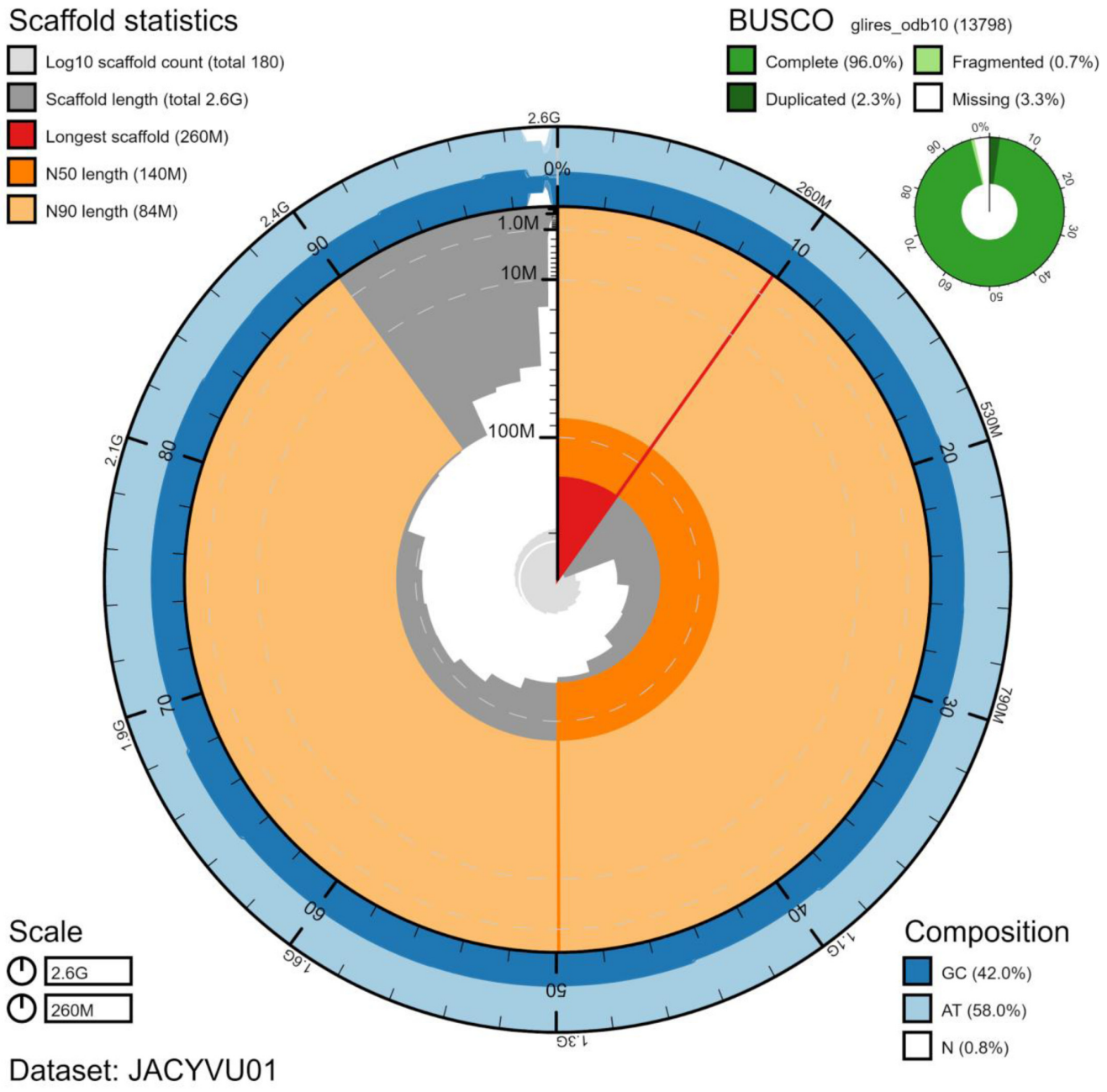

Figure 1. Genome assembly of *Rattus norvegicus*, mRatBN7.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Rattus%20norvegicus/dataset/JACYVU01/snail.

Methods

The Norway rat specimen (strain BN/NHsdMcwi, generation F61) was a male individual housed in a standard rodent microisolator cage at the Medical College of Wisconsin, Milwaukee, Wisconsin, USA. The animal was euthanised by CO_2 inhalation. This procedure was approved by the Medical College of Wisconsin Institutional Animal Care and Use Committee.

DNA was extracted using an agarose plug extraction from kidney tissue following the Bionano Prep Animal Tissue DNA Isolation Soft Tissue Protocol. Pacific Biosciences CLR long

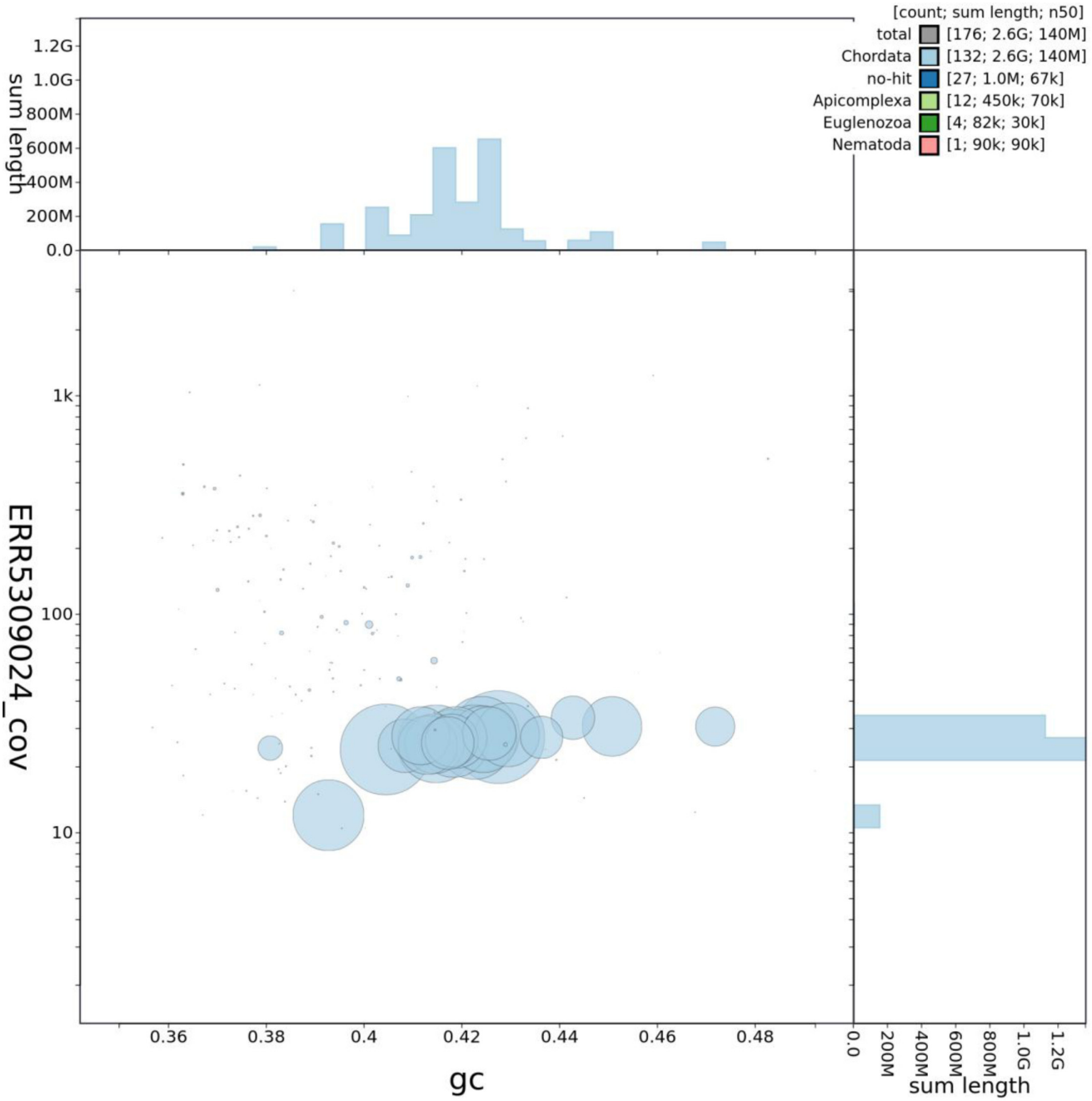

Figure 2. Genome assembly of *Rattus norvegicus*, mRatBN7.2: GC coverage. BlobToolKit GC-coverage plot. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Rattus%20norvegicus/dataset/JACYVU01/blob?plotShape=circle.

read and 10X Genomics read cloud sequencing libraries were constructed according to the manufacturers' instructions. Hi-C data were generated using the Arima v2 Hi-C kit. Sequencing was performed by the Scientific Operations DNA Pipelines at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II and Illumina HiSeq X instruments. DNA was labeled for Bionano Genomics optical mapping following the Bionano Prep Direct Label and Stain (DLS) Protocol and run on one Saphyr instrument chip flowcell.

Assembly was carried out following the Vertebrate Genome Project pipeline v1.6 (Rhie *et al.*, 2020) with Falcon-unzip (Chin *et al.*, 2016), haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020) and a first round

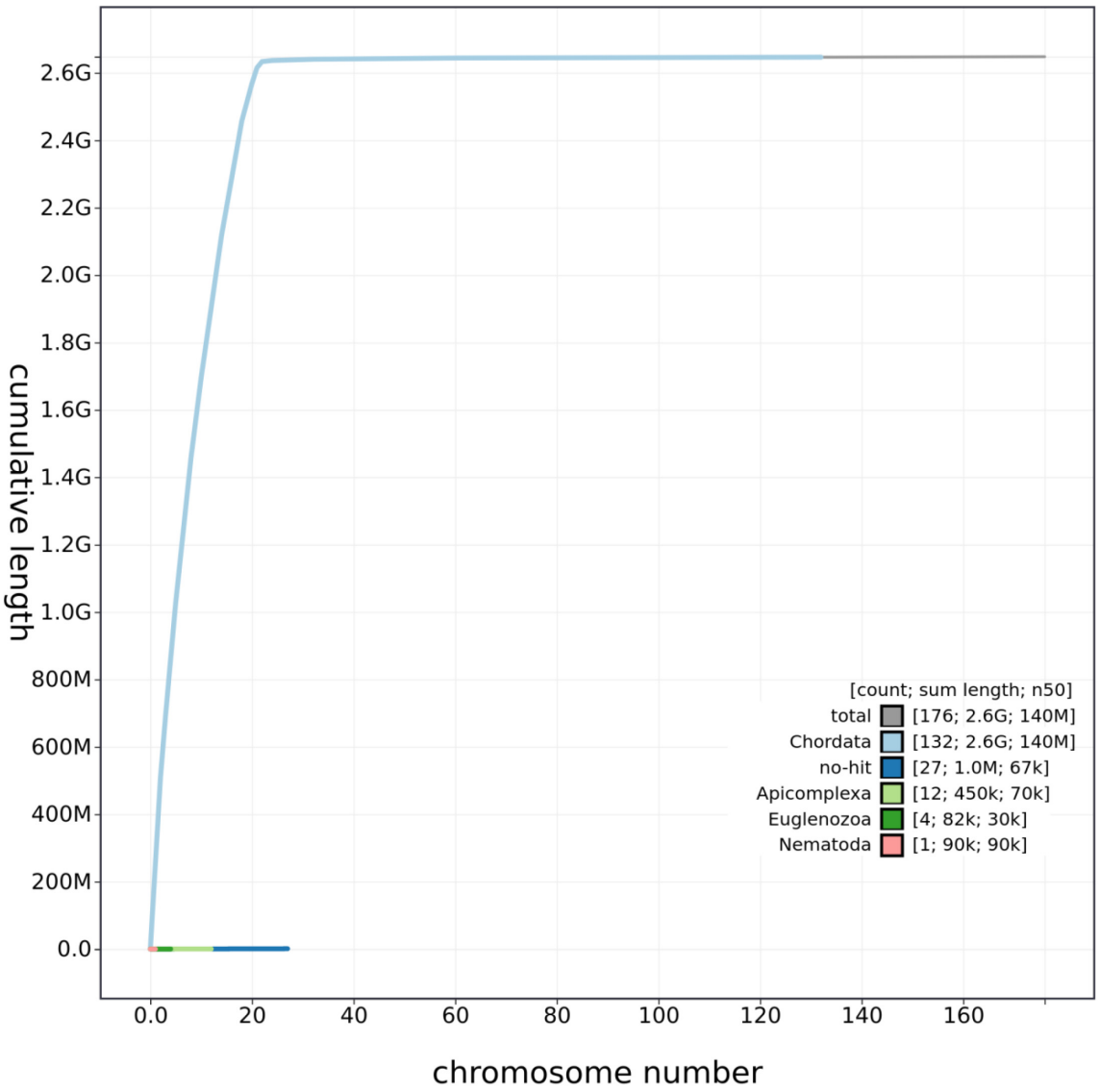

Figure 3. Genome assembly of *Rattus norvegicus*, mRatBN7.2: cumulative sequence. BlobToolKit cumulative sequence plot. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Rattus%20norvegicus/dataset/JACYVU01/cumulative.

of scaffolding carried out with 10X Genomics read clouds using scaff10x (see Table 3 for software versions and sources). Hybrid scaffolding was performed using the BioNano DLE-1 data and BioNano Solve. Scaffolding with Hi-C data (Rao *et al.*, 2014) was carried out with SALSA2 (Ghurye *et al.*, 2019). The Hi-C scaffolded assembly was polished with arrow using the PacBio data, then polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012) and applying homozygous non-reference edits using bcftools consensus. Two rounds of the Illumina polishing were applied. The assembly was checked for contamination and analysed using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, Bionano Access, HiGlass and Pretext. In addition, we used 10X *longranger* and genetic mapping data provided by LS, RWW, HC and AK to identify and resolve regions of concern. Figure 1–Figure 3 were generated using BlobToolKit (Challis *et al.*, 2020).

The mitochondrial genome was assembled as part of assembly mRatBN7.1, but was replaced with the pre-existing mitochondrial assembly MT AY172581.1, which is identical. This replacement occurred as annotation already existed for the

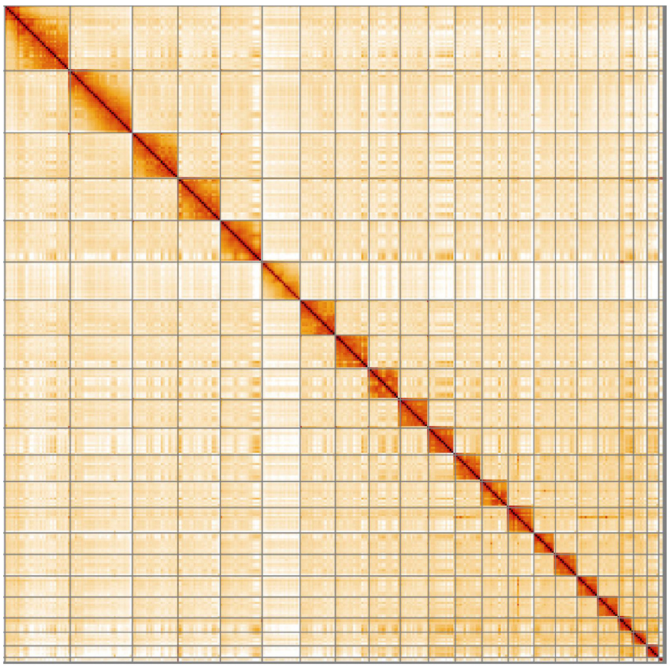

Figure 4. Genome assembly of *Rattus noevegicus*, mRatBN7.2: Hi-C contact map. Hi-C contact map of the mRatBN7.2 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the primary genome assembly of *Rattus norvegicus* mRatBN7.2.

Accession	Chromosome	Size (Mb)	GC%
CM026974.1	1	260.52	42.8
CM026975.1	2	249.05	40.5
CM026976.1	3	169.03	42.5
CM026977.1	4	182.69	41.6
CM026978.1	5	166.88	42.3
CM026979.1	6	140.99	41.9
CM026980.1	7	135.01	42.4
CM026981.1	8	123.90	43
CM026982.1	9	114.18	41.9
CM026983.1	10	107.21	45.1
CM026984.1	11	86.24	40.8
CM026985.1	12	46.67	47.2
CM026986.1	13	106.81	41.5
CM026987.1	14	104.89	41.3
CM026988.1	15	101.77	41.2
CM026989.1	16	84.73	41.8
CM026990.1	17	86.53	42.6
CM026997.1	18	83.83	41.7
CM026992.1	19	57.34	44.3
CM026993.1	20	54.44	43.7
CM026994.1	X	152.45	39.5
CM026995.1	Y	18.32	42.2

Table 3. Software tools used.

Software tool	Version	Source
Falcon-unzip	falcon-kit 1.8.0	(Chin *et al.*, 2016)
purge_dups	1.0.0	(Guan *et al.*, 2020)
Bionano Solve	Solve3.4.1_09262019	https://bionanogenomics.com/downloads/bionano-solve/
SALSA2	2.1	(Ghurye *et al.*, 2019)
scaff10x	4.2	https://github.com/wtsi-hpag/Scaff10X
arrow	GCpp-1.9.0	https://github.com/PacificBiosciences/GenomicConsensus
longranger align	longranger align (2.2.2)	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	v1.3.1-17-gaa2ace8	(Garrison & Marth, 2012)
bcftools consensus	1.11-88-g71d744f8	http://samtools.github.io/bcftools/bcftools.html
HiGlass	1.11.6	(Kerpedjiev *et al.*, 2018)
PretextView	0.0.4	https://github.com/wtsi-hpag/PretextView
gEVAL	N/A	(Chow *et al.*, 2016)
BlobToolKit	1.2	(Challis *et al.*, 2020)

pre-existing assembly. As such, the primary assembly is now mRatBN7.2.

Data availability

Underlying data

NCBI BioProject: Rattus norvegicus (Norway rat) genome assembly, mRatBN7, Accession number PRJNA662962: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA662962/

NCBI Assembly: mRatBN7.2 primary assembly, Accession number GCA_015227675.2: https://www.ncbi.nlm.nih.gov/assembly/GCF_015227675.2

NCBI Assembly: mRatBN7.1 alternate haplotype, Accession number GCA_015244455.1: https://www.ncbi.nlm.nih.gov/assembly/GCA_015244455.1

The genome sequence is released openly for reuse. The *R. norvegicus* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project and the Vertebrate Genome Project (VGP) ordinal references programme. All raw data and the assemblies have been deposited in INSDC databases under BioProject PRJNA662962. Raw data and assembly accession identifiers are reported in Table 1.

References

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

Chin CS, Peluso P, Sedlazeck FJ, *et al.*: **Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing.** *Nat Methods.* 2016; **13**(12): 1050–54.
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

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv:1207.3907. 2012.
Reference Source

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

Guan 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

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

Modlinska K, Pisula W: **The Norway Rat, from an Obnoxious Pest to a Laboratory Pet.** *eLife.* 2020; **9**: e50651.
PubMed Abstract | Publisher Full Text | Free Full Text

Ramdas S, Ozel AB, Treutelaar MK, *et al.*: **Extended Regions of Suspected Mis-Assembly in the Rat Reference Genome.** *Sci Data.* 2019; **6**(1): 39.
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): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species.** *bioRxiv.* 2020; 2020.05.22.110833.
Publisher Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 06 October 2021

https://doi.org/10.21956/wellcomeopenres.18591.r46049

✔ **Kim Pruitt**
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

The article describes generating a new reference genome assembly for *Rattus norvegicus*. This update is a significant improvement over the previous reference assembly for this organism and will provide critical support to the rat research community. The article provides structured information on the biological sample, the sequencing and assembly methods, and database identifier citations.

I have two revision requests:

1. Table 1:
- Remove BioSample ID SAMEA5928170 from the table. It is a duplicate of SAMN16261960 which is the BioSample ID linked to the primary and alternate haplotype assemblies.

2. References:
- Rhie *et al.* (2021)[1] is published now. Please update this reference to indicate the Nature citation and link to PubMed 33911273 (Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species).

- I note there is a newer BUSCO publication[2], please consider if it is more relevant to cite.

References

1. Rhie A, McCarthy SA, Fedrigo O, Damas J, et al.: Towards complete and error-free genome assemblies of all vertebrate species.*Nature*. 2021; **592** (7856): 737-746 PubMed Abstract | Publisher Full Text
2. Manni M, Berkeley MR, Seppey M, Simão FA, 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

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: My area of expertise is genome annotation, gene and sequence curation, data management, and product management.

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 28 September 2021

https://doi.org/10.21956/wellcomeopenres.18591.r46050

Zhihua Jiang
Comparative Genome Biology Program, Department of Animal Sciences, Washington State University, Pullman, WA, USA

Howe and colleagues report an updated reference genome assembly for the Norway rat, *Rattus norvegicus*. The team extracted DNA from kidney tissue collected from a male rat (BN/NHsdMcwi), and sequenced the DNA at 80x genome coverage with PacBio long-reads and at 31x genome coverage with 10X genomics short reads, followed by chromosomal confirmation of primary assembly using Hi-C reads at 29x genome coverage. Like all “Data Notes” published in *Wellcome Open Research*, the manuscript involves four core figures reporting genome assembly: 1) metrics, 2) GC coverage, 3) cumulative sequence and 4) Hi-C contact map plus three core tables demonstrating 1) genome data accession numbers, 2) chromosomal assembly information and 3) software tools used in the study. Recently, we mapped alternative polyadenylation sites to the newest reference genome and found dramatic improvements as compared to previous versions. The advanced genome resources will certainly facilitate functional annotation of the rat genome

and promote the initiation of new research fronts to understand the complicated relationships between genome and phenome for better use of the species to model health and diseases in humans.

Genome assembly nomenclature. Based on the NCBI collection, there are ten assemblies of the *Rattus norvegicus* species deposited there so far. As shown in Comment Table 1, each submitter was free to name their assembly. Rnor_6.0 and its previous versions have served as representative reference genomes for a while, which were, however, replaced by mRatBN7.2. As stated by the authors, a female from generation F14 contributed to Rnor_6.0, while a male from generation F61 was used to build the assembly of mRatBN7.2. In fact, both individuals belonged to the same colony, or the BN/NHsdMcwi strain. Perhaps that is why the authors assigned the version as 7.2, rather than 1.2, for example. My guess is that mRatBN would mean something like a male (m) rat representing Brown Norway (BN). Although the genome is indeed derived from a male, its sequences of autosomes and chromosome X can be used for any female research. As such, labeling a male-specific assembly is not necessary. In addition, the word "rat" is rather simple, because it is not specific to the *Rattus norvegicus* species. For example, the *Rattus rattus* species is the black rat, which has a nuclear genome with 18 autosomes and sex chromosomes X and Y. Therefore, I would suggest that genome assembly nomenclature be standardized for the *Rattus norvegicus* species. For example, we may use this format: Rnor_Strain (abbreviation for a strain)_xx (version number). Accordingly, mRatBN7.2 may be renamed as Rnor_BN_7.2. Hopefully, the community can discuss this further.

Genome description consistency. Generally speaking, assembly and annotation of a genome is an endless task as information evolves. Some inconsistencies need to be addressed or explained in order for the manuscript to be officially published. In terms of genome size, the authors stated that "The genome sequence is 2.44 gigabases in span" in the Abstract, but "a total length of 2.65 Gb" was presented in the Genome Sequence Report section. As shown in Comment Table 1, the latter claim is inconsistent with the NCBI report. In addition, the authors also need to double check the numbers of contigs, scaffolds and their N50 and L50 values as discrepancies exist between Table 1 (reported by the authors) and Comment Table 1 (collected from NCBI). Interestingly, the authors listed PRJNA662962 as the BioProject number, which is different from what is listed at NCBI (PRJNA677964). In fact, PRJNA662962 is not wrong either, but it contains four sub-projects: PRJNA662791, PRJNA663241, PRJNA677964 and PRJEB43118. Nevertheless, PRJNA677964 is directly linked to the assembly GCA_015227675.2. As listed in Table 2, the nuclear genome of the *Rattus norvegicus* species is split into 22 chromosomal pseudomolecules, including 20 autosomes and 2 sex chromosomes. As such, the claim on "20 chromosomal-level scaffolds representing 20 autosomes and the X and Y sex chromosomes" would certainly cause confusion.

Genome report expansion? No doubt, the current version of the manuscript strictly follows the Data Note styles so its focus is on assembly more than annotation. If possible, the team should report any changes in 1) genome structure – genes and gene-related sequences (exons, introns, UTRs and pseudogenes, for example) and intergenic DNA (genome-wide repeats and other intergenic regions) and 2) gene collection – how many genes are terminated, how many genes are renamed (based on new gene nomenclature), how many genes are overlapped and how many new genes are added to the reported assembly.

Comment Table 1. Genome assemblies deposited at NCBI for the Norway rat, *Rattus norvegicus*.

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative Genome Biology; Genome Sequencing; Functional Analysis; Alternative Transcriptome

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.