Received: 21 December 2023 | Accepted: 30 June 2024

DOI: 10.1111/age.13463

SHORT COMMUNICATION

ANIMAL GENETICS WILEY

EquCab_Finn: A new reference genome assembly for the domestic horse, Finnhorse

Abstract

Finnhorse is Finland's native and national horse breed and it has genetic affinities to northern European and Asian horses. It has historical importance for agriculture, forest work and transport and as a war horse. Finnhorse has four breeding sections in the studbook and is under conservation and characterisation efforts. We sequenced and annotated the genome of a Finnhorse mare from the working horse section using PacBio and Omni-C data. This genome can complement the existing Thoroughbred reference genome (EquCab 3.0) and facilitate genetic studies of horses from northern Eurasia. We assembled 2.4 Gb of the genome with an N50 scaffold length of 83.8 Mb and the genome annotation resulted in a total of 19 748 protein coding genes of which 1200 were Finnhorse specific. The assembly has high quality and synteny with the current horse reference genome. We manually curated five genes of interest and deposited the final assembly in the European Nucleotide Archive under the accession no. PRJEB71364.

The first reference genome for the domestic horse was published nearly 15 years ago (Wade et al., 2009) and further improved in recent years (Jevit et al., 2023; Kalbfleisch et al., 2018). Reference genomes and the advancement in modern genomic tools have greatly advanced the equine research, especially in identifying novel genetic variants and the genetic basis of several Mendelian and complex traits (Abri et al., 2020; Raudsepp et al., 2019). Given the remarkable diversity of this species, there is a need for a pangenomic approach where genomes from several breeds could be utilised (Clark et al., 2020). The current reference genome for domestic horse is from the Thoroughbred, which is closely related to other Warmblood breeds, such as the Swiss Warmblood, Hanoverian and American Quarter Horse. However, the Finnhorse is genetically distinct from the Thoroughbred and shows genetic affinity with several northern European native horse breeds, such as the Icelandic horse, Norwegian Fjord horse and North Swedish horse, as well as northern Eurasian and Asian breeds, such as the Mezen horse from west Russia, the Yakutian horse, the Tuva horse and the Mongolian horse (Petersen et al., 2013; Sild et al., 2019). The genetic ancestries of the Finnhorse and eastern and Asian horses may be associated with the dispersal of Eurasian domestic steppe horses during prehistoric times (Librado et al., 2021). Here, we present the assembly and annotation of a genome from a Finnhorse mare that will also act as a good addition to the existing reference genome. We believe that our assembly is better suited to genetic studies of horses originating from northern Eurasia.

Finnhorse (*suomenhevonen* in Finnish) is a native horse breed of Finland and the national breed (Figure 1) with historical importance. During early civilisation, Finnhorses were used for agriculture, forest work and transport. They played a significant role during the Second World War. In the 1950s, there were approximately 400 000 Finnhorses in Finland. However, owing to internal migration and the adoption of motorised horsepower, the golden era for Finnhorse ended after the 1960s and currently, there are approximately 20 000 Finnhorses in Finland (mares, stallions and geldings). The studbook for Finnhorse was established in 1907 and presently has four breeding sections: trotter, riding horse, pony-type horse and draft horse (Solala, 2021). Currently, there is an ongoing effort to conserve and characterise this breed. Natural Resources Institute Finland (Luke) oversees the conservation actions regarding Finnhorse genetic resources under the Finnish National Genetic Resources Programme for Agriculture, Forestry and Fishery. As part of an effort to better characterise the breed, we have sequenced the genome of 'Tähden Piirros', a Finnhorse mare. The mare belongs to the working horse section in the Finnhorse studbook (born 14 July 2007, herd book number 2294-07T, withers height 158 cm), and has won the Finnish championship in the annually organised Finnish working horse competitions.

We used a combination of PacBio and Omni-C to assemble the genome of a Finnhorse mare. DNA samples were quantified using Qubit 2.0 Fluorometer

FIGURE 1 The new horse reference genome assembly EquCab_Finn is the annotated genome of a Finnhorse mare, Tähden Piirros (2294-07T). Finnhorse is a multipurpose horse used for trotting, riding and working. Photo by Juha Kantanen.

(Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) with the manufacturer recommended protocol and sequenced on PacBio Sequel II 8M SMRT cells. Wtdbg2 (Ruan & Li, 2020) was run to generate a primary assembly. A total of 298.7 Gbp of PacBio CLR reads were used as an input to wtdbg2 v2.5 with genome size 3.0 g, minimum read length 20 000 and minimum alignment length 8192. The blast results of the WTDBG2 output assembly against the nt database were used as input for blobtools v1.1.1 (Laetsch & Blaxter, 2017) and scaffolds identified as possible contamination were removed from the assembly. Finally, purge_dups v1.2.3 (Guan, 2023) was used to remove haplotigs and contig overlaps.

The Omni-C library was prepared using the Dovetail Omni-C Proximity Ligation Assay (Dovetail Genomics, Scotts Valley, CS, USA) following the manufacturer's protocol and sequenced on an Illumina HiSeqX platform to produce approximately 30× sequence coverage. The draft PacBio assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016). Dovetail OmniC library sequences were aligned to the draft assembly using bwa (https://github.com/lh3/bwa). We performed busco analysis to assess the completeness of the assembly. Moreover, the quality of the Finnhorse genome was assessed using quast-lg v5.2.0 (Gurevich et al., 2013) using EquCab3.0 as a reference. quast-lg is a homology-based method that predicts genomic features (genes, transcripts, coding sequences [CDS]) by aligning the genome with the reference genome. A feature count is considered partially covered if the assembly contains incomplete features but has at least 100 bp of a given feature.

Repeat families found in the genome assemblies of *Equus ferus caballus* were identified *de novo* and classified using the software package repeatmodeler (version 2.0.1) (Flynn et al., 2020). repeatmodeler depends on the programs recon (version 1.08) (Bao & Eddy, 2002) and repeatscout (version 1.0.6) (Price et al., 2005) for the *de novo* identification of repeats within the genome. The custom repeat library obtained from repeatmodeler was used to discover, identify and mask the repeats in the assembly file using repeatmasker (Version 4.1.0) (Tarailo-Graovac & Chen, 2009). Coding sequences from *Bos taurus*, *Equus asinus*, *Equus caballus*, *Homo sapiens* and *Rangifer tarandus* were used to train the initial *ab initio* model for our assembly using the augustus software (version 2.5.5). Six rounds of prediction optimisation were done with the software package provided by augustus. The same coding sequences were also used to train a separate *ab initio* model for the assembly using snap (version 2006-07-28). RNA sequencing reads were mapped onto the genome using the star aligner software (version 2.7) (Dobin et al., 2013) and intron hints generated with the bam2hints tools within the augustus software (Stanke et al., 2006). maker (Cantarel et al., 2008), snap (Zaharia et al., 2011) and augustus (with intron–exon boundary hints provided from RNA-sequencing) were then used to predict genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from *B. taurus*, *E. asinus*, *E. caballus*, *H. sapiens* and *R. tarandus* to generate peptide evidence in the maker pipeline. Only genes that were predicted by both snap and augustus software were retained in the final gene sets. To help assess the quality of the gene prediction, annotation edit distance (AED) scores were generated for each of the predicted genes as part of the maker pipeline. Genes were further characterised for their putative function by performing a blast search of the peptide sequences against the UniProt database. Transfer RNA was predicted using the software trnascan-se (version 2.05) (Lowe & Eddy, 1997). The protein sequences of Finnhorse genome were compared with the protein sequences from the horse reference genome and other mammalian species namely: Arabian camel (CamDro2), cattle (ARS-UCD1.3), donkey (ASM1607732v2), human (GRCH38.p14) and sheep (ARS-UI_Ramb_v2.0) using orthofinder (Emms & Kelly, 2019).

The final assembly comprises 3749 scaffolds of which 3705 are more than 1 kb and the largest is 153 953 531 bp, which corresponds to horse chromosome 1. The sizes of individual scaffolds are relatively smaller than those of EquCab3.0 chromosomes (Table S1). Scaffolds N50 and N90 are 83.8 and 39 Mb, respectively, which are similar to the current horse reference genome EquCab3.0 (Table 1). The assembly has 319 gaps and covers over 96% of the eukaryotic genes (lineage dataset: eukaryote_odb10) according to busco analysis (Seppey et al., 2019). In total 246 out of 255 busco eukaryotic genes (busco version 4.0.5) were detected in our assembly, thus indicating the high quality of EquCab_Finn. Meanwhile, three busco genes were fragmented and six were missing. The *ab initio* genome annotation process resulted in 19 748 protein coding genes, including a total of 1176 single-exon genes. The average sequence length of the genes is 1443 bp, and the total coding region of the assembly spans 28 505 616 bp. Based on the orthology analysis (Table S2), 145 455 genes (96.3% of total) were assigned to 19 121 orthogroups of which 8007 consisted of single-copy genes and 11 102 had all species present. Out of 19 237 genes of Finnhorse, 17 685 were assigned to 13 120 orthogroups and 1552 were not assigned. Moreover, Finnhorse has 1200 (6.2%) species-specific (i.e only present in the EquCab_Finn) genes represented by 153 orthogroups. In comparison, EquCab3.0 has 278 (1.3%) species-specific genes and 35 orthogroups. Five genes (*MSTN*, *DRD4*, *THRAP3*, *PRKG1* and *TRPV3*) of interest were manually curated as part of the annotation process. These genes were selected based on their association with racing performance (*MSTN*; Hill et al., 2019), behaviour (*DRD4*; Hori et al., 2013) and adaptation to northern environments (*THRAP3*, *PRKG1* and *TRPV3*; Librado et al., 2015; Su et al., 2023), which are important characteristics relevant for the Finnhorse. The final assembly is publicly available at the European Nucleotide Archive under the accession no. PRJEB71364.

TABLE 1 Assembly statistics: Summary statistics for the new Finnhorse assembly (EquCab_Finn) and comparison with the reference genome of horse (Equcab3.0).

Statistic	EquCab_Finn	EquCab3.0
Assembly length (bp)	2 373 605 502	2 506 949 475
Number of scaffolds	3749	4700
Scaffold N50 (bp)	83 774 284	87 230 776
Scaffold L50	12	12
Scaffold N90 (bp)	39 010 579	40 254 690
Scaffold L90	27	

We performed pairwise alignment of our assembly with the EquCab3.0 using minimap2 (Li, 2018) and visualised the consistency plot using jupiterplot (Chu, 2018). The genome alignment of our assembly with EquCab3.0 indicated high synteny between the assemblies. The top 32 scaffolds from EquCab_Finn mapped to the 31 autosomes and X chromosome of Equcab3.0 on a 1:1 ratio (Figure 2).

The quast-lg analysis indicated that 93.01% of EquCab_Finn assembly mapped to the EquCab3.0 genome assembly, encompassing a total alignment length of 2 332 061 753 bp (Table 2). A total of 35 153 637 bp were unaligned, or 974 contigs could not be aligned to the EquCab3.0 genome, and 2143 contigs could only be partially aligned (Table 2), which represents the discrepancy between the total length of the Finnhorse genome and the total aligned length to the reference EquCab3.0 genome. Partially aligned and unaligned contigs could have resulted from structural variations between the Finnhorse genome and the reference EquCab3.0 genome, such as large indels (insertion/deletions), as well as repetitive regions and/or alternative haplotypes causing assembly errors.

The Finnhorse genome assembly was found to have 2 239 539 (94.99%) complete and 20 345 (0.86%) partial features out of the 2 357 683 genomic feature (genes, transcripts and CDS) annotations of the reference assembly in EquCab3.0 (Table 2). This indicates that the Finnhorse genome assembly is comparable with the EquCab3.0 reference genome in terms of genomic feature annotations. Moreover, homology-based gene prediction identified 20 184 genes in the Finnhorse genome assembly, representing 94.02% of the genes annotated in EquCab3.0 ($n = 21 468$) (Table 2). Of these, 87.51% (17664) were complete and 12.49% (2520) were partial, probably reflecting the level of fragmentation of the Finnhorse genome assembly. Homology-based gene predictions performed on the Finnhorse genome assembly ($n = 20 184$) based on EquaCab3.0 annotated

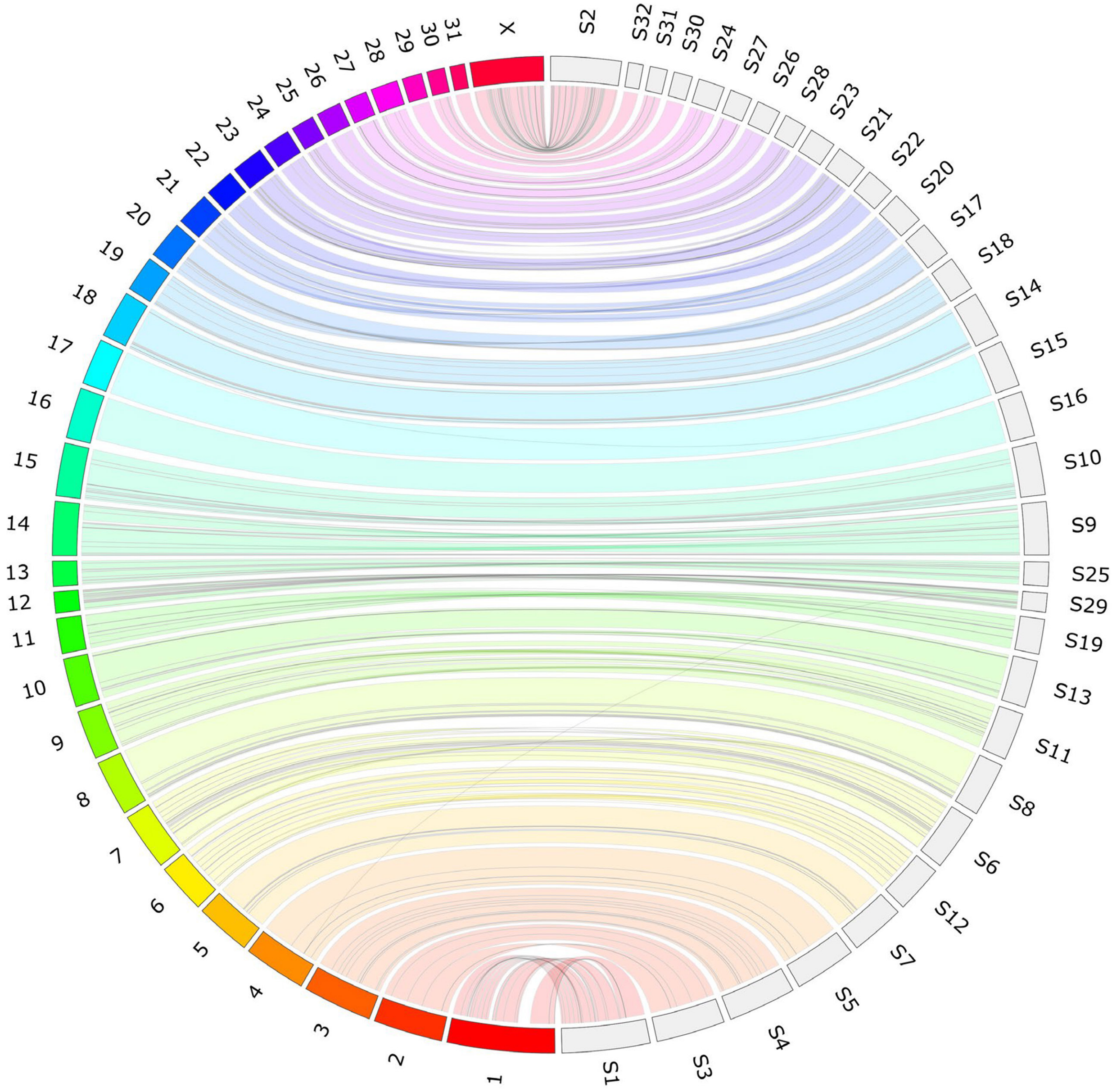

FIGURE 2 Jupiter consistency plot: genome alignment between the Finnhorse assembly EquCab_Finn and EquCab3.0. The top 32 scaffolds represent 95% of the assembly and cover the entire chromosomes of EquCab3.0. Please note in the legend that numbers 1, 2, 3, … denote EquCab3.0 chromosomes and S1, S2, S3, … denote EquCab_Finn scaffolds.

TABLE 2 QUAST-LG statistics of EquCab_Finn based on the pairwise alignment with EquCab3.0.

Assembly quality metrics of Finnhorse						
Genome fraction percentage	Total aligned length	Largest alignment	Number of fully unaligned contigs	Fully unaligned length	Number of partially unaligned contigs	Partially unaligned length
93.01	2 332 061 753	44 483 216	974	8 745 623	2143	26 408 014
Genes predicted with homology-based prediction method						
Genes	Partial genes	Total	Percentage of reference's annotated genes (n=21 468)			
17 664	2520	20 184	94.02			
Genomic features predicted with homology-based prediction method						
Number of complete genomic features	Number of partial genomic features	Total	Percentage of reference's genomic features (n=2 357 683)			
2 239 539	20 345	2 259 884	95.85			

genome were in agreement with *ab initio* genome annotation (n = 19 748) and indicated a good level of genome completeness.

In summary, we present the first complete genome assembly of Finnhorse, a native horse breed from Finland. The genome was sequenced using a combination of ILLUMINA and PACBIO technologies and assembled using WTBG2 and HIRISE assemblers. The resulting assembly consists of 2 373 605 502 bp and 3749 scaffolds, with a scaffold N50 of 83.8 Mb. The final annotated assembly contains 19 748 protein-coding genes. The Finnhorse genome is publicly available at the European Nucleotide Archive under the accession PRJEB71364. This genome provides a valuable resource for studying the genetic diversity and evolution of horses, especially those from northern Europe, and will facilitate future pangenome analyses of equine species.

KEYWORDS

Finnhorse, reference genome

AUTHOR CONTRIBUTIONS

Kisun Pokharel: Data curation; formal analysis; methodology; visualization; writing – original draft; writing – review and editing. **Melak Weldenegodguad:** Data curation; formal analysis; methodology; writing – review and editing. **Tiina Reilas:** Investigation; methodology; resources; writing – review and editing. **Juha Kantanen:** Conceptualization; funding acquisition; investigation; project administration; resources; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the funding provided by the Ministry of Agriculture and Forestry in Finland. We thank the owner of Tähden Piirros, Mr Kauko Tuominen, for collaboration in blood sampling.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

ETHICS STATEMENT

Animal handling procedures and sample collections were performed in accordance with the legislation approved by the Animal Experiment Board in Finland (ESAVI/7034/04.10.05.2015).

DATA AVAILABILITY STATEMENT

The final assembly is publicly available at the European Nucleotide Archive under the accession no. PRJEB71364.

Kisun Pokharel[1]
Melak Weldenegodguad[2]
Tiina Reilas[1]
Juha Kantanen[1]

[1]*Natural Resources Institute Finland (Luke), Jokioinen, Finland*
[2]*Natural Resources Institute Finland (Luke), Helsinki, Finland*

Correspondence
Juha Kantanen, Natural Resources Institute Finland (Luke), Tietotie 4, FI-31600 Jokioinen, Finland.
Email: juha.kantanen@luke.fi

ORCID

Kisun Pokharel https://orcid.org/0000-0002-4924-946X
Melak Weldenegodguad https://orcid.org/0000-0002-2876-6353
Juha Kantanen https://orcid.org/0000-0001-6350-6373

REFERENCES

Abri, M.A.A., Holl, H.M., Kalla, S.E., Sutter, N.B. & Brooks, S.A. (2020) Whole genome detection of sequence and structural polymorphism in six diverse horses. *PLoS One*, 15, e0230899. Available from: https://doi.org/10.1371/journal.pone.0230899

Bao, Z. & Eddy, S.R. (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, 12, 1269–1276. Available from: https://doi.org/10.1101/gr.88502

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B. et al. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18, 188–196. Available from: https://doi.org/10.1101/gr.6743907

Chu, J. (2018) *Jupiter plot: A circos-based tool to visualize genome assembly consistency*. https://doi.org/10.5281/zenodo.1241235

Clark, E.L., Archibald, A.L., Daetwyler, H.D., Groenen, M.A.M., Harrison, P.W., Houston, R.D. et al. (2020) From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biology*, 21, 285. Available from: https://doi.org/10.1186/s13059-020-02197-8

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21. Available from: https://doi.org/10.1093/bioinformatics/bts635

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 238. Available from: https://doi.org/10.1186/s13059-019-1832-y

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C. et al. (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 9451–9457. Available from: https://doi.org/10.1073/pnas.1921046117

Guan, D. (2023) *Purge_Dups*. Available from: https://github.com/dfguan/purge_dups [Accessed 20th December 2023].

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072–1075. Available from: https://doi.org/10.1093/bioinformatics/btt086

Hill, E.W., McGivney, B.A., Rooney, M.F., Katz, L.M., Parnell, A. & MacHugh, D.E. (2019) The contribution of myostatin (MSTN) and additional modifying genetic loci to race distance aptitude in thoroughbred horses racing in different geographic regions. *Equine Veterinary Journal*, 51, 625–633. Available from: https://doi.org/10.1111/evj.13058

Hori, Y., Ozaki, T., Yamada, Y., Tozaki, T., Kim, H.-S., Takimoto, A. et al. (2013) Breed differences in dopamine receptor D4 gene

(DRD4) in horses. *Journal of Equine Science*, 24, 31–36. Available from: https://doi.org/10.1294/jes.24.31

Jevit, M.J., Castaneda, C., Paria, N., Das, P.J., Miller, D., Antczak, D.F. et al. (2023) Trio-binning of a hinny refines the comparative organization of the horse and donkey X chromosomes and reveals novel species-specific features. *Scientific Reports*, 13, 20180. Available from: https://doi.org/10.1038/s41598-023-47583-x

Kalbfleisch, T.S., Rice, E.S., DePriest, M.S., Walenz, B.P., Hestand, M.S., Vermeesch, J.R. et al. (2018) Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, 1, 197. Available from: https://doi.org/10.1038/s42003-018-0199-z

Laetsch, D.R. & Blaxter, M.L. (2017) *BlobTools: Interrogation of genome assemblies*. https://doi.org/10.12688/f1000research.12232.1

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100. Available from: https://doi.org/10.1093/bioinformatics/bty191

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A. et al. (2015) Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America*, 112, E6889–E6897. Available from: https://doi.org/10.1073/pnas.1513696112

Librado, P., Khan, N., Fages, A., Kusliy, M.A., Suchan, T., Tonasso-Calvière, L. et al. (2021) The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*, 598, 634–640. Available from: https://doi.org/10.1038/s41586-021-04018-9

Lowe, T.M. & Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25, 955–964. Available from: https://doi.org/10.1093/nar/25.5.0955

Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., Bailey, E. et al. (2013) Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*, 8, e54997. Available from: https://doi.org/10.1371/journal.pone.0054997

Price, A.L., Jones, N.C. & Pevzner, P.A. (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics*, 21(Suppl 1), i351–i358. Available from: https://doi.org/10.1093/bioinformatics/bti1018

Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R. et al. (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, 26, 342–350. Available from: https://doi.org/10.1101/gr.193474.115

Raudsepp, T., Finno, C.J., Bellone, R.R. & Petersen, J.L. (2019) Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Animal Genetics*, 50, 569–597. Available from: https://doi.org/10.1111/age.12857

Ruan, J. & Li, H. (2020) Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, 17, 155–158. Available from: https://doi.org/10.1038/s41592-019-0669-3

Seppey, M., Manni, M. & Zdobnov, E.M. (2019) BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962, 227–245. Available from: https://doi.org/10.1007/978-1-4939-9173-0_14

Sild, E., Rooni, K., Värv, S., Røed, K., Popov, R., Kantanen, J. et al. (2019) Genetic diversity of Estonian horse breeds and their genetic affinity to northern European and some Asian breeds. *Livestock Science*, 220, 57–66. Available from: https://doi.org/10.1016/j.livsci.2018.12.006

Solala, H. (2021) *Suomalaisen hevosrodun synty: Maatiaishevonen ja kotieläinjalostuksen kansainvälinen murros 1893–1907*. Tampere University. Available from: https://trepo.tuni.fi/handle/10024/133436 [Accessed 18th December 2023].

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. & Morgenstern, B. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34, W435–W439. Available from: https://doi.org/10.1093/nar/gkl200

Su, W., Qiao, X., Wang, W., He, S., Liang, K. & Hong, X. (2023) TRPV3: structure, diseases and modulators. *Molecules*, 28, 774. Available from: https://doi.org/10.3390/molecules28020774

Tarailo-Graovac, M. & Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 25, 4.10.1–4.10.14. Available from: https://doi.org/10.1002/0471250953.bi0410s25

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F. et al. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, 326, 865–867. Available from: https://doi.org/10.1126/science.1178158

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S. et al. (2011) *Faster and more accurate sequence alignment with SNAP*. ArXiv11115572 Cs Q-Bio. Available from: http://arxiv.org/abs/1111.5572 [Accessed 15th June 2020].

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.