The genome of Przewalski's horse (*Equus ferus przewalskii)*

Nicole Flack[1†], Lauren Hughes[2†], Jacob Cassens[3], Maya Enriquez[4], Samrawit Gebeyehu[5], Mohammed Alshagawi[4], Jason Hatfield[4], Anna Kauffman[4], Baylor Brown[4], Caitlin Klaeui[4], Islam F. Mabrouk[4], Carrie Walls[5], Taylor Yeater[4], Anne Rivas[6], Christopher Faulk[5*]

1 Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota Saint Paul, MN, USA
2 Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota Saint Paul, MN, USA
3 Division of Environmental Health Sciences, School of Public Health, University of Minnesota Minneapolis, MN, USA
4 ANSC 8520 Students, University of Minnesota Minneapolis, MN, USA
5 Department of Animal Science, College of Food, Agricultural and Natural Resource Sciences, University of Minnesota Saint Paul, MN, USA
6 Minnesota Zoo Apple Valley, MN, USA

†These authors contributed equally.

* Corresponding author. Email: cfaulk@umn.edu

Keywords
horse, Przewalksi, nanopore, genome assembly, DNA methylation

Running Title
The Genome of Przewalski's Horse

Abstract

The Przewalski's horse (*Equus ferus przewalskii*) is an endangered equid native to the steppes of central Asia. After becoming extinct in the wild, multiple conservation efforts convened to preserve the species including captive breeding programs, reintroduction and monitoring systems, protected lands, and cloning. Availability of a highly contiguous reference genome is essential to support these continued efforts. We used Oxford Nanopore sequencing to produce a scaffold-level 2.5 Gb nuclear assembly and 16,002 bp mitogenome from a captive Przewalski's mare. All assembly drafts were generated from 111 Gb of sequence from a single PromethION R10.4.1 flow cell. The mitogenome contained 37 genes in the standard mammalian configuration and was 99.63% identical to the domestic horse (*Equus caballus*). The nuclear assembly, EquPr2, contained 2,146 scaffolds with an N50 of 85.1 Mb, 43X mean depth, and BUSCO quality score of 98.92%. EquPr2 successfully improves upon the existing Przewalski's horse reference genome (Burgud), with 25-fold fewer scaffolds, a 166-fold larger N50, and phased pseudohaplotypes. Modified basecalls revealed 79.5% DNA methylation and 2.1%

hydroxymethylation globally. Allele-specific methylation analysis between pseudohaplotypes revealed 226 differentially methylated regions (DMRs) in known imprinted genes and loci not previously reported as imprinted. The heterozygosity rate of 0.165% matches previous estimates for the species and compares favorably to other endangered animals. This improved Przewalski's horse assembly will serve as a valuable resource for conservation efforts and comparative genomics investigations.

Introduction

The Przewalski's horse (*Equus ferus przewalskii*), also called the tahki, is an endangered equid native to central Asia [1] whose lineage diverged from the domestic horse (*Equus caballus*) tens of thousands of years ago [2, 3, 4, 3, 5]. *E. f. przewalskii* has a distinct short and stocky build with dun coloring and an erect mane. These horses were initially native to the steppes of central Asia, and by the 19th century inhabited only Mongolia, Tibet, and China [6, 7]. Introgression from the domestic horse, loss of natural habitat and resources, harsh climates, and hunting contributed to a severe population bottleneck and subsequent extinction in the wild in the 1960s [7, 8]. All current Przewalski's horses are descendants of 12 wild-caught individuals and several domesticated horses [6, 7, 9]. Massive conservation efforts have focused on preserving the genetic diversity of this endangered species [10, 11, 12, 13, 14, 15].

Targeted captive breeding and management programs, including published studbooks since 1959, have increased the Przewalski's horse population to over 2,000 individuals [14, 9, 3]. Reintroduction efforts starting in the 1980s have successfully established wild herds in protected lands of Mongolia, China and Kazakhstan [16, 9, 15, 17, 18, 19, 20, 21], and have improved the species' status from critically endangered to endangered [1]. Recent cloning efforts through the Przewalski's Revive & Restore Project have produced two males from a cryopreserved cell line in the San Diego Zoo Wildlife Alliance Frozen Zoo [22, 23, 24], further bolstering captive breeding efforts.

Przewalski's horses possess additional chromosomes (2n=66) when compared to the domestic horse (2n=64) [25, 26, 27, 28]; this chromosome difference is thought to be derived from a Robertsonian translocation event [29, 30]. Przewalski's and domestic horse crosses produce viable, fertile offspring with odd numbered chromosomes (n=65), in contrast to the infertile offspring of domestic horse and donkey (*Equus asinus*) crosses [31]. Investigation of the differences between domestic and Przewalski's horse chromosome structures may facilitate improved understanding of chromosome fusion.

Recent technological advances have increased accessibility and reduced costs for generating high-quality whole genome sequencing data [32]. Long-read Oxford Nanopore sequencing provides genomic and epigenomic data simultaneously without amplification or bisulfite conversion [33]. Two major limitations of the technology, high per-base error rates and homopolymer indels, have been addressed with improved basecalling models, self-correction algorithms, and adequate depth of coverage [34]. In addition to direct capture of epigenetic base modifications, pseudohaplotype phasing permits heterozygosity estimation and allele-specific DNA methylation analysis; strict allele-specific methylation is a signature of genomic imprinting, where genes are monoallelically expressed based on parent of origin [35, 36].

The epigenome is also relevant to species evolutionary biology due to the increased mutation rate of methylated cytosines via deamination [37, 38, 39, 40, 41, 42].

Here, we highlight the exclusive use of Oxford Nanopore sequencing reads to provide a high-quality, highly contiguous diploid nuclear genome assembly, updated mitogenome, and DNA methylation analysis for the endangered Przewalski's horse.

Methods

Sample collection

Varuschka, a 10-year-old captive-bred Przewalski's mare (Figure 1), was subject to routine veterinary care under anesthesia during which 10 ml of whole blood was collected by zoo veterinarians. The Minnesota Zoo has been active in Przewalski's horse breeding and management, with over 50 foals born since the 1970s, and contributed a stallion to reintroduction efforts in Mongolia's Hustai National Park [43, 44].

Figure 1. *Equus ferus przewalskii* specimen photo depicting Varuschka, the 10-year-old captive-bred mare sampled for genome assembly. Image courtesy of the Minnesota Zoo.

DNA extraction and sequencing

Genomic DNA was extracted from blood using a MagAttract Blood DNA/RNA Kit (Qiagen, Venlo, Netherlands) according to manufacturer's instructions. Sequencing was performed on a P2 Solo instrument (Oxford Nanopore Technologies, Oxford, UK) using a single PromethION R10.4.1 flow cell. We created two libraries using the LSK-114 ligation sequencing kit for native DNA sequencing. For the first library, 3 µg of DNA in 100 µl elution buffer was sheared by passage through a 28 gauge needle 30 times, library prepped, then split into three aliquots of 15 µl each. For the second library, unsheared DNA was eluted and prepped as a single 15 µl library. The first aliquot of the sheared library was loaded onto the flowcell and run for 24 hours, after which the flowcell was washed using the manufacturer's wash kit. The remaining two sheared aliquots were loaded and sequenced in the same fashion on days 2 and 3, respectively. After sequencing the three sheared library aliquots, the unsheared library was loaded into the flowcell and run for 24 hours. Data were collected using 5 kHz minKNOW version 23.07.12 (Oxford Nanopore Technologies, Oxford, UK).

Raw nanopore data from both libraries were basecalled together using Dorado v0.4.3 (https://github.com/nanoporetech/dorado) with "super accuracy" model dna_r10.4.1_e8.2_400bps_sup@v4.2.0; base modifications called simultaneously with flag `--modified-bases 5mC_5hmC`. Read quality was assessed using the Nanoq package (https://github.com/esteinig/nanoq).

Genome assembly

Detailed computational methods are available in Supplementary File 2. The genome was *de novo* assembled using Flye v2.9 [45], followed by polishing using Medaka v1.6.0 (https://github.com/nanoporetech/medaka). Duplicate contigs were removed using Purge Dups v1.2.6 (https://github.com/dfguan/purge_dups). Additional manual curation was performed to remove contigs with less than 15X or greater than 500X mean coverage. One contig representing the mitochondrial genome was also removed at this stage. The resulting draft assembly was scaffolded onto the reference domestic horse genome, EquCab3.0 (GCF_002863925.1), using the RagTag package v2.1.0 [46]. Gaps were closed using TGS-GapCloser v1.2.1 [47].

Scaffolds for *E. ferus przewalskii* were assigned chromosome names based on homology to *E. caballus* synteny from EquCab3.0 as demonstrated previously [30]. The homologous *E. f. przewalskii* scaffold to *E. caballus* chromosome 5 was split on its scaffold gaps with AWK, and the largest blocks syntenic to the p- and q-arms were labeled with their chromosome names (Supplementary File 1). The remaining smaller contigs syntenic to EquCab3.0 chromosome 5 were labeled as ChrUns in the scaffolded assembly. The FCS-adapter tool from the NCBI Foreign Contamination Screening program suite was used to detect and remove adapter and vector contamination from the final haplotype assemblies (https://github.com/ncbi/fcs).

The quality of all draft assemblies was evaluated by detecting Benchmarking Universal Single-Copy Ortholog (BUSCO) genes within the *Cetartiodactyla* lineage [48]. We used the Compleasm package to calculate BUSCO scores as it is a faster, more accurate implementation of BUSCO [49]. We combined Compleasm's single and duplicate BUSCO counts to provide a direct comparison to the standard BUSCO program's 'complete' value. On average, Compleasm

detected 3% more BUSCOs than the BUSCO tool v3.4.7 (Supplementary File 1). We calculated assembly N50, L50, and other statistics using the packages Assembly-Stats (https://github.com/sanger-pathogens/assembly-stats) and the Quality Assessment Tool for Genome Assemblies (QUAST v5.2.0) [50], using EquCab3.0 as reference for the latter. Assembly statistics were compared to other publicly available equid genomes including the preexisting Przewalski's horse (Burgud), domestic horse (EquCab3.0), and plains zebra genome (*Equus quagga*; UCLA_HA_Equagga_1.0) references.

Methylation

Global DNA methylation and hydroxymethylation (5mC and 5hmC) at cytosine-guanine dinucleotides (CpGs) was determined using the modified base information stored in the basecalled BAM files, i.e., modBAMs. Since the original basecalling was performed prior to a reference existing, Dorado stored modified bases per-read as unmapped modBAMs. These were concatenated together and aligned to the final scaffold-level assembly using Minimap2 [51, 52]. The resulting mapped modBAMs were converted to bedMethyl format using Modkit v0.2.2 (https://github.com/nanoporetech/modkit) with global 5mC and 5hmC percentages summarized using AWK (https://www.gnu.org/software/gawk/manual/gawk.html). For allele-specific DNA methylation, Modkit was applied to the phased modBAM generated by variant calling to generate two bedMethyl files separated by haplotag. Count filtering, normalization, tiling, and differential methylation testing were performed with MethylKit [53, 54]. Nearest genes to significant regions were found with Bedtools [55]. Differentially methylated regions (DMRs) between haplotypes were visualized with Methylartist [56].

Repeats

RepeatMasker v4.1.4 (https://www.repeatmasker.org/) was used to identify repetitive sequences with the complete Dfam library v3.6 (https://www.dfam.org/home) as described previously [57, 58]. For consistency of comparison, Repeatmasker was also run locally on the existing equid reference genomes with the same parameters.

Variant calling and diploidization

Clair3 v1.05 [59] was run to determine the number and type of variants. Heterozygosity was calculated by counting the number of variants divided by the total genome size. Variants were phased with Whatshap [60] and haplotagged with Longphase [61]. BCFTools was used to swap out the phased variants in the primary assembly to generate the secondary pseudohaplotype assembly (https://github.com/samtools/bcftools).

Gene annotation

Homology-based gene prediction was performed with Gene Model Mapper (GeMoMa, https://doi.org/10.1007/978-1-4939-9173-0_9) using EquCab3.0 transcripts as the reference. BUSCO was used in protein mode to assess gene prediction accuracy and completeness.

Mitochondrial assembly

The mitogenome was extracted from the *E. ferus przewalskii* assembly using MitoHiFi v3.2 [62, 63]. The program identifies mitogenome contigs by comparison to known mitogenomes from related species; in this case we used the EquCab3.0 mitogenome (NC_001640.1). MitoHiFi also circularizes and annotates the putative mitogenome contig.

Results and Discussion

Assembly

We generated a total of 111 Gb of DNA sequencing data for Przewalski's horse with a read N50 of 10,829 bp and mean quality of Q18.49. The quality of each draft assembly was assessed with parameters including N50 (i.e., length of the shortest contig at 50% of the total assembly length), L50 (i.e., smallest number of contigs whose length sum to 50% of the total assembly length), and the count of benchmark universal single-copy ortholog (BUSCO) genes. Our initial Flye run yielded a 2.59 Gb draft assembly with 6,808 contigs and an N50 of 13.6 Mb (Table 1). With an L50 of 55, the majority of the genome was assembled into relatively few contigs representing large portions of chromosomes.

Table 1. Contiguity and quality statistics for draft *E. ferus przewalskii* assemblies, final EquPr2 assemblies, and existing equid reference genomes. Benchmark universal single-copy ortholog (BUSCO) scores were calculated with the Compleasm tool [49]. HP: pseudohaplotype.

Draft	Size (Gb)	Contigs/ Scaffolds	N50 (Mb)	L50	BUSCO complete	BUSCO single	BUSCO duplicate	Gaps (n)
Przewalski's horse (Burgud)	2.396	53,097	0.514	1,223	89.88%	89.51%	0.37%	80,963
Domestic horse (EquCab3.0)	2.507	4,700	87.231	12	98.92%	97.74%	1.18%	6,294
Zebra (UCLA_HA-_Equagga-_1.0)	2.501	12,506	122.556	9	98.47%	95.13%	3.34%	14,637
Initial assembly (Flye)	2.594	6,808	13.697	55	98.84%	98.17%	0.67%	0
Polished (Medaka)	2.594	6,808	13.696	55	98.85%	98.19%	0.66%	0

Draft	Size (Gb)	Contigs/ Scaffolds	N50 (Mb)	L50	BUSCO complete	BUSCO single	BUSCO duplicate	Gaps (n)
Purged (Purge_dups)	2.503	3,411	13.974	52	98.77%	98.22%	0.55%	0
Curated	2.498	2,796	13.976	52	98.77%	98.22%	0.55%	0
EquPr2 HP1	2.500	2,146	85.15	12	98.92%	98.37%	0.55%	306
EquPr2 HP2	2.499	2,146	85.13	12	98.92%	98.35%	0.57%	306

The initial assembly was polished, purged of duplicates, and manually curated to remove contigs with coveage below 15X and 500X as they are unlikely to represent single-copy nuclear regions [64]. This procedure resulted in the removal of 4,012 contigs and 96 Mb of sequence. Polishing with Medaka increased the complete BUSCO score by reducing the percentage of duplicate BUSCOs by 0.01%. In our initial draft, the duplicate BUSCO count was already low at less than 1% of the total; haplotig purging reduced the number of duplicate BUSCOs by 0.11% and increased N50 from 13.70 Mb to 13.97 Mb.

We scaffolded the curated assembly onto the domestic horse genome (EquCab3.0); gaps were filled with reads placed by TGS-GapCloser where spanning reads could be identified. This final assembly, EquPr2, was 2.50 Gb in length with a chromosome-level N50 of 85.1 Mb and L50 of 12. EquPr2 contained 306 gaps spanning 2.03 Mb, a 5-fold and 11-fold reduction in gap length compared to EquCab3.0 and Burgud, respectively (Table 1). The contiguity of EquPr2 also matches EquCab3.0's L50 of 12, in contrast to Burgud's L50 of 1,223 (Figure 2).

Our Przewalski's horse genome had a BUSCO completeness score of 98.92%, improving from 89.88% in the previous Przewalski's horse assembly with 25-fold fewer scaffolds and a 166-fold increase in N50. It is important to deconstruct the complete BUSCO score reported by most new genome assemblies into its component parts of single and duplicate copy percentages. Examining only the completeness score obscures the presence of haplotig missassemblies represented by high duplicate counts. For instance, the reference horse genome, EquCab3.0, is a high-quality, highly contiguous assembly with a complete BUSCO score of 98.92%, the same as our Przewalski's horse assembly EquPr2. However EquCab3.0 has a duplicate rate of 1.18% versus EquPr2's duplicate rate of 0.55%. Similarly, the plains zebra genome, UCLA_HA_Equagga1.0, is even more contiguous with an L50 of 9 but has a 3.34% BUSCO duplication rate. While these differences are relatively small, duplication misassemblies have contributed to erroneous gene gain and gene family expansion findings in high-heterozygosity vertebrate genomes [65].

The Przewalski's horse has 33 sets of chromosomes versus the domestic horse's 32 due to a Robertsonian translocation where *E. caballus* chromosome 5 is homologous to *E. f. przewalskii* chromosomes 23 and 24 [30]. Other EquPr2 chromosomes were named based on homology to domestic horse consistent with previous karyotyping [66, 67]. We manually split the EquPr2 scaffold homologous to *E. caballus* chromosome 5 at every gap from the contig-level assembly and named the largest contigs as chromosomes 23 and 24 based on homology to the p and q arms of its species ortholog. Due to lack of positional certainty, the remaining contigs mapping

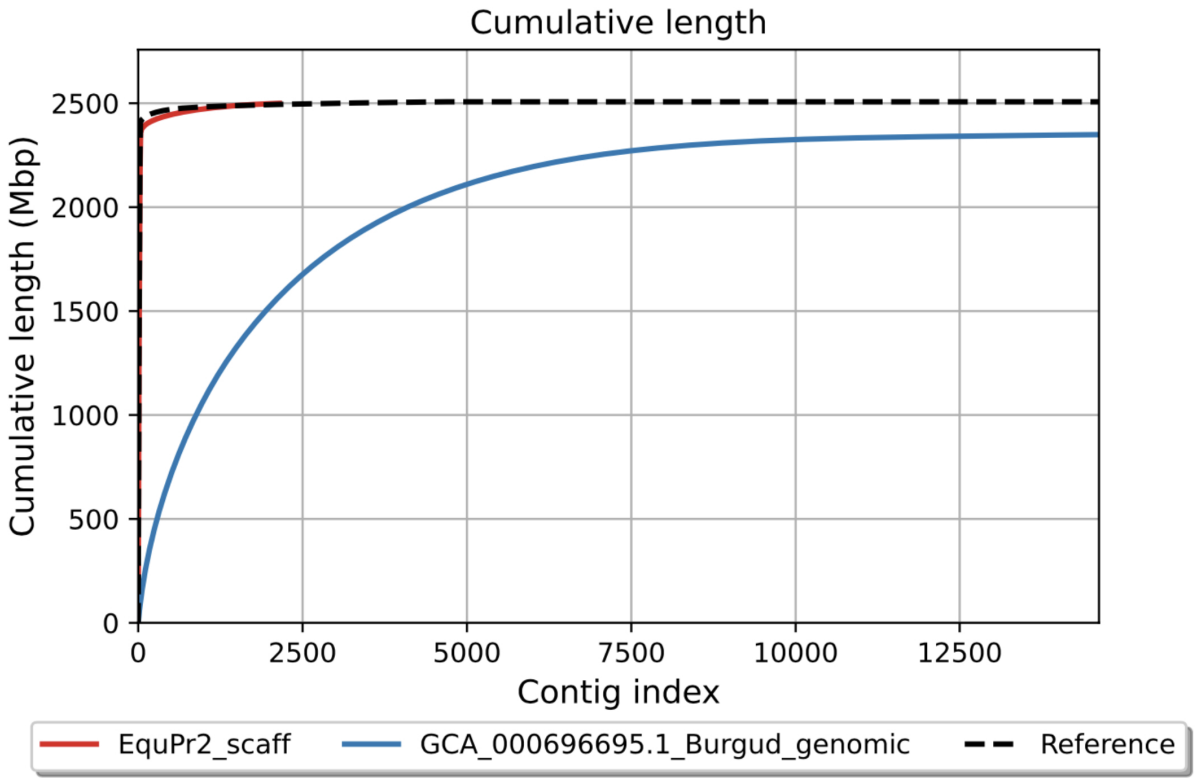

Figure 2. Cumulative length of EquPr2 scaffolds.

to EquCab3.0's chromosome 5 were named as chromosome unknown (ChrUn).

Repetitive DNA

We used RepeatMasker to detect repetitive element content within the new assembly and compared it to the previous assembly, Burgud, and the domestic horse (Table 2). Given the genetic similarity of Przewalski's horse to the domestic horse, global transposon content is nearly identical across all categories. Identification of single repeat insertions facilitated by a more contiguous Przewalski's horse reference may be valuable for future comparative genomics investigations of these two closely related species.

Table 2. Repetitive DNA content for EquPr2, the existing Przewalski's horse reference (Burgud), and the domestic horse reference (EquCab3.0). Only RepeatMasker categories reaching >1% genomic content are shown; smaller families are collapsed into the relevant parent category.

RepeatMasker	**EquPr2 (%)**	**Burgud (%)**	**EquCab3.0 (%)**
Retroelements	30.89	31.66	32.21
SINEs:	3.55	3.70	3.66
LINEs:	21.35	21.65	22.27
L2/CR1/Rex	5.26	5.47	5.43
L1/CIN4	15.88	15.96	16.63
LTRs:	6.00	6.32	6.28
Retroviral	5.57	5.88	5.84

DNA transposons	3.44	3.62	3.56
hobo-Activator	2.6	2.74	2.69
Total repeats	34.36	35.32	35.8

Heterozygosity

Given the extreme population bottleneck that occurred during the near-extinction of Przewalski's horse, it is critical to understand the genetic diversity remaining for captive breeding efforts. To this end and due to lack of samples for Varuschka's sire and dam, we called and phased EquPr2 variants to build an alternate pseudohaplotype assembly. Variant calling with Clair3 found 4,114,297 variants, the majority of which were single-nucleotide variants (SNVs). Heterozygosity was estimated to be 0.165. This level of genetic diversity is concordant with previous microarray data from nine Przewalski's horses where average heterozygosity was 0.168, the highest estimate among nine members of *Hippomorpha* excluding the domestic horse [68].

Gene annotation

We detected 21,552 putative genes in EquPr2 with the *ab initio* gene prediction tool Gene Model Mapper (GeMoMa) [69] using the EquCab3.0 annotation (GCF_002863925.1) as reference. The resulting protein BUSCO score was 85.6% complete (84.6% single copy and 1.0% duplicates); this score appears to be near GeMoMa's maximum performance based on previous demonstration of similar results for existing high-quality reference genomes [64]. The EquPr2 annotation would likely be significantly improved with by the application of NCBI's annotation pipeline, which requires the availability of RNA-seq data.

Mitochondrial genome

The mitochondrial genome was extracted from the assembly and characterized with MitoHiFi, a tool designed for long-read sequencing mitochondrial contig identification and annotation [63]. The resulting mitogenome for EquPr2 was 16,002 bp, 8 bp shorter than EquCab3.0, possibly resulting from incomplete assembly. It contained 37 genes in the standard mammalian configuration and was 99.63% identical to *E. caballus* isolate TN9488 (Supplementary File 1).

DNA methylation

Oxford Nanopore sequencing can natively detect base modifications including 5-methylcytosine (5mC). This feature has been applied in previous mammalian genome assemblies to evaluate allele-specific DNA methylation and identify known and putative novel imprinted genes [64]. Globally, we found that whole blood leukocyte DNA methylation in Przewalski's horse was 79.5% methylated and 2.1% hydroxymethylated at CpG sites genome-wide. After filtering to include CpGs with at least 10x coverage in both pseudohaplotypes, there were 18,928,678 sites available to test for allele-specific differential methylation. Counts were tiled into 100 bp

windows; windows containing at least 10 CpGs with an absolute difference in methylation >=50% and Benjamini-Hochberg-adjusted p-value < 0.05 were deemed differentially methylated regions (DMRs). With these parameters, we identified 226 DMRs between EquPr2 pseudohaplotyes with a mean absolute methylation difference of 64.1% (Figure 3). Nearest features to DMRs included known imprinted genes (e.g., *IGF2R*, *INPP5F*, *PEG3*, *DIRAS3*) and loci not previously reported as imprinted (e.g., *DLG3*). Seventy-seven DMRs (34.1%) directly overlapped predicted genes. As a consequence of random X-chromosome inactivation (XCI) in a female animal, 111 of the 226 DMRs (49.1%) were located on ChrX. Imprinted genes are unique to mammals and strongly influence growth, making this information valuable for future investigations of Przewalski's horse evolution, comparative genomics, and conservation.

Conclusions

The availability of a high-quality reference genome is imperative for improved understanding of the genetic diversity of Przewalski's horse. The lineage of the horse sampled for this paper, Varuschka, traces back to founders from Mongolia; her dam was imported from the Cologne Zoo in Germany as a part of the Species Survival Plan (SSP), and her sire was transferred from the Smithsonian Zoo. Here, we provide a 2.5 Gb nanopore-only genome assembly for Przewalski's horse with an improved BUSCO score of 98.92%, 25-fold fewer scaffolds, and a 166-fold increase in N50 compared to the existing reference genome. Modified basecalls additionally facilitated allele-specific methylation analysis; significantly differentially methylated regions included known imprinted genes and potential novel loci. This genome will aid Przewalski's horse conservation by providing a higher quality, more contiguous foundation for captive breeding, population genomics, and other efforts.

Funding

This work was supported by the NIH Office of the Director T32OD010993 (Flack), National Institute on Aging L70AG079467 (Flack), National Institute on Aging R21AG071908 (Faulk), Impetus Grant (Norn Group) (Faulk), USDA-NIFA MIN-16-129 (Faulk).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Author Contributions

This work was a collaborative effort by the members of graduate course ANSC 8520 taught by Dr. Faulk in the Department of Animal Science at the University of Minnesota in Fall 2023. Students were Cassens, Enriquez, Gebeyehu, Alshagawi, Hatfield, Kauffman, Klaeui, Mabrouk, Walls, Brown, and Dr. Yeater. Students performed analyses and provided text for the manuscript. Drs. Hughes, Martinson, Rivas, and Koch provided the horse sample, analytical

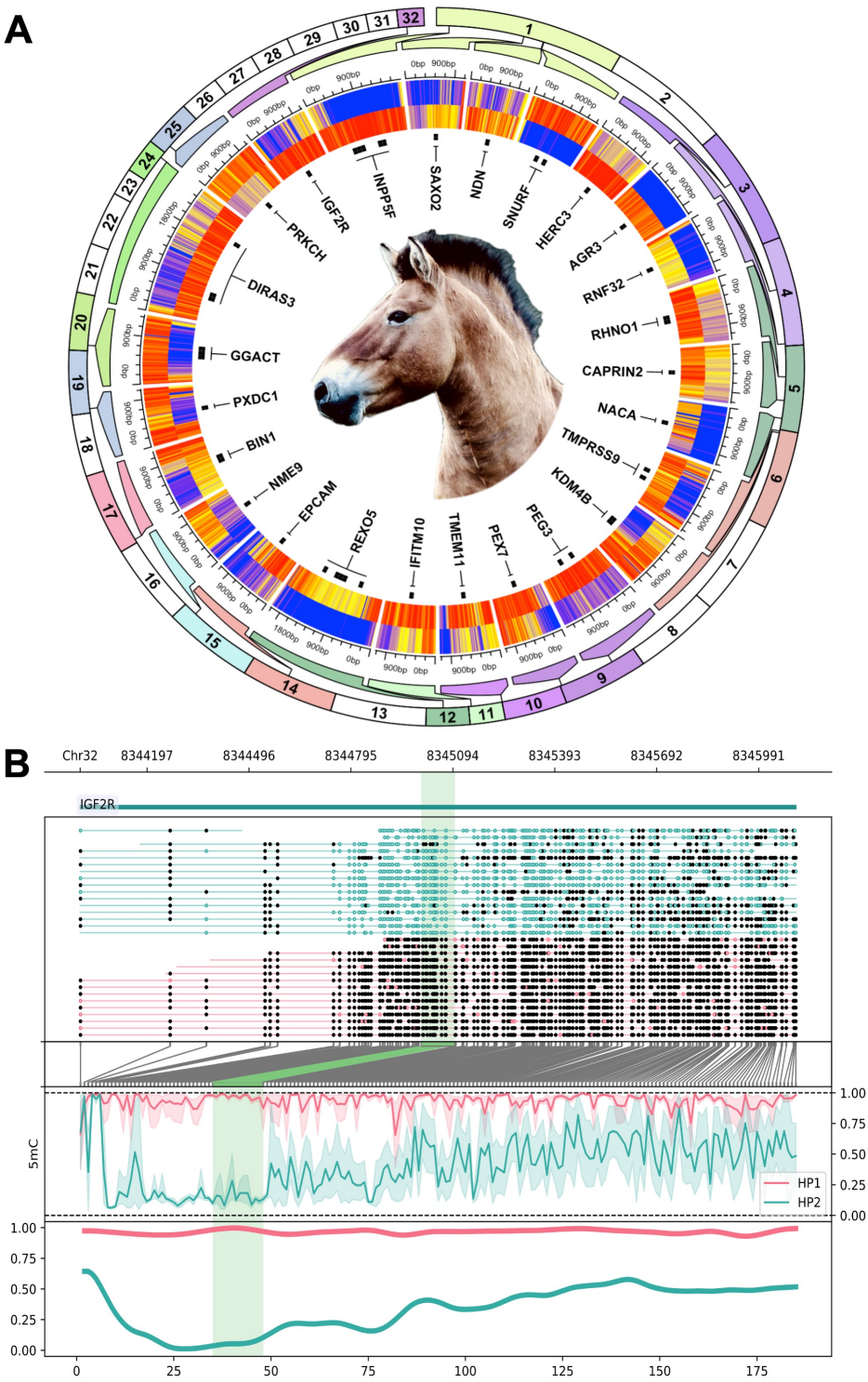

Figure 3. Genome-wide allele-specific DNA methylation analysis for EquPr2. (A) Circular heatmap of 115 differentially methylated regions (DMRs) identified between EquPr2 pseudohaplotypes and labeled by nearest gene symbol. Regions are zoomed from their chromosomal position as indicated by outer track color. Each heatmap row is a pseudohaplotype and each column is a single CpG within the DMR. Methylation values range from 0% (blue) to 100% (red). (B) Example DMR overlapping the known imprinted gene *IGF2R*. Filled circes are methylated CpGs and open circles are unmethylated CpGs.

guidance, and edited the manuscript. Dr. Flack analyzed methylation data, provided text for the manuscript, and edited the manuscript. Dr. Faulk performed analyses and edited the manuscript. All authors have read and approved the manuscript prior to submission.

Data availability

Supplementary File 2 contains code used to generate data. The diploid assembly and mitogenome are available under NCBI umbrella BioProject PRJNA1073944 and submissions JAZHEL000000000 and JAZHEM000000000. Variant calls and other supplementary material are available on figshare.

Acknowledgments

We thank the Minnesota Zoo for generously donating the sample from Varuschka, as well as the Animal Care and Animal Health staff who provided sample collection assistance and knowledge of Przewalski's horse history. We thank Krishona Martinson for helpful discussions on the biology of the horse.

References

1. King SRB, Boyd L, Zimmerman W, Kendall BE. Equus ferus; 2014. Available from: http://dx.doi.org/10.2305/IUCN.UK.2015-2.RLTS.T41763A45172856.en.

2. King SRB. Extinct in the Wild to Endangered: the History of Przewalski's Horse (Equus ferus przewalskii) and its Future Conservation. Mongolian Journal of Biological Sciences. 2005;3(2). doi:10.22353/mjbs.2005.03.11.

3. Der Sarkissian C, Ermini L, Schubert M, Yang M, Librado P, Fumagalli M, et al. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Current Biology. 2015;25(19):2577–2583. doi:10.1016/j.cub.2015.08.032.

4. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, Schubert M, et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science. 2018;360(6384):111–114. doi:10.1126/science.aao3297.

5. MacHugh DE, Larson G, Orlando L. Taming the Past: Ancient DNA and the Study of Animal Domestication. Annual Review of Animal Biosciences. 2017;5(1):329–351. doi:10.1146/annurev-animal-022516-022747.

6. Bouman J. Particulars About the Przewalski's Horse; 1986. Available from: https://przewalskihorse.nl/wp-content/uploads/2016/03/booklet.pdf [cited 12-02-2024].

7. Wakefield S, Knowles J, Zimmermann W, van Dierendonck M. Status and Action Plan for the Przewalski's Horse (Equus ferus przewalskii); 2002. Available from: https://is.muni.cz/el/1431/jaro2007/Bi8710/E-przewalskii.pdf [cited 12-02-2024].

8. The Przewalski horse and restoration to its natural habitat in Mongolia; 1986. Available from: https://www.fao.org/3/AC148E/AC148E00.htm [cited 13-02-2024].

9. Turghan MA, Jiang Z, Niu Z. An Update on Status and Conservation of the Przewalski's Horse (Equus ferus przewalskii): Captive Breeding and Reintroduction Projects. Animals. 2022;12(22):3158. doi:10.3390/ani12223158.

10. Bowling AT, Zimmermann W, Ryder O, Penado C, Peto S, Chemnick L, et al. Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. Cytogenetic and Genome Research. 2003;102(1-4):226–234. doi:10.1159/000075754.

11. Orlando L. Ancient Genomes Reveal Unexpected Horse Domestication and Management Dynamics. BioEssays. 2019;42(1). doi:10.1002/bies.201900164.

12. Goto H, Ryder OA, Fisher AR, Schultz B, Kosakovsky Pond SL, Nekrutenko A, et al. A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski's Horses. Genome Biology and Evolution. 2011;3:1096–1106. doi:10.1093/gbe/evr067.

13. Ryder OA, Wedemeyer EA. A cooperative breeding programme for the Mongolian wild horse Equus przewalskii in the United States. Biological Conservation. 1982;22(4):259–271. doi:10.1016/0006-3207(82)90021-0.

14. Ryder OA. Przewalski's horse — putting the wild horse back in the wild. Oryx. 1988;22(3):154–157. doi:10.1017/s0030605300027770.

15. King SRB, Gurnell J. Habitat use and spatial dynamics of takhi introduced to Hustai National Park, Mongolia. Biological Conservation. 2005;124(2):277–290. doi:10.1016/j.biocon.2005.01.034.

16. Jiang Z, , Zong H, , . REINTRODUCTION OF THE PRZEWALSKI'S HORSE IN CHINA: STATUS QUO AND OUTLOOK. Nature Conservation Research. 2019;4(Suppl.2). doi:10.24189/ncr.2019.045.

17. Robert N, Walzer C, Rüegg SR, Kaczensky P, Ganbaatar O, Stauffer C. PATHOLOGIC FINDINGS IN REINTRODUCED PRZEWALSKI'S HORSES (EQUUS CABALLUS PRZEWALSKII) IN SOUTHWESTERN MONGOLIA. Journal of Zoo and Wildlife Medicine. 2005;36(2):273–285. doi:10.1638/03-035.1.

18. Bernátková A, Oyunsaikhan G, Šimek J, Komárková M, Bobek M, Ceacero F. Influence of weather on the behaviour of reintroduced Przewalski's horses in the Great Gobi B Strictly Protected Area (Mongolia): implications for conservation. BMC Zoology. 2022;7(1). doi:10.1186/s40850-022-00130-z.

19. Tang Y, Liu G, Zhao S, Li K, Zhang D, Liu S, et al. Major Histocompatibility Complex (MHC) Diversity of the Reintroduction Populations of Endangered Przewalski's Horse. Genes. 2022;13(5):928. doi:10.3390/genes13050928.

20. Przewalski's Horse (Equus ferus przewalskii) Re-introduction in the Great Gobi B Strictly Protected Area: from Species to Ecosystem Conservation. Mongolian Journal of Biological Sciences. 2007;5(1-2). doi:10.22353/mjbs.2007.05.03.

21. Xia C, Cao J, Zhang H, Gao X, Yang W, Blank D. Reintroduction of Przewalski's horse (Equus ferus przewalskii) in Xinjiang, China: The status and experience. Biological Conservation. 2014;177:142–147. doi:10.1016/j.biocon.2014.06.021.

22. ; 2022. Available from: https://sandiegozoowildlifealliance.org/pr/kurtandholly [cited 13-02-2024].

23. Two of a Kind;. Available from: https://stories.sandiegozoo.org/zoonooz/two-of-a-kind/ [cited 13-02-2024].

24. The Przewalski's Horse Project; 2024. Available from: https://reviverestore.org/projects/przewalskis-horse/ [cited 13-02-2024].

25. Benirschke K, Malouf N, Low RJ, Heck H. Chromosome Complement: Differences between *Equus caballus* and *Equus przewalskii* , Poliakoff. Science. 1965;148(3668):382–383. doi:10.1126/science.148.3668.382.

26. DELHANTY JDA, MATTHEWS JG, WEST L, BRIGHTLEY G, HAWKEY CM, ASHTON DG, et al. Chromosome studies on 21 Przewalski horses. International Zoo Yearbook. 1979;19(1):236–239. doi:10.1111/j.1748-1090.1979.tb00571.x.

27. Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, Makova KD. Horse Domestication and Conservation Genetics of Przewalski's Horse Inferred from Sex Chromosomal and Autosomal Sequences. Molecular Biology and Evolution. 2008;26(1):199–208. doi:10.1093/molbev/msn239.

28. Myka JL, Lear TL, Houck ML, Ryder OA, Bailey E. FISH analysis comparing genome organization in the domestic horse (Equus caballus) to that of the Mongolian wild horse (E. przewalskii). Cytogenetic and Genome Research. 2003;102(1-4):222–225. doi:10.1159/000075753.

29. Huang J, Zhao Y, Shiraigol W, Li B, Bai D, Ye W, et al. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. Scientific Reports. 2014;4(1). doi:10.1038/srep04958.

30. Ahrens E, Stranzinger G. Comparative chromosomal studies of *E. caballus* (ECA) and *E. przewalskii* (EPR) in a female F1 hybrid. Journal of Animal Breeding and Genetics. 2005;122(s1):97–102. doi:10.1111/j.1439-0388.2005.00494.x.

31. Lippold S, Matzke NJ, Reissmann M, Hofreiter M. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evolutionary Biology. 2011;11(1):328. doi:10.1186/1471-2148-11-328.

32. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology. 2021;39(11):1348–1365. doi:10.1038/s41587-021-01108-x.

33. Arakawa K, editor. Nanopore Sequencing. Springer US; 2023. Available from: http://dx.doi.org/10.1007/978-1-0716-2996-3.

34. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PLOS ONE. 2021;16(10):e0257521. doi:10.1371/journal.pone.0257521.

35. Bajrami E, Spiroski M. Genomic Imprinting. Open Access Macedonian Journal of Medical Sciences. 2016;4(1):181–184. doi:10.3889/oamjms.2016.028.

36. Elhamamsy AR. Role of DNA methylation in imprinting disorders: an updated review. Journal of Assisted Reproduction and Genetics. 2017;34(5):549–562. doi:10.1007/s10815-017-0895-5.

37. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics. 2003;33(S3):245–254. doi:10.1038/ng1089.

38. Skvortsova K, Iovino N, Bogdanović O. Functions and mechanisms of epigenetic inheritance in animals. Nature Reviews Molecular Cell Biology. 2018;19(12):774–790. doi:10.1038/s41580-018-0074-2.

39. Chang C, Lu Q, editors. Epigenetics in Allergy and Autoimmunity. Springer Singapore; 2020. Available from: http://dx.doi.org/10.1007/978-981-15-3449-2.

40. Yamaguchi M. Regulation of animal behavior by epigenetic regulators. Frontiers in Bioscience. 2019;24(6):1071–1084. doi:10.2741/4769.

41. Xu B, Yang G, Jiao B, Zhu H. Analysis of ancient and modern horse genomes reveals the critical impact of lncRNA-mediated epigenetic regulation on horse domestication. Frontiers in Genetics. 2022;13. doi:10.3389/fgene.2022.944933.

42. Xia J, Han L, Zhao Z. Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. BMC Genomics. 2012;13(S8). doi:10.1186/1471-2164-13-s8-s7.

43. Mallinger M, Markle T, Minerich B, Nordmeyer C, Runquist E, Stapleton S. Understanding How the Unique Context of the Minnesota Zoo Shapes Our Local Conservation Initiatives. Journal of Zoological and Botanical Gardens. 2023;4(2):427–444. doi:10.3390/jzbg4020032.

44. Lessard K. A Conservation Milestone: Minnesota Zoo Welcomes First Asian Wild Horse in 20 Years; 2008. http://mnzoo.org/wp-content/uploads/2014/08/ZooTracks_2008_Fall_A-Conservation-Milestone_Phorse.pdf.

45. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology. 2019;37(5):540–546. doi:10.1038/s41587-019-0072-8.

46. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology. 2022;23(1). doi:10.1186/s13059-022-02823-7.

47. Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, et al. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience. 2020;9(9). doi:10.1093/gigascience/giaa094.

48. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution. 2021;38(10):4647–4654. doi:10.1093/molbev/msab199.

49. Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023;39(10). doi:10.1093/bioinformatics/btad595.

50. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142–i150. doi:10.1093/bioinformatics/bty266.

51. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

52. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37(23):4572–4574. doi:10.1093/bioinformatics/btab705.

53. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology. 2012;13(10):R87. doi:10.1186/gb-2012-13-10-r87.

54. Wreczycka K, Gosdschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. Strategies for analyzing bisulfite sequencing data. Journal of Biotechnology. 2017;261:105–115. doi:10.1016/j.jbiotec.2017.08.007.

55. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

56. Cheetham SW, Kindlova M, Ewing AD. Methylartist: tools for visualizing modified bases from nanopore sequence data. Bioinformatics. 2022;38(11):3109–3112. doi:10.1093/bioinformatics/btac292.

57. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences. 2020;117(17):9451–9457. doi:10.1073/pnas.1921046117.

58. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mobile DNA. 2021;12(1). doi:10.1186/s13100-020-00230-y.

59. Zheng Z, Li S, Su J, Leung AWS, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nature Computational Science. 2022;2(12):797–803. doi:10.1038/s43588-022-00387-x.

60. Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau GW, et al.. WhatsHap: fast and accurate read-based phasing; 2016. Available from: http://dx.doi.org/10.1101/085050.

61. Lin JH, Chen LC, Yu SC, Huang YT. LongPhase: an ultra-fast chromosome-scale phasing algorithm for small and large variants. Bioinformatics. 2022;38(7):1816–1822. doi:10.1093/bioinformatics/btac058.

62. Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Molecular Ecology Resources. 2020;20(4):892–905. doi:10.1111/1755-0998.13160.

63. Uliano-Silva M, Gabriel R N Ferreira J, Krasheninnikova K, Formenti G, Abueg L, Torrance J, et al.. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads; 2022. Available from: http://dx.doi.org/10.1101/2022.12.23.521667.

64. Flack N, Drown M, Walls C, Pratte J, McLain A, Faulk C. Chromosome-level, nanopore-only genome and allele-specific DNA methylation of Pallas's cat, *Otocolobus manul*. NAR Genomics and Bioinformatics. 2023;5(2). doi:10.1093/nargab/lqad033.

65. Ko BJ, Lee C, Kim J, Rhie A, Yoo DA, Howe K, et al. Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biology. 2022;23(1). doi:10.1186/s13059-022-02764-1.

66. RICHER CL, POWER MM, KLUNDER LR, McFEELY RA, KENT MG. Standard karyotype of the domestic horse (Equus caballus). Hereditas. 2008;112(3):289–293. doi:10.1111/j.1601-5223.1990.tb00069.x.

67. Musilova P, Kubickova S, Zrnova E, Horin P, Vahala J, Rubes J. Karyotypic relationships among Equus grevyi, Equus burchelli and domestic horse defined using horse chromosome arm-specific probes. Chromosome Research. 2007;15(6):807–813. doi:10.1007/s10577-007-1164-8.

68. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. PLoS Genetics. 2012;8(1):e1002451. doi:10.1371/journal.pgen.1002451.

69. Keilwagen J, Hartung F, Grau J. In: GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. Springer New York; 2019. p. 161–177. Available from: http://dx.doi.org/10.1007/978-1-4939-9173-0_9.