Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids

Brenda Larison, Gabriela M. Pinho, Amin Haghani, Joseph A. Zoller, Caesar Z. Li, Carrie J. Finno, Colin Farrell, Christopher B. Kaelin, Gregory S. Barsh, Bernard Wooding, Todd R. Robeck, Dewey Maddox, Matteo Pellegrini & Steve Horvath

Effective conservation and management of threatened wildlife populations require an accurate assessment of age structure to estimate demographic trends and population viability. Epigenetic aging models are promising developments because they estimate individual age with high accuracy, accurately predict age in related species, and do not require invasive sampling or intensive long-term studies. Using blood and biopsy samples from known age plains zebras (Equus quagga), we model epigenetic aging using two approaches: the epigenetic clock (EC) and the epigenetic pacemaker (EPM). The plains zebra EC has the potential for broad application within the genus Equus given that five of the seven extant wild species of the genus are threatened. We test the EC’s ability to predict age in sister taxa, including two endangered species and the more distantly related domestic horse, demonstrating high accuracy in all cases. By comparing chronological and estimated age in plains zebras, we investigate age acceleration as a proxy of health status. An interaction between chronological age and inbreeding is associated with age acceleration estimated by the EPM, suggesting a cumulative effect of inbreeding on biological aging throughout life.
Effective management of threatened species relies on the ability to estimate demographic trends, which depend, in turn, on accurate information about age distributions within populations. Growth rates shape age distributions, can reflect past and current environmental and anthropogenic perturbations, and can also be used to predict future population growth. However, age is challenging to quantify in wild animals. Age estimation typically requires either investment in long-term field studies or invasive approaches that may not be feasible in live animals. Another problem is the limited accuracy of some methods, which may negatively impact conservation efforts. The challenges and importance of obtaining accurate age information have motivated efforts to develop an accurate and non-invasive approach to aging wild animals.

Epigenetic aging models, particularly epigenetic clocks (ECs), promise to improve the aging of wild animals and thereby make valuable contributions to wildlife conservation and population biology. These highly accurate clocks use information from genomic methylation patterns and have been studied extensively in humans,1–9 and mice10–12. The development and use of epigenetic models in other species are still limited but are becoming increasingly common. A critical limitation to developing epigenetic aging models for wildlife is that these models need to be trained on samples from individuals of known age. Therefore, populations of non-model organisms with known-age individuals are of extreme importance. Here, we develop epigenetic models for a wild equid (plains zebra, Equus quagga) using both blood and remote biopsy samples collected from known-age individuals in a captive-bred population.

Besides their high accuracy, four other features of epigenetic aging models should make them attractive for wildlife managers. First, they can be developed from many different tissue types. Second, they can be created based on very few genomic sites. The most accurate clocks for humans involve only a few hundred CpG sites, and far fewer CpGs have been used to build ECs in some wild vertebrates. Third, an accurate clock can be developed using relatively few individuals of known age. Finally, ECs developed for one species have been shown to predict age accurately in closely related taxa (e.g., humans and chimps), and therefore can be developed for a less threatened species with the intent of using them in threatened sister species. Using this rationale, we aim to test the performance of the EC developed for plains zebras to predict age in domestic horses (Equus caballus) and two threatened species: Grevy’s zebra (E. grevyi) and Somali ass (E. africanus somaliensis).

Since individual chronological ages are known for our plains zebra population, it is possible to estimate how fast individuals are aging compared to others by the discrepancy between epigenetic age and chronological age, dubbed age acceleration. Positive age acceleration indicates that an individual is biologically older than expected based on its chronological age. Age acceleration is predictive of all-cause mortality in humans,20–22 which suggests that epigenetic models can be a powerful approach to study the impact of different factors on biological aging. Age-acceleration has also been associated with stress and adversity,23–25 elevated glucocorticoids,26,27 and inbreeding,28–31 all of which are relevant for managing wild populations. The plains zebra population sampled in this study has a complex pedigree due to semi-captive breeding, which creates the opportunity to test whether inbreeding is associated with accelerated epigenetic aging in this population. The strong correlation of ECs to age can sometimes negatively affect their ability to detect age acceleration associated with biological variation.22,23 We therefore also develop a second model for the plains zebra, the epigenetic pacemaker model (EPM). The EPM has previously been found to be useful for investigating how environmental and life-history factors influence aging.24,34,35

Our main goals for this study are to (1) develop epigenetic aging models for plains zebras; (2) test the ability of an EC developed for plains zebras to predict age in three equid species; and (3) estimate the influence of inbreeding levels on plains zebra aging patterns by correlating it with age acceleration predicted by both an EC and EPM. Our main results include an EC that predicts age accurately in plains zebras and three congeners tested, including domestic horse and two endangered sister species. We further show that inbreeding associated age acceleration increases with age, suggesting that inbreeding may have a cumulative effect on age acceleration throughout life. The development of epigenetic aging models in a wild equid stands to have broad impact because the crown group of the genus Equus comprises a closely related group of six extant species, five of which range from near threatened to critically endangered.

### Results

**Epigenetic aging models.** We developed epigenetic models using methylation data profiles from three plains zebra data sets: (1) 76 blood samples, (2) 20 biopsy samples, and (3) 96 blood and biopsy samples combined (Table 1). For each data set, we developed both an EC and an EPM. We evaluated the effectiveness of applying the blood-based zebra EC to predict age in other equids using known-age domestic horses (E. caballus, n = 188), Grevy’s zebras (E. grevyi, n = 5), and Somali asses (E. africanus somaliensis, n = 7).

To develop the ECs we fit a generalized linear model with elastic-net penalization using leave-one-out (LOO) cross-validation. To improve EC fit, we square root transformed chronological age prior to fitting the models. The blood EC (Pearson’s r = 0.96, median absolute error (MAE) = 0.56 years, Fig. 1a) and the combined tissue EC (r = 0.89, MAE = 0.62, Fig. 1c) predicted age more accurately than the biopsy EC (r = 0.62, MAE = 1.79 years, Fig. 1b). The blood EC selected 70 CpG sites, the biopsy clock 31 CpGs, and the combined clock selected 99 CpGs. The biopsy EC had no CpG sites in common with the blood and combined ECs. The blood EC and combined tissue EC shared only two CpGs. We report coefficients, intercepts, and lambdas in Supplementary Data 1 and present the results from using untransformed ages in Supplementary Fig. 1.

Cross-species predictive ability was high (Fig. 1d–f). The zebra-blood EC predicted horse age with high accuracy (r = 0.93, MAE = 1.82). The error when predicting the ages of horses younger than 15 years was lower (MAE = 1.15) than when predicting the ages of older horses (MAE = 3.97). While prediction errors for Grevy’s zebra and Somali wild ass ages were even lower (MAE of 1.08 and 1.15 respectively), this should be viewed with caution as we had limited sample sizes from Grevy’s zebra (n = 5) and Somali wild ass (n = 7).

To construct plains zebra EPMs we used sites in which methylation levels were highly correlated with individual chronological age. Epigenetic states were estimated using a leave-one-out cross-validation. The blood, biopsy, and combined-tissue EPMs included 391, 242, and 248 CpG sites, respectively. We provide details of the CpGs selected by each model in

| Tissue | N | No. female | Mean age | Min. age | Max. age |
|--------|---|------------|----------|----------|----------|
| Blood  | 76 | 42         | 5.21     | 0.156    | 20.2     |
| Biopsy | 20 | 9          | 5.87     | 0.162    | 24.8     |

*We restrict the description to animals whose ages could be estimated with high confidence (90% or higher).* Tissue type, N = Total number of samples/arrays. Number of females. Age: mean, minimum and maximum.
Supplementary Data 1. Epigenetic state was strongly correlated with chronological age in both tissue types: blood ($r = 0.97$, Fig. 1g) and biopsy ($r = 0.95$, Fig. 1h). EPMs based on the two tissues used largely distinct sets of CpG sites, sharing only 40 sites. Despite retaining a strong overall correlation in the combined EPM ($r = 0.96$), the difference in the model’s performance for the two sample types is apparent (Fig. 1i). The combined tissue pacemaker shared 138 sites with blood and only 34 with biopsies.

**Association of inbreeding with biological aging.** Inbreeding was estimated both as the inbreeding coefficient F and by the proportion of the genome in runs of homozygosity (FROH). We derived the genotypes used to estimate inbreeding from two sources: RAD sequencing data (42 samples) and genotypes imputed at the same set of loci from low-coverage whole-genome sequencing data using GLIMPSE\textsuperscript{38} (28 additional samples). GLIMPSE produced high-quality imputations (mean dosage $r^2$ of 80%; mean concordance between imputed and true genotype of...
92%) as assessed by running leave-one-out imputations on 35 samples present in both the RADseq and low coverage data sets (Supplementary Fig. 2a, b). The Mendelian error rate across loci averaged 0.06 (+/−0.06) in the full data set. ROH were discoverable across 90% of the genome and ranged in size from 1.5 MB (minimum allowed size) up to >60 MB (Supplementary Fig. 2c). ROH over 10 MB are expected to reflect inbreeding loops occurring within the last five generations. Individuals from more recent generations showed a pattern of excessive total ROH relative to the number of ROH segments as indicated by an upward shift in ROH size relative to number (Supplementary Fig. 2d). \( F_{\text{ROH}} \) ranged from 0 to 0.37, and \( F \) statistics ranged from −0.21 to 0.32 (one outlier with an extreme negative \( F \) value was removed from the analysis). The Pearson correlation between \( F \) and \( F_{\text{ROH}} \) was 0.84.

We used multiple linear regressions to assess whether inbreeding is associated with age acceleration in the plains zebra population. Age acceleration was calculated by subtracting each individual’s chronological age from its predicted age, and was calculated separately for the EPM and EC. Sex was added to the linear models as a covariate. EPM age acceleration was calculated separately for the EPM and EC. Sex was not associated with either measure of inbreeding. Sex was not associated with age acceleration in any model. A re-run of our analyses using only the RADseq samples gave similar results (Supplementary Table 1a–h).

**EWAS and functional analysis of plains zebra tissues.** We performed the EWAS analysis on the 31,836 probes that could be uniquely aligned to specific adjacent loci in the horse genome. Since the mammalian methylation array is based on stretches of DNA that are conserved in all mammals, the horse annotation of 9757 and 331 probes were related to age in blood (\( n = 76 \), age-range 0.15–20.2 years) and biopsies (\( n = 20 \), age-range 0.16–24.8 years) respectively (Fig. 3a). The top age-related changes per tissue are as follows (Fig. 3a): blood, hypomethylation in FANCI upstream, MAF downstream, ZNF608 upstream, and PBX3 intron; biopsy, hypermethylation in PLCBI, NEUROD1, and BARHL2 upstream. DNA methylation-based age estimators for any wild equid. The high accuracy of the EC models reflects that we used a custom array that profiled 36 thousand probes that were highly conserved across numerous mammalian species. These robust data allowed us to construct highly accurate epigenetic aging models for plains zebras. The best model to predict chronological age was the EC developed from blood samples, which predicted individual age with a 6-month error.

Developing a highly accurate biopsy-based EC may be challenging due to the variability of tissue types within such a sample. Biopsy samples consist of three skin layers—the epidermis, the dermis, and the hypodermis—and may even contain deeper tissues. These individual layers can exhibit different methylation patterns and are often present in different proportions across samples because they vary in thickness across the body and among individuals. A comparison between blood and skin-based (dermis and epidermis only) odontocete clocks also found the skin-based clocks to be less accurate. Despite the inherent difficulties of using biopsy samples and the small sample size and skewed age distributions, the associations of CpG sites with chronological age in blood and biopsy were relatively similar, with a moderate positive correlation between the z-scores from the EWAS for each tissue type (\( r = 0.25 \), Fig. 3e). Of the CpGs with significant association with chronological age, only 81 overlapped between blood and biopsy samples (Fig. 3d). Some of these shared CpGs include hypermethylation in PLCBI exon, RIMS1 exon, and hypomethylation in NOVA2 intron and NFI intron (Supplementary Fig. 3a, b). In contrast to results based on specific genes, functional enrichment analysis using GREAT identified that age-related CpGs in both blood and biopsies were significantly related with regard to biological pathways, specifically development (e.g., nervous system) and survival, and were enriched with polycomb repressor complex 2 (e.g., EED, SUZ12, PCR2) target genes (Supplementary Fig. 4d).

**Discussion** To the best of our knowledge, this is the first study to present DNA methylation-based age estimates for any wild equid. The high accuracy of the EC models reflects that we used a custom array that profiled 36 thousand probes that were highly conserved across numerous mammalian species. These robust data allowed us to construct highly accurate epigenetic aging models for plains zebras. The best model to predict chronological age was the EC developed from blood samples, which predicted individual age with a 6-month error.

Developing a highly accurate biopsy-based EC may be challenging due to the variability of tissue types within such a sample. Biopsy samples consist of three skin layers—the epidermis, the dermis, and the hypodermis—and may even contain deeper tissues. These individual layers can exhibit different methylation patterns and are often present in different proportions across samples because they vary in thickness across the body and among individuals. A comparison between blood and skin-based (dermis and epidermis only) odontocete clocks also found the skin-based clocks to be less accurate. Despite the inherent difficulties of using biopsy samples and the small sample size and skewed age distributions, the associations of CpG sites with chronological age in blood and biopsy were relatively similar, with a moderate positive correlation between the z-scores from the EWAS for each tissue type (\( r = 0.25 \), Fig. 3e). Of the CpGs with significant association with chronological age, only 81 overlapped between blood and biopsy samples (Fig. 3d). Some of these shared CpGs include hypermethylation in PLCBI exon, RIMS1 exon, and hypomethylation in NOVA2 intron and NFI intron (Supplementary Fig. 3a, b). In contrast to results based on specific genes, functional enrichment analysis using GREAT identified that age-related CpGs in both blood and biopsies were significantly related with regard to biological pathways, specifically development (e.g., nervous system) and survival, and were enriched with polycomb repressor complex 2 (e.g., EED, SUZ12, PCR2) target genes (Supplementary Fig. 4d).
distribution of the biopsy samples (Supplementary Fig. 5), our biopsy clock predicted age with an error of $\pm 2\text{ years}$.

The plains zebra blood EC accurately predicted the chronological ages of horses, Grevy’s zebras, and Somali wild asses. This was expected since caballine and non-caballine equids are somewhat closely related ($4 - 4.5$ MYA)\cite{4,46}. Non-caballine species of equids are more closely related to plains zebras ($1.28 - 1.75$ MYA) than domestic horses, which may explain the lower errors found for age estimation in Grevy’s zebras and Somali asses. In fact, chimpanzees and bonobos have a similar divergence time to those observed within the non-caballine equids\cite{47} and align more closely to each other in DNAm age than either does with humans\cite{14}.

The extraordinary accuracy of ECs stems from their utilization of sites that maximize a linear relationship between epigenetic age and chronological age. This accuracy can sometimes negatively affect their ability to detect age acceleration associated with biological variation\cite{32,33}. Methylation levels change in a non-linear fashion throughout individuals’ lifetimes in several species, with accelerated changes in early life and slower changes once individuals reach adulthood\cite{14,34}. The EPM was developed to model these non-linear changes\cite{34,48,49}. The EPM estimates epigenetic age by maximizing the similarity between estimated and observed methylation levels, and therefore does not make any assumptions about linearity but rather identifies the shape of the relationship between age and methylation directly from the data. In this sense, the EPM is potentially more associated with biological than chronological aging\cite{34}, which may be particularly useful for investigating how environmental and life-history factors influence aging\cite{35}.

**Fig. 3 Epigenome wide association study (EWAS) of chronological age in blood and skin of plains zebras. a** Manhattan plots of the EWAS of chronological age. Since a genome assembly was not available for zebra, the coordinates are estimated based on the alignment of Mammalian array probes to EquCab3.0.100 (domestic horse) genome. The direction of associations with $p < 10^{-4}$ (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors. The top 15 CpGs were labeled by the neighboring genes. **b** Location of top age-related CpGs in each tissue relative to the closest transcriptional start site. Top CpGs were selected at $p < 10^{-4}$ and further filtering based on z-score of association with chronological age for up to 500 in a positive and negative direction. The number of selected CpGs: blood, 1000; biopsy, 331; meta-analysis, 1000. The gray color represents the location of 3,1836 mammalian BeadChip array probes mapped to EquCab3.0.100 genome. **c** Boxplot of z-scores from a correlation of age with CpG location (within or outside CpG islands). The median Z statistics are significantly different ($p < 10^{-4}$). **d** Venn diagram of the top age-related CpGs in blood and biopsy samples from plains zebras. **e** Sector plot of DNA methylation aging in plains zebra blood and biopsy tissues. Red dotted line: $p < 10^{-4}$; blue dotted line: $p > 0.05$; Red dots: shared CpGs; black dots: tissue-specific changes; blue dots: CpGs whose age correlation differs between blood and biopsy tissue.
The EPM models developed here reveal that epigenetic changes occur in a non-linear fashion throughout the lifespan of plains zebras. As has been found in humans and other species\(^\text{65,66}\), young zebras undergo faster epigenetic changes than adult zebras. We also observed that the variance in the estimates of epigenetic age is lower in young compared to adult zebras. Increased variation in epigenetic age in adults is observed in humans and other species as well and may be a consequence of lifetime accumulation of environmental and physiological factors on the epigenome\(^\text{7,14}\). In agreement with the idea of cumulative effects on aging, the effects of inbreeding on epigenetic aging were more apparent in older individuals. We found a significant effect of the interaction between inbreeding measures and chronological age on age acceleration estimated from the blood EPM, wherein inbred individuals exhibited higher age acceleration at older chronological ages. The association between inbreeding and DNA methylation has been described in plants\(^\text{28,30}\), chickens\(^\text{31}\), and salmon\(^\text{29}\). Increases in inbreeding effects with age, such as those we describe, are predicted by theory\(^\text{50}\) and have been shown empirically in an insect\(^\text{51}\) and mammal\(^\text{52}\) species. Because the probes in the mammalian array were selected based on conservation in mammalian genomes, we expect our findings will have high translatability into humans and other mammals. The age-related gain of methylation in promoters (Fig. 3b) is consistent with observations in humans and many other species\(^\text{53,54}\). The low overlap of significant CpGs between tissues (Fig. 3d) may reflect the relatively low sample size (\(n = 20\) skin biopsy samples) or biological differences between the biopsy and whole blood samples. Some of the genes close to the most significant CpGs in the EWAS of blood and biopsy samples play key roles in the DNA damage pathway and maintenance of genomic integrity (FANCL)\(^\text{55}\), regulation of cellular and/or developmental processes (MAF\(^\text{56}\), PBX3\(^\text{57}\), NEUROD1\(^\text{58}\), BARHL2\(^\text{59,60}\), NOVA2\(^\text{61,62}\), NFIA\(^\text{63,64}\)), and extra- and/or intra- cellular signaling (PLCB1\(^\text{65,66}\), RIMS1\(^\text{67,68}\)).

Most of these genes are near to the top significant CpG sites from an EWAS that included multiple mammalian species and tissues: MAF, NEUROD1, BARHL2, NFIA, PLCB1, and RIMS1\(^\text{69}\). In mice, MAF promotes osteoblast differentiation and may be an important gene for age-related bone disease therapy\(^\text{70,71}\). PBX3 is at the center of the most enriched network associated with aging based on differentially methylated regions in human blood\(^\text{72}\), and its expression levels in rat frontal cortex with aging may affect the relatively low sample size (\(n = 20\) skin biopsy samples) or biological differences between the biopsy and whole blood samples. Some of the genes close to the most significant CpGs in the EWAS of blood and biopsy samples play key roles in the DNA damage pathway and maintenance of genomic integrity (FANCL)\(^\text{55}\), regulation of cellular and/or developmental processes (MAF\(^\text{56}\), PBX3\(^\text{57}\), NEUROD1\(^\text{58}\), BARHL2\(^\text{59,60}\), NOVA2\(^\text{61,62}\), NFIA\(^\text{63,64}\)), and extra- and/or intra- cellular signaling (PLCB1\(^\text{65,66}\), RIMS1\(^\text{67,68}\)).

Many of these genes are near to the top significant CpG sites from an EWAS that included multiple mammalian species and tissues: MAF, NEUROD1, BARHL2, NFIA, PLCB1, and RIMS1\(^\text{69}\). In mice, MAF promotes osteoblast differentiation and may be an important gene for age-related bone disease therapy\(^\text{70,71}\). PBX3 is at the center of the most enriched network associated with aging based on differentially methylated regions in human blood\(^\text{72}\), and its expression levels in rat frontal cortex with aging may depend on the amount of stress experienced by the previous generations\(^\text{73}\). Many of these genes are potential targets for therapies to prevent and treat age-related cognitive decline and neurodegeneration. BARHL2 is differentially expressed in the hippocampus of young and old rats\(^\text{74}\), and methylation levels in nearby CpGs are associated with aging across different tissues in naked mole rats\(^\text{75}\). PLCB1’s expression levels are associated with aging in the human prefrontal cortex\(^\text{76}\). SNPs near ZNF608 have been associated with early stage of cognitive decline\(^\text{77}\), Alzheimer’s disease risk\(^\text{78}\), and body mass index\(^\text{79,80}\) in humans. NEUROD1 is differentially expressed with aging in the mouse hippocampus and is a critical regulator of neurogenesis\(^\text{81}\). RIMS1 is important for synaptic transmission at neuromuscular junctions in mammals\(^\text{82}\). The abundance of the NOVA2 protein in the cytoplasm decreases with aging in human neurons\(^\text{83}\). NFIA modulates the plasticity of local circuits in the adult hippocampus and may be involved in the cortical atrophy associated with Alzheimer’s disease\(^\text{84}\). In addition to the EWAS results, the GREAT enrichment analysis implicated developmental pathways, bivalent chromatin, and regions suppressed by polycomb repressor complex 2. These results are consistent with observations in many other mammalian species and corroborate findings on DNAm aging in many other mammalian species\(^\text{3,14,17,32,33,69,73,85}\).

We expect that epigenetic aging models will be valuable for ecological studies and population management in wild species since age estimates can be used to assess reproductive potential and population viability. In known age populations, both EGS\(^\text{71,80,87}\) and EPMs\(^\text{14,48,49}\) have the potential to identify causes of individual accelerated aging. These models can be especially useful when combined with ecological or other genetic data\(^\text{4}\). Given the small number of CpG sites required, an aging project in a wild population could be done relatively inexpensively using bisulfite sequencing or pyrosequencing\(^\text{48–50}\), and the costs of the mammalian array are decreasing. While epigenetic models based on blood are more accurate and less invasive than many other options for aging mammals, there are drawbacks in that animals must be immobilized to obtain the samples. Because biopsy samples can be obtained with minimal disruption\(^\text{85}\), a highly reliable clock based on biopsy samples will be a worthwhile direction for future research. Since the methods for extracting genomic DNA from feces have improved\(^\text{92–94}\), it will be worthwhile to explore whether epigenetic aging models can be adapted to this non-invasive source of DNA.

Methods

Samples. We obtained both whole blood (96) and remote biopsy (24) samples from a captive population of zebras maintained in a semi-wild state by the Quagga Project66 in the Western Cape of South Africa. The population was founded in 1989 with 19 wild individuals (9 from Etosha National Park in Namibia and 10 from the KwaZulu-Natal in South Africa). Since its inception, the population has undergone artificial selection to reproduce the phenotype of the extinct quagga subspecies: no stripes on legs and hindquarters, and thinner and paler stripes in the head and barrel region. At sampling, we identified individuals by their unique stripe patterns and derived their chronological ages from studbook information, in which dates of birth are typically accurate to within one month. One exception is a biopsy sample from a founder that was captured for the project as a young mare and would have been at least 25 years old at sampling. We obtained remote biopsies using an air-powered rifle affixed with a 1 mm wide by 20–25 mm deep biopsy dart and preserved in RNAlater (Qiagen). Veterinarians collected blood opportunistically during activities of the Quagga Project and preserved them in EDTA tubes. All but four samples were collected from different individuals; two individuals were sampled twice at different ages (one and three years apart, respectively). We stored all samples at \(-20^\text{\circ}C\). After eliminating samples (24 of 120) with <90% confidence for individual identity or age, we retained 76 blood samples and 20 biopsy samples, totaling 96 plains zebra samples (Table 1).

The collection of 188 whole-blood samples from domestic horses is described in detail in\(^\text{7}\). The Grevy’s zebra (\(n = 5\)) and Somali wild ass (\(n = 7\)) are samples from zoo-based animals that were opportunistically collected and banked during routine health exams. The DNA methylation profiles from these samples have been reported previously\(^\text{69}\).

Ethics approval. We collected plains zebra samples under a protocol approved by the Research Safety and Animal Welfare Administration, University of California Los Angeles: ARC # 2009-090-31, approved initially in 2009.

DNA methylation data. We generated all DNA methylation data (plains zebra, horse, Somali wild ass, Grey’s zebra) using a custom illumina methylation array (HorvathMammalMethylChip40)\(^\text{96}\). The array contains 36 thousand probes, 31 of which mapped unique to the horse genome\(^\text{97,98}\). We normalized methylation values from each species (plains zebra, horse, Somali wild ass, and Grey’s zebra) and tissue (blood and biopsy) using SeSAMe\(^\text{99}\). Unsupervised hierarchical clustering revealed that the plains zebra samples clustered by tissue (Supplementary Fig. 6).

Epigenetic aging models. We studied epigenetic aging in plains zebras using both EC\(^\text{56,95}\) and EPM models\(^\text{14}\). For the ECs we fit generalized linear models in glmnet v.4.0-2 in R v.1.1.0\(^\text{101}\) and treated the methylation data from other equids (domestic horse, Grey’s zebra, Somali wild ass) independently and did not use them for zebra clock development. In addition to the LOO cross-validation presented here, we also conducted analyses using 10-fold cross-validation. The two forms of cross-validation did not produce appreciably different results. We tested the ability of the blood-based EC to predict chronological age for other equids (domestic horse, Grey’s zebra, Somali wild ass) by inputting the DNA methylation profiles of these species into the plains zebra model. We used the
median absolute error of chronological age estimation and the Pearson correlation
between predicted ages and known ages to assess accuracy for each species.
To construct the models, we used sites in which methylation levels were
highly correlated with individual chronological age. The Pearson correlation (r)
thresholds for entry into the model were absolute values of 0.75 for blood and
biopsy, and 0.6 for the combined EPM. The threshold used to select the sites for
input into the combined EPM was lower because only one CpG site had r higher
than 0.7. Epigenetic states were estimated using a leave-one-out cross-validation
with EpigeneticPacemaker 0.3.48 in Python 3.7.1. Supplementary Data 2
provides all Pearson coefficients for methylation levels against chronological age.

Association of inbreeding with biological aging. We prepared libraries and
constructed 2 × 150 bp paired-end RAD sequencing as described in105. To maximize
the number of SNP’s ascertainment, we sequenced reads produced by two restriction
enzymes, SfiI and PacI (74155 and 127429 cut sites in the horse genome,
respectively). We aligned reads to the horse genome EquCab7 using BWA106. We
called genotypes using the haplotype-based callers Freebayes107 and SnpEff108
retaining only the intersection of variants called by these callers. We retained only
variants within the first read of each paired-end read (a total of ~60.5 Mb). We also
removed indels, multiallelic and non-autosomal sites, loci genotyped in <10% of
individuals, and individuals genotyped at <20% of loci. We followed GATK’s basic
guidelines for additional filtering (https://gatk.broadinstitute.org/hc/en-us/articles/
360035890471-Hard-filtering-germline-short-variants), excluding SNPs with
QD < 2, FS > 60,SOR > 3,MQ < 40, MRQRankSum <-12.5, and ReadPosRankSum
<8. Finally, we removed SNPs with MAF < 0.01. Our filtering strategy resulted in
56 samples genotyped at 322542 loci. The number of loci is consistent with
levels of heterozygosity observed in inbred wild populations of plains zebras105.
The 56 samples with RADseq genotypes were used as a reference panel to
impute these same 322542 SNPs in 89 individuals sequenced at low coverage. Low
coverage libraries were constructed from 200 ng genomic DNA using TrueSeq Nano
kit (Illumina), indexed with unique dual indices (Integrated DNA Technologies),
and sequenced at 2 x 31 reads per lane on an Illumina HiSeq platform (Illumina). 6.2 ± 1.4 raw Gbp per sample. We aligned sequences to the horse genome
EquCab7 using BWA106. Genotypes were then imputed using GIMPLOMPE108. As
part of the GIMPLOMPE pipeline, genotype likelihoods were called from the low
coverage bams using mpileup in bcftools109, ignoring indels and duplicate reads,
and recalculating base alignment quality on the fly. Thirty-five of the 89 imputed
individuals were also present in the RADseq data and were used to assess
imputation quality using a leave-one-out approach. Forty-two individuals from the
RAD-seq data and 28 from the imputed data had associated epigenetic data. The combined sample of 70 individuals spans seven generations of the Quagga Project. With the exception of removing singletons and private doubletons, we did not further MAF prune or LD prune the combined RAD
sequence data. We filtered genotypes using “standardScreeningNumericTrait” from the “WGSCNA” R package111. The results were combined across tissues using Stouffer’s meta-analysis method.112 We estimated the distance of each significant CpG site to the closest transcriptional start site. We retained only the 500 CpG sites with the most positive
z-scores and the 500 with the most negative z-scores from each EWAS (blood,
biopsy, and the combined meta-analysis). Restricting the number of analyzed CpG sites did not have a drastic impact on the enriched pathway results. We used these CpG sites as the input for GREAT analysis software113. The background was the human Hg19 genome with 31,832 horse genomic loci. The results of the analysis included “Basal plus extension” and a maximum of 50 kb flanking window for the CpG sites coordinates.

EWAS and functional analysis of plains zebra tissues. To identify genes
potentially associated with aging, we performed EWAS in each tissue separately
using the R function “standardScreeningNumericTrait” from the “WGSCNA” R package.111 The results were combined across tissues using Stouffer’s meta-analysis method.112 We estimated the distance of each significant CpG site to the closest transcriptional start site. We retained only the 500 CpG sites with the most positive
z-scores and the 500 with the most negative z-scores from each EWAS (blood,
biopsy, and the combined meta-analysis). Restricting the number of analyzed CpG sites did not have a drastic impact on the enriched pathway results. We used these CpG sites as the input for GREAT analysis software.113 The background was the human Hg19 genome with 31,832 horse genomic loci. The results of the analysis included “Basal plus extension” and a maximum of 50 kb flanking window for the CpG sites coordinates.

Statistics and reproducibility. Statistical analyses were performed for the epige-
netic models, inbreeding and age-acceleration, and EWAS and functional analysis.
The analyses are described in the corresponding Methods sections, including all
parameters used to allow reproducibility.

Reporting summary. Further information on research design is available in the Nature
Research Reporting Summary linked to this article.

Data availability. Methylation data for plains zebras can be downloaded from Gene Expression Omnibus GSE184223. RAD sequencing data is available as fastq files on SRA, BioProject ID:
PRJNA670933. Data and scripts associated with model development and the creation of
Figs. 1 and 2 are published on DRYAD (doi:10.5068/D1W39K).

Code availability. Scripts used to run EWAS and functional analyses are at https://github.com/shorvath/
MammalianMethylation Consortium/

Received: 27 August 2021; Accepted: 2 December 2021; Published online: 17 December 2021

References
1. Beissinger, S. R. & Westphal, M. I. On the use of demographic models of
population viability in endangered species management. J. Wildl. Manag. 62,
821–841 (1998).
2. Campana, S. Accuracy, precision and quality control in age determination,
including a review of the use and abuse of age validation methods. J. Fish.
Biol. 59, 197–242 (2001).
3. Polanowski, A. M., Robbins, J., Chandler, D. & Jarman, S. N. Epigenetic
estimation of age in humpback whales. Mol. Ecol. Resour. 14, 976–987 (2014).
4. Jarman, S. N. et al. Molecular biomarkers for chronological age in animal
ecology. Mol. Ecol. 24, 4826–4847 (2015).
5. Thompson, M. J., vonHoldt, B. M., Horvath, S. & Pellegrini, M. An epigenetic
aging clock for dogs and wolves. Aging 9, 1055–1068 (2017).
6. De Paoli-Ippoliti, R. et al. Measuring animal age with DNA methylation: from
humans to wild animals. Front. Genet. 8, 106 (2017).
7. Bell, C. G. et al. DNA methylation aging clocks: challenges and
recommendations. Genome Biol. 20, 249 (2019).
8. Field, A. E. et al. DNA methylation clocks in aging: categories, causes, and
consequences. Mol. Cell 71, 882–895 (2018).
9. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic
clock theory of ageing. Nat. Rev. Genet. 19, 371–384 (2018).
10. Petkovich, D. A. et al. Using DNA methylation profiling to evaluate biological
age and longevity interventions. Cell Metab. 25, 954–960 e956 (2017).
11. Stubbs, T. M. et al. Multi-tissue DNA methylation age predictor in mouse.
Genome Biol. 18, 68 (2017).
12. Wang, T. et al. Epigenetic aging signatures in mice livers are slowed by
dwarfism, calorie restriction, and rapamycin treatment. Genome Biol. 18, 57
(2017).
13. Nussey, D. H., Froy, H., Lemaître, J. F., Gaillard, J. M. & Austin, S. N.
Senescence in natural populations of animals: widespread evidence and its
implications for bio-gerontology. Ageing Res. Rev. 12, 214–225 (2013).
14. Horvath, S. DNA methylation age of human tissues and cell types. Genome
Biol. 14, R115 (2013).
15. Voisin, S. et al. An epigenetic clock for human skeletal muscle. J. Cachexia
Sarcopenia Muscle Metab. 960 e956 (2020).
16. De Paoli-Ippoliti, R. et al. Age estimation in a long-lived seabird (Ardenna
tenuirostris) using DNA methylation-based biomarkers. Mol. Ecol. Resour. 19,
411–425 (2019).
17. Ito, H., Udono, T., Hirata, S. & Inoue-Murayama, M. Estimation of
chimpanzee age based on DNA methylation. Sci. Rep. 8, 9998 (2018).
18. Chen, B. H. et al. DNA methylation-based measures of biological age meta-
analysis predicting time to death. Aging 8, 1844–1865 (2016).
19. Christiansen, L. et al. DNA methylation age is associated with mortality in a
longitudinal Danish twin study. Aging Cell 15, 149–154 (2016).
20. Horvath, S. et al. Decreased epigenetic age of PBMCs from Italian semi-
supercentenarians and their offspring. Aging 7, 1139–1170 (2018).
21. Marioni, R. E. et al. DNA methylation age of blood predicts all-cause mortality
in later life. Genome Biol. 16, 25 (2015).
22. Perna, L. et al. Epigenetic age acceleration predicts cancer, cardiovascular,
and all-cause mortality in a German case cohort. Clin. Epigenetics 8, 64 (2016).
23. Mitchell, C., Schneider, L. M. & Notterman, D. A. DNA methylation, early life
environment, and health outcomes. Pediatr. Res. 79, 212–219 (2016).
Pérez, R. F., Santamarina, P., Fernández, A. F., & Fraga, M. F. Epigenetics and Lifestyle: The Impact of Stress, Diet, and Social Habits on Tissue Homeostasis. In Epigenetics and Reparation (ed. Palacios, D.) pp. 461–489 (Academic Press, 2019).

Szyf, M., Tang, Y. Y., Hill, K. G. & Musci, R. The dynamic epigenome and its implications for behavioral interventions: a role for epigenetics to inform disorder prevention and health promotion. Transl. Behav. Med. 6, 55–62 (2016).

Lee, R. S. et al. Chronic corticosterone exposure increases expression and decreases deoxyribonucleic acid methylation at promoters of Chinnok salmon. Mol. Ecol. 25, 4521–4532 (2016).

Verger, P., Wagemaker, N. C. & Ouborg, N. J. Evidence for an epigenetic role in inbreeding depression. Biol. Lett. 8, 798–801 (2012).

Han, W. et al. Genome-wide analysis of the role of DNA methylation in inbreeding depression of reproduction in Langshan chicken. Genomics 112, 2677–2687 (2020).

Thompson, M. J. et al. A multi-tissue full lifespan epigenetic clock for mice. Aging 10, 2832–2854 (2018).

Zhang, Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. 11, 54 (2019).

Snr, S., Farrell, C. & Pellegrini, M. Human epigenetic ageing is logarithmic with time across the entire lifespan. Epigenetics 14, 912–926 (2019).

Ceballos, F. C., Hazelhurst, S. & Ramsay, M. Runs of homozygosity in sub-Saharan African populations provide insights into complex demographic histories. Hum. Genet. 138, 1123–1142 (2019).

Curik, I., Ferenczakovit, M. & Sölkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. Livest. Sci. 166, 26–34 (2014).

Lynch, J. A. et al. The costs of competition: high social status males experience accelerated epigenetic ageing in wild baboons. eLife 10, e66128 (2020).

McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory elements. Nat. Genet. 48, 234–243 (2016).

McClain, A. T. & Faulk, C. The evolution of CpG density and lifespan in conserved primate and mammalian promoters. Aging 10, 561–572 (2018).

Moehlman, P. D. Equids: Zebras, Asses, and Horses Status Survey and Conservation Action Plan (ed. Palacios, D.) pp. 461–489 (Academic Press, 2019).
83. Tollery, J. R. et al. Analysis of alternative splicing associated with aged and neurodegeneration in the human brain. Genome Res. 21, 1572–1582 (2011).

84. Kim, B. H., Nho, K. & Lee, J. M. Alzheimer’s Disease Neuroimaging Initiative. White matter hyperintensity and supratentorial atrophy in Alzheimer’s disease. Neurobiol. Aging 102, 200 e201–200 e211 (2021).

85. Horvath, S. et al. DNA methylation aging and transcriptomic studies in horses. Preprint at bioRxiv https://doi.org/10.1101/2021.03.11.435032 (2021).

86. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of aging: linking environmental inputs to genomic stability. Nat. Rev. Mol. Cell. Biol. 16, 593–610 (2015).

87. Quach, A. et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging 9, 419–446 (2017).

88. Crary-Dooley, F. K. et al. A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. Epigenetics 12, 206–214 (2017).

89. Reed, K., Poulin, M. L., Yan, L. & Parissanti, A. M. Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. Anal. Biochem. 397, 96–106 (2010).

90. Tost, J., Dunker, J. & Gut, I. G. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. Biotechniques 35, 152–156 (2003).

91. Kalb, S. M. et al. Use of the Epigenetic Clock for primary prevention. Neurobiology of Aging 593, 41–54 (2017).

92. Snyder-Mackler, N. et al. Effects of food insecurity on DNA methylation and telomere length. Am. J. Clin. Nutr. 106, 1553–1561 (2017).

93. Quach, A. et al. Quantifying and mapping DNA methylation changes at the gene level: a methodological approach. J. Proteomics 76, 17–27 (2011).

94. Snyder-Mackler, N. et al. Effect of food insecurity on DNA methylation and telomere length. Am. J. Clin. Nutr. 106, 1553–1561 (2017).

95. Harley, E. H., Knight, M. H., Lardner, C., Wooding, B. & Gregor, M. The Epigenetic Clock Development Foundation which plans to license several patents from its employer UC Regents. These patents list S.H. as inventor. The other authors declare no competing interests.

96. Snyder-Mackler, N. et al. Effect of food insecurity on DNA methylation and telomere length. Am. J. Clin. Nutr. 106, 1553–1561 (2017).

97. Kalb, S. M. et al. Use of the Epigenetic Clock for primary prevention. Neurobiology of Aging 593, 41–54 (2017).

98. Wade, C. M. et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 323, 699–714 (2010).

99. Zhou, W., Triche, T. J. Jr., Laird, P. W. & Shen, H. SeSAMe: reducing stochasticity and increasing assembly contiguity and composition. Commun. Biol. 1, 197 (2018).

100. Alam, T. et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 323, 699–714 (2010).

101. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

102. McQuillin, R. et al. Runs of homoyogosity in European populations. Hum. Genet. 83, 595–578 (2007).

103. Zeileis, A. & Hothorn, T. Diagnostic checking in regression relationships. R News 2, 7–10 (2002).

104. Zeileis, A. Econometric computing with HC and HAC covariance matrix estimates. J. Stat. Softw. 11, 1–17 (2004).

105. Zeileis, A., Koller, S. & Graham, N. Various versatile variances: an object-oriented implementation of clustered covariances in R. J. Stat. Softw. 95, 1–36 (2020).

106. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 9, 559 (2008).

107. Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. & Williams, R. M. The Adjustment During Army Life (Princeton University Press, 1949).

Acknowledgements
This work was supported by the Paul G. Allen Frontiers Group (SH). G.P. was supported by the Science Without Borders program of the National Counsel of Technological and Scientific Development of Brazil. Sample collection was supported by National Geographic Grant 8941–11 (BL). We thank the following people for their assistance with acquiring the plains zebra blood and tissue samples: Colleen O’Ryan, Stephen Mitchell, Mick D’Alton, Tom Turner, Boet le Roux, Evert Grobbeler, Ewald Groenewald, Basie and Coenraad Beuzdenhout, Linda Mason, Patricia Swanepoel, Fernando Rueda, Ross Cowlin, Melissa Stander, Hanna Lindstad, Ansel Abels, J.P. Hugo, Cobus van Coller, Jannie du Plessis, and SANParks. White Oak Conservation kindly provided Grey’s zebra and Somali wild ass samples.

Contributing authors
B.L., S.H. and G.P. conceived of the study; B.L. and G.P. analyzed data, and B.L., G.P., and S.H. co-wrote the article. The remaining authors helped with the statistical analysis (A.H., J.Z., C.L., C.F., and M.P.), or the data generation (B.W., C.J.F., C.K., G.B., T.R., and D.M.). All authors reviewed and edited the article.

Competing interests
The authors declare the following competing interests: S.H. is a founder of the non-profit Epigenetic Clock Development Foundation which plans to license several patents from its employer UC Regents. These patents list S.H. as inventor. The other authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02935-z.

Correspondence and requests for materials should be addressed to Brenda Larison or Steve Horvath.

Peer review information Communications Biology thanks Aaron Shafer and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Christopher Hine and Luke R. Grinham. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.