Speciation with gene flow in equids despite extensive chromosomal plasticity

Hákon Jónsson[a,1], Mikkel Schubert[a,1], Andaine Seguin-Orlando[a,b,1], Aurélien Ginolhac[a], Lillian Petersen[b], Matteo Fumagalli[c,d], Anders Albrechtsen[e], Bent Petersen[f], Thorfinn S. Korneliussen[a], Julia T. Vilstrup[a], Teri Lear[g], Jennifer Leigh Myka[g], Judith Lundquist[g], Donald C. Miller[h], Ahmed H. Alfarhan[i], Saleh A. Alquraishi[i], Khaled A. S. Al-Rasheid[i], Julia Stagegaard[j], Günter Strauss[k], Mads Frost Bertelsen[l], Thomas Sicheritz-Ponten[f], Douglas F. Antczak[h], Ernest Bailey[g], Rasmus Nielsen[c], Eske Willerslev[a], and Ludovic Orlando[a,2]

[a]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen K, Denmark; [b]National High-Throughput DNA Sequencing Center, DK-1353 Copenhagen K, Denmark; [c]Department of Integrative Biology, University of California, Berkeley, CA 94720; [d]UCL Genetics Institute, Department of Genetics, Evolution, and Environment, University College London, London WC1E 6BT, United Kingdom; [e]The Bioinformatics Centre, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; [f]Centre for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark; [g]Maxwell H. Gluck Equine Research Center, Veterinary Science Department, University of Kentucky, Lexington, KY 40546; [h]Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853; [i]Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; [j]Ree Park, Ebeltoft Safari, DK-8400 Ebeltoft, Denmark; [k]Tierpark Berlin-Friedrichsfelde, 10319 Berlin, Germany; and [l]Centre for Zoo and Wild Animal Health, Copenhagen Zoo, DK-2000 Frederiksberg, Denmark

Edited by Andrew G. Clark, Cornell University, Ithaca, NY, and approved October 27, 2014 (received for review July 3, 2014)

Horses, asses, and zebras belong to a single genus, *Equus*, which emerged 4.0–4.5 Mya. Although the equine fossil record represents a textbook example of evolution, the succession of events that gave rise to the diversity of species existing today remains unclear. Here we present six genomes from each living species of asses and zebras. This completes the set of genomes available for all extant species in the genus, which was hitherto represented only by the horse and the domestic donkey. In addition, we used a museum specimen to characterize the genome of the quagga zebra, which was driven to extinction in the early 1900s. We scan the genomes for lineage-specific adaptations and identify 48 genes that have evolved under positive selection and are involved in olfaction, immune response, development, locomotion, and behavior. Our extensive genome dataset reveals a highly dynamic demographic history with synchronous expansions and collapses on different continents during the last 400 ky after major climatic events. We show that the earliest speciation occurred with gene flow in Northern America, and that the ancestor of present-day asses and zebras dispersed into the Old World 2.1–3.4 Mya. Strikingly, we also find evidence for gene flow involving three contemporary equine species despite chromosomal numbers varying from 16 pairs to 31 pairs. These findings challenge the claim that the accumulation of chromosomal rearrangements drive complete reproductive isolation, and promote equids as a fundamental model for understanding the interplay between chromosomal structure, gene flow, and, ultimately, speciation.

equids | evolutionary genomics | speciation | admixture | chromosomal rearrangements

The rich fossil record of the equid family has provided one of the most famous examples of evolutionary transition. It illustrates 55 million years of anatomic changes, starting from small three- to four-toed ancestors and leading to the one-toed ungulates that survive today (1). Although equids originated in the New World, they diversified into several dozens of recognized genera adapted to a diversity of environments in both the Old World and the New World during the Miocene and Oligocene (1). Most of this past diversity is now extinct, and all living members of the equid family belong to a single genus, *Equus*, which most likely emerged some 4.0–4.5 Mya (2).

Along with domestic horses and donkeys, current living equids include asses and zebras, with natural habitats spread across Africa and Eurasia. In Eurasia, hemiones (*Equus hemionus*) and Tibetan kiangs (*Equus kiang*) represent the so-called Asiatic wild asses. Although historically found across a wide geographic range, including Mongolia, Central Asia, Anatolia, and Russia, hemiones are now classified as endangered by the International Union for Conservation of Nature. Their current range is restricted to the border of Mongolia and China, as well as small distant regions in which a number of local subspecies are recognized, including the Iranian onager (3). In contrast, Tibetan kiangs, whose International Union for Conservation of Nature status is "least concern," inhabits a wide range of the Tibetan plateau, reaching altitudes as high as 5,400 m.

Wild asses also can be found in Africa, where they are represented by the critically endangered Somali wild asses (*Equus africanus somaliensis*) and Nubian wild asses (*Equus africanus africanus*) from arid and desert habitats of Egypt, Erithrea, Ethiopia, and Somalia (4). African equids also are represented by three species of zebras, one of which, the Grevy's zebra (*Equus grevyi*), is currently endangered. Historically, the latter was distributed in the desert of Erithrea, overlapping with the range of Somali wild asses, but is now restricted to more southern territories, namely Ethiopia

Significance

Thirty years after the first DNA fragment from the extinct quagga zebra was sequenced, we set another milestone in equine genomics by sequencing its entire genome, along with the genomes of the surviving equine species. This extensive dataset allows us to decipher the genetic makeup underlying lineage-specific adaptations and reveal the complex history of equine speciation. We find that *Equus* first diverged in the New World, spread across the Old World 2.1–3.4 Mya, and finally experienced major demographic expansions and collapses coinciding with past climate changes. Strikingly, we find multiple instances of hybridization throughout the equine tree, despite extremely divergent chromosomal structures. This contrasts with theories promoting chromosomal incompatibilities as drivers for the origin of equine species.

Author contributions: R.N. and L.O. designed research; H.J., M.S., A.S.-O., L.P., J.T.V., T.L., J.L.M., J.L., D.C.M., D.F.A., and L.O. performed research; H.J., M.S., A.S.-O., A.H.A., S.A.A., K.A.S.A.-R., J.S., G.S., M.F.B., T.S.-P., E.B., E.W., and L.O. contributed new reagents/analytic tools; H.J., M.S., A.G., M.F., A.A., B.P., T.S.K., T.L., E.B., and L.O. analyzed data; and L.O. wrote the paper, with input from H.J., M.S., K.A.S.A.-R. and all coauthors.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All sequences are deposited in the European Nucleotide Archive (accession no. PRJEB7446).

[1]H.J., M.S., and A.S.-O. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: Lorlando@snm.ku.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412627111/-/DCSupplemental.

and Kenya (5). A second species of zebra, the plains zebra (*Equus quagga*), represents the most abundant African equid, found from Ethiopia to South Africa and as far west as the coast of Namibia. The last African equid, the mountain zebra (*Equus zebra*), also is found in Namibia, along with patchy isolates in South Africa.

The succession of events that gave rise to the full diversity of extant species remains unclear, mainly because the paleontologic record has been split into an excessive number of taxonomic assemblages and is incomplete in some regions of the world (6, 7). In addition, the available genetic information is relatively limited, with complete genomes available only for the horse and the domestic donkey (2, 8, 9). Moreover, only a handful of mitochondrial genomes (10) and nuclear genes (11), along with limited genome-wide single nucleotide variants (SNVs) ascertained to horses (12), have been characterized. Although wild populations have been genetically monitored for mitochondrial variation (13, 14) and microsatellite data (15), the amount of genetic information available is incompatible with detailed reconstructions of their population history.

In this study, we generated the first complete whole-genome dataset of extant species in the equid family, complemented by the genome of the quagga zebra, a recently extinct conspecific of the plains zebra (16). In the process, we also have provided a valuable resource for the genetic management of endangered species. We used the genome information to investigate the tempo and mode of speciation of equids and to identify their past population history over the last 4 million years at an unprecedented level of detail. Equid species are known to exhibit very divergent karyotypes, ranging from 16 to 33 chromosomal pairs (17), which might have acted as a barrier to gene flow (18, 19); thus, we also used our genome dataset to investigate whether this chromosomal plasticity influenced the population history of equids.

Results and Discussion

We prepared indexed DNA libraries from the blood of captive animals for the Somali wild ass (SOM), onager (ONA), Tibetan kiang (KIA), and Grevy's (GRE), mountain (HAR), and plains (BOE) zebras. Ancient DNA was extracted from the hairs of the extinct quagga (QUA) in dedicated clean laboratory facilities, where indexed DNA libraries were prepared as well (*SI Appendix*, sections 1 and 2). We generated and aligned ∼7.3 billion reads against the horse reference genome (8) and a draft assembly for the domestic donkey (2). Applying stringent filters, we achieved average genome coverage of 13.3×–21.5× for living equids and 7.9×–8.1× for the extinct quagga (*SI Appendix*, Tables S8 and S9). Sequencing error rates were low after quality filtering (0.009–0.040% per base) for all except the extinct quagga (0.454%), for which typical signatures of DNA degradation support data authenticity (*SI Appendix*, sections 2.6 and 2.7).

We identified ∼55.9 million SNVs across the extant species (Fig. 1*A*), most of which were intergenic (*SI Appendix*, Tables S12–S18). We compared those against a dataset of ∼54,000 SNVs (12) and found that our individuals clustered with conspecific members (*SI Appendix*, Figs. S19–S22). From our full set of SNVs, we identified ∼14.9 million potentially ancestry-informative markers as variants segregating in a single species (*SI Appendix*, section 3.3). This represents a valuable resource for wildlife management, given the limited genetic information hitherto available (12–14) and the fact that three species are either classified as endangered (ONA and GRE) (3, 5) or critically endangered (SOM) (4).

We next compared the inbreeding present in our genomes and previously reported horse (TWI) and donkey (DON) genomes (Fig. 1*B*). No inbreeding was detected for the plains zebra or mountain zebra. In contrast, the Somali wild ass appeared to be extremely inbred, as expected from its pedigree (*SI Appendix*, section 3.4). In all other extant species, levels of inbreeding were greater than expected based on pedigree information.

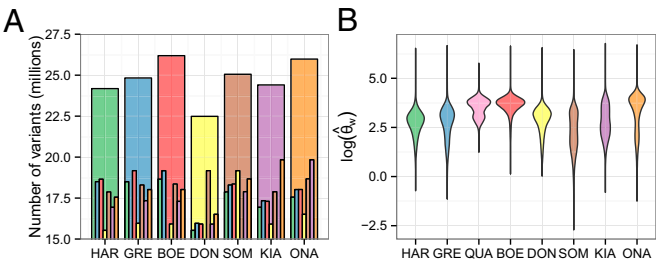

Fig. 1. SNV overlap, heterozygosity, and inbreeding. (*A*) Number of SNVs identified with respect to horse, in millions. Smaller bars indicate the overlap with the other species sequenced in this study and the donkey (DON) from Orlando et al. (2). (*B*) Genome-wide distribution of heterozygosity values inferred from the θ-Watterson estimator in 50-kb blocks. The analysis was restricted to transversions for comparison with the QUA sample.

Interestingly, the extinct quagga showed a similar level of inbreeding as a captive Przewalski's horse sequenced previously (2).

We reconstructed the equine phylogeny using maximum likelihood inference and 20,374 protein-coding genes (Fig. 2*B*). The topology is in line with previous reports based on complete mitochondrial sequences (10) and ∼50 k genome-wide SNVs (12). Noncaballine equids were found to cluster into monophyletic zebras and asses, separating ∼1.69–1.99 Mya. This is significantly younger than previous estimates based on limited genetic information (10, 11). African and Asiatic asses diverged soon after (∼1.47–1.75 Mya), in agreement with the appearance of donkeys in the fossil record (20). A parallel radiation occurred in Africa 1.28–1.59 Mya, leading to the three main species of zebras within ∼200 ky (*SI Appendix*, section 6). The extinct quagga was confirmed to be closely related to the plains zebra (10, 16, 21).

Living equids exhibit strikingly divergent karyotypes, ranging from 16 chromosome pairs in mountain zebras to 33 chromosome pairs in Przewalski's horses (17). We inferred the most likely karyotype of each phylogenetic branch and found overall higher rates of chromosomal loss than chromosomal gain, in line with an *Equus* ancestor having 38–42 pairs of chromosomes and the known prevalence of Robertsonian fissions/fusions in equids (22) (Fig. 2*A*). Particularly high rates of chromosomal changes were observed for three branches. The first two of these branches correspond to the caballine/noncaballine divergence, suggesting extremely karyotypic plasticity 4.0–4.5 Mya with 8.34 and 17.7 chromosomal gains and losses already accumulated by the time of the most recent common noncaballine ancestor (∼1.69–1.99 Mya). The third branch, leading to mountain zebras, experienced almost four times more chromosome losses than gains, resulting in the smallest number of chromosomes in the entire genus ($2n = 32$).

We next mapped mutations on the equine tree and identified signatures of selection. We found an acceleration of gene loss in the branch ancestral to asses (Fig. 2*B* and *SI Appendix*, section 6), with rates remaining high in African asses but returning to zebra-like levels in Asiatic asses. We identified 84 mutations in microRNAs, 58 of which were specific to a single lineage and representing potential candidates for regulatory changes. We detected 48 genes with dN/dS rates suggestive of positive selection (Fig. 2*B* and *SI Appendix*, section 7). This finding reveals key adaptive changes that participated in the genetic makeup of each equine species, with functions involved in cellular interactions (*DSG1*, *NCAN*, *TBC1D2*) and trafficking (*MYO5C*, *ZW10*), metabolism (*ACCS*, *SLC9A4*, *HMGCR*), development (*SEMA5A*, *FAT2*), olfaction (*BPIFB4*), and immunity (*ITIH4*, *SIGLEC1*). Interestingly, mountain zebras showed positive selection signatures at *VPS13A*, a gene associated with locomotory and behavioral disorders (23, 24), possibly in relation to their high sociability. The antiporter *SLC9A4*, involved in pH regulation and countering of adverse environmental

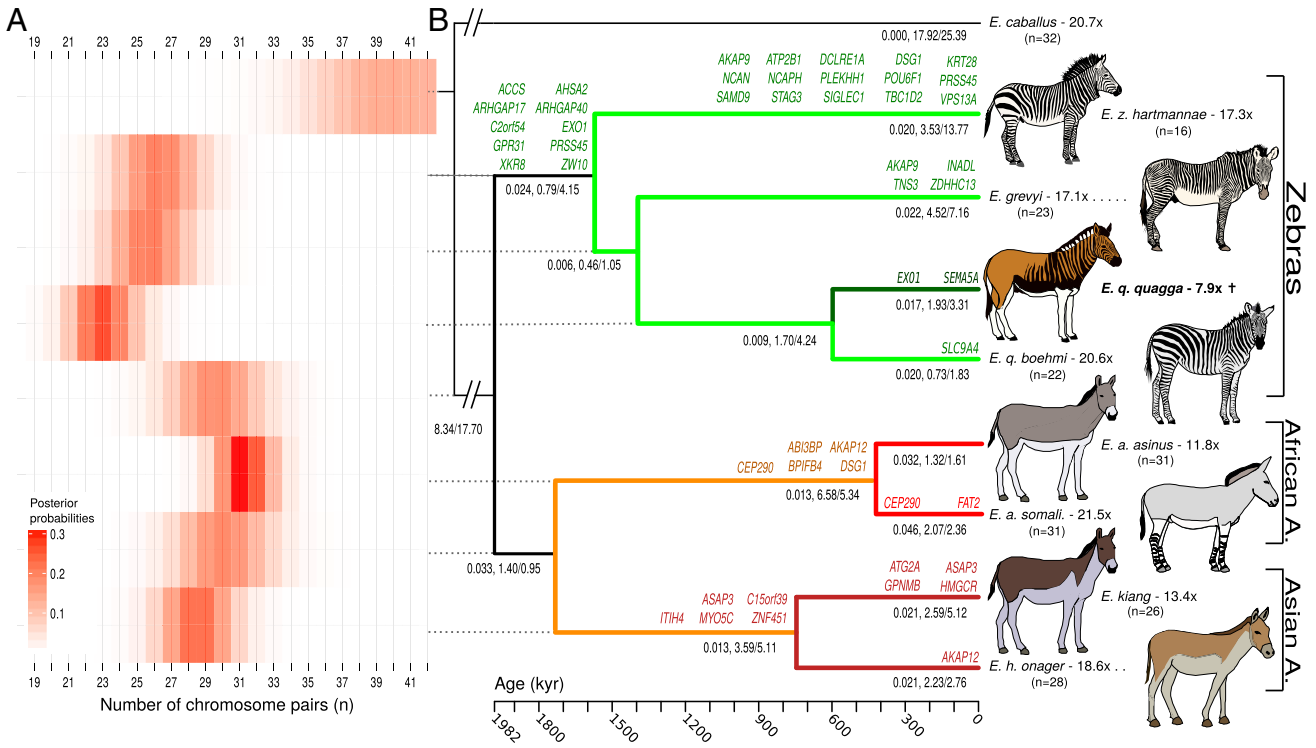

Fig. 2. Equine phylogeny, selection scan, and karyotypic and mutational changes. (*A*) Inferred number of chromosome pairs for each ancestral node. (*B*) Phylogenetic chronogram of lineage divergence in equids based on a relaxed molecular clock. All nodes received 100% bootstrap support. The names of the genes showing evidence of positive selection are reported above the branches concerned. The numbers provided below branches refer to rates of gene loss and chromosome gains and losses, respectively. The numbers of chromosome pairs (dominant form) are indicated below species names.

conditions, was found to be positively selected in plains zebras, which experience a wide range of environments throughout their geographical range. Similarly, the axonal guidance factor *SEMA5A*, associated with cranial vascular patterning in mice (25) and hippocampal volume (26) and autism in humans (27), was found to have undergone positive selection in the extinct quagga.

We next reconstructed the population history of each species, which revealed extremely dynamic demographic profiles (Fig. 3). Most species expanded after the Eemian (~125 kya) before collapsing during the last 30 ky, possibly related to the climatic changes of the Late Pleistocene. This post-Eemian expansion signal was weaker for Grevy's zebras and Somali wild asses and absent for kiangs. An earlier period of expansion–collapse, peaking ~200–250 kya, was inferred for horses, onagers, and plains zebras during the interglacial marine isotope stage 7 (28), a time when the population size of other species was limited. This suggests that major climatic changes allowed synchronous expansions of some species living in different continents, where they successfully exploited the increasing availability of their ecological niches. Earlier demographic trajectories were asynchronous for horses, asses, and zebras, suggesting complex ecological dynamics in Africa, Eurasia, and America before 400 kya.

We estimated population time splits using the first divergence date in the demographic profiles of sister species and coalescent-based simulations (Figs. 3 and 4 and *SI Appendix*, section 9). We found that the early population split between Asiatic and African asses occurred ~1.7 Mya. Onagers and kiang populations diverged more recently (~266–392 kya), whereas the three populations of living zebras had already diverged ~1.1 Mya. The extinct quagga split from the plains zebra population only ~233–356 kya.

Using the *D* statistics approach (29), which tests for an excess of shared polymorphisms between one of two closely related lineages (E1 or E2) and a third lineage (E3), we investigated whether equine populations are connected by gene flow (*SI Appendix*, Tables S39 and S40). Admixture was supported between the kiang and the donkey [*D*(ONA,KIA;DON,TWI) = 0.02, *z*-score = 6.05], but not with the Somali wild ass. This indicates that gene flow was not reciprocal and occurred mainly from the kiang (or a closely related population) into the donkey population after the African ass population split ~126–188 kya (Fig. 4). The *D* statistics also support an excess of shared polymorphisms between asses and zebras, with a large majority of the significant tests (19 of 24) involving the Grevy's zebra and/or Somali wild ass [*D*(asses,SOM;GRE,TWI) = 0.03–0.144; *z*-score = 5.95–29.71]. *D* statistics within 1-Mb windows were strongly correlated when varying the species considered for E3 (*SI Appendix*, Figs. S43 and S44), as expected for closely related species (*SI Appendix*, section 8.3). This finding suggests that the same genomic blocks, and thus the same admixture event, was in fact driving these results. Given the predominance of Grevy's zebra and the Somali wild ass in significant

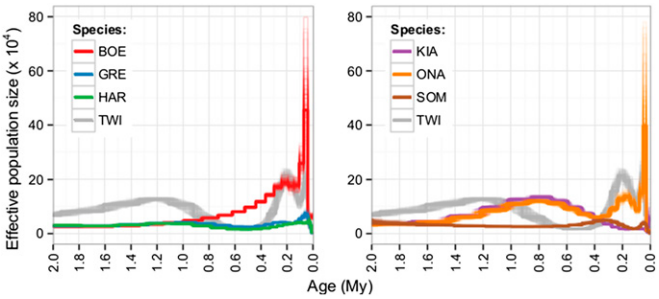

Fig. 3. PSMC demographic profiles for all extant equid species over the last 2 million years.

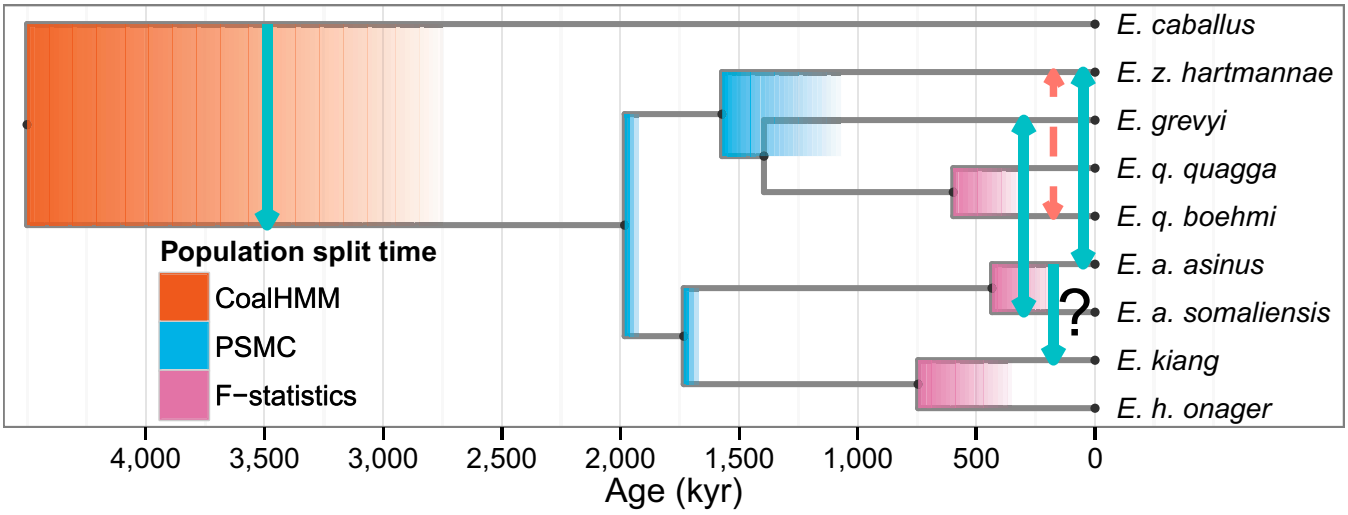

Fig. 4. Proposed population model for equids. Events of gene flow between populations are represented by arrows with directionality if available. The exact timing of such events is not known at present. The dashed arrow indicates the result of a significant admixture test that likely reflects the consequence of the gene flow detected between *E. a. somaliensis* and *E. grevyi*. Divergence (time to the most recent common ancestor, TMRCA) and population split times are indicated by darker and lighter ends of the colored rectangles, respectively.

admixture tests, we conclude that the admixture between the two species occurred once both were established. This is in line with the significant overlap between the historical ranges of both species and their shared territorial mating system (4, 5), which contrasts with the more common harem-like social structure found in other equids.

In addition, the *D* statistics indicated admixture with the mountain zebra [*D*(BOE,HAR;asses,TWI) = 0.014–0.024; *z*-score = 3.11–5.13]. This was the case for all asses except the Somali wild ass. The latter is likely a consequence of the gene flow from Grevy's zebra that brought into the Somali wild ass alleles shared with the plains zebra, which compensated for the influx of alleles from the mountain zebra. Because the mountain zebra emerged after divergence of the Asiatic and African ass populations (Fig. 4), we conclude that the admixture could not have occurred between the mountain zebra and the population ancestral to all asses but with one of the ass lineages. This conclusion is consistent with the strong correlation among the *D* statistics for all four tests, suggesting the presence of a single admixture event (*SI Appendix*, Fig. S45). Given that zebras are exclusively African, we propose that this admixture occurred in Africa rather than in Asia.

Finally, we uncovered the signature of a possible admixture between plains and mountain zebras [*D*(BOE,GRE;HAR,TWI) = −0.06; *z*-score = −17.98]. *D* statistics were significantly correlated when considering mountain zebras or Somali wild asses as E3, however (*SI Appendix*, Fig. S43). This indicates that the outcome of the *D*(BOE,GRE;HAR,TWI) test likely reflects the gene flow from the Somali wild ass into Grevy's zebra, and not admixture between plains and mountain zebras.

We next used the coalescent hidden Markov model approach HMMCoal (30) to detect gene flow during the first divergence within *Equus* (*SI Appendix*, section 10.1). Our analysis suggests that noncaballine equids did not emerge after a founder event into the Old World, but rather maintained gene flow in North America with caballine equids. We estimate that this gene flow ceased 2.1–3.4 Mya, which closely matches the paleontological evidence for the noncaballine dispersal out of America (31). The site frequency spectrum between horses and African asses also supports gene flow between caballines and noncaballines, mostly from caballines into noncaballines, until 2.9–3.8 Mya (*SI Appendix*, section 10.2), Given that hinnies ($\male_{horse} \times \female_{donkey}$) show significantly lower pregnancy rates than mules ($\male_{donkey} \times \female_{horse}$; *SI Appendix*, section 10.3), we suggest that this gene flow was sex-biased and involved mostly male noncaballines and female caballines and further hybrid backcrosses with noncaballines. Interestingly, although mules are generally sterile, a similar quartet involving a fertile female mule, a male donkey, and two offspring was reported recently (32).

Overall, we found evidence for four main episodes of gene flow among equids: one during the earliest *Equus* divergence, one from kiang ($2n = 51–52$) into the donkey lineage (2n = 62–64), one between the Somali wild ass ($2n = 62$) and Grevy's zebra ($2n = 46$), and one between African asses (SOM-DON; $2n = 62–64$) and the mountain zebra ($2n = 32$). We excluded contamination and inadvertent crosses in captivity as sources of the detected signals (*SI Appendix*, sections 4 and 8.4). Thus, we conclude that such massive karyotypic changes have not resulted in full reproductive isolation, in stark contrast with theories assuming that chromosomal impairment during meiosis is responsible for complete sterility in hybrids (18), but in agreement with the description of fertile offspring across equine species (32, 33).

We next used previously published data (17) and FISH to perform the first cytogenetic characterization of the onager. We defined 14 large-scale (>10 Mb) regions with chromosomal changes and tested whether they influenced gene flow (*SI Appendix*, section 11). A linear model for the *D* statistics (1-Mb windows), with the cytogenetic regions as the covariate, was preferred over a model disregarding the structural information when using *D*(DON,SOM;GRE,TWI) ($P < 0.0001$), supporting the contention that chromosomal rearrangements impact gene flow (19). Interestingly, a 46-Mb region at chromosome 2 appeared to have introduced heterogeneity in the gene flow (Fig. 5) and was significantly enriched in genes involved in lipid metabolism.

High-resolution structural characterization (34, 35) will be required to fully explore the precise influence of chromosomal rearrangements on gene flow. For now, our study provides the first empirical evidence connecting gene flow and chromosomal rearrangements in equids. The extreme plasticity of their karyotype and the presence of multiple gene flows make equids a promising model for deciphering the role of chromosomal rearrangements in speciation.

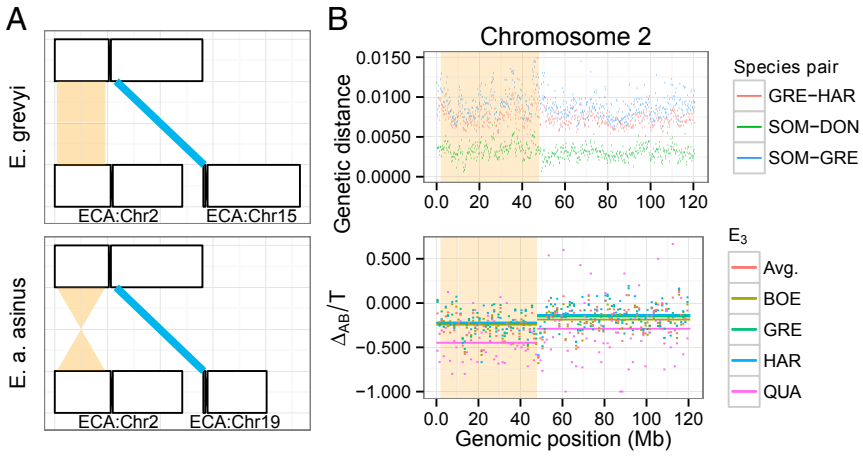

Fig. 5. Karyotypic rearrangements and gene flow in equids. (*A*) An example of karyotypic rearrangement between positions 1,920,626 and 47,900,194 on the second horse chromosome. The filled segment encloses the area between two BAC markers from Musilova et al. (17). (*B*) Genetic distance (*Upper*) and *D* statistics (Δ_{AB}/T, *Lower*) across the genomic region showing a chromosomal rearrangement. The genetic distance is calculated within 200-kb nonoverlapping windows for all species pairs involved in the proposed SOM-GRE admixture event. The *D* statistics values are calculated within 1-Mb nonoverlapping windows for *D*(SOM,DON;X,TWI), where X is DON, KIA, ONA, or SOM.

Methods

Genome Sequencing. DNA from hairs of the extinct quagga were extracted and prepared into indexed Illumina libraries in the ancient DNA facilities of the Centre for GeoGenetics. DNA from the six extant equine species was extracted in other laboratory facilities, one species per extraction session, together with one extraction blank. Illumina sequencing was performed on a HiSEq 2000 platform at the Danish National High-Throughput DNA Sequencing Centre (*SI Appendix*, sections 1 and 2).

Read trimming, mapping, and variant calling were performed with the PALEOMIX pipeline (36). Postmortem DNA damage was quantified and visualized using mapDamage2.0 (37). Average sequencing error rates were determined relative to the genome of the Somali wild ass, assuming that the species are equally distant to the horse reference genome (*SI Appendix*, section 2).

Genomic Variation. The variant calls from PALEOMIX were used to define species or clade specific markers among the sequenced individuals, including ancestry informative markers. Heterozygosity per individual was quantified with the θ-Watterson estimator in 50-kb windows (10-kb step size) using angsd (www.popgen.dk/angsd). The level of inbreeding was estimated based on the genomic coverage of low-heterozygosity tracts, as described previously (38). Contamination levels were quantified by the fraction of non-majority bases at clade-informative SNV, defined with markers from McCue et al. (12) and the known mitochondrial diversity for all equid groups (*SI Appendix*, sections 3 and 4).

Demographic Inference. Demographic profiles were estimated with a pairwise sequential Markov coalescent (PSMC) (39) for all species except quagga, owing to insufficient coverage. The spread of the estimates was assessed with 100 bootstraps corresponding to sampling with replacement of 100 random genomic segments (500 kb). For low-coverage (<20×) genomes, uniform false-negative rates were determined empirically using down-sampled datasets, by recovering the original PSMC profiles for samples with higher coverage (>20×) (*SI Appendix*, section 5).

Phylogenomic Inference. A maximum likelihood framework as implemented in PALEOMIX was used for phylogenomic reconstructions (36), and molecular dating was done using r8s (40). Nonsynonymous mutations in coding sequences and putative regulatory variants were mapped parsimoniously on the resulting phylogenetic tree. The most likely karyotype for each node of the equine tree was estimated using chromEvol (41) and known karyotypes for extant species (*SI Appendix*, section 6).

Selection Scans. Selection scans based on dN/dS ratios were performed using PAML (42) and a false-discovery rate-modulated Monte Carlo framework (43) to assess the significance. Putative regions under selection were identified by tracking local declines in the genetic diversity of predefined equine groups using θ-Watterson estimates from angsd, calculated within 50-kb sliding windows (*SI Appendix*, section 7).

Population Admixture. Admixture tests were implemented using *D* statistics (29) calculated in 10-Mb windows, and subsequently jackknifed (44) to statistically assess significance. Windows of 1 Mb were used to quantify the correlation among the *D* statistics when varying the E3 species in the subsequent tests (E1, E2;E3,Outgroup). The f_3 statistics were calculated using the TreeMix implementation (45), with the spread of the estimate assessed by the accompanying block resampling using groups of 10,000 variants (*SI Appendix*, section 8).

Population Divergence. Recent population splits were timed using F-statistics (46) coupled with coalescent simulations and approximate Bayesian computation (*SI Appendix*, section 9), following the procedure described by Orlando et al. (2).

Speciation Mode. The hidden Markov coalescent framework CoalHMM (30) was used to estimate the mode (allopatry or with gene flow) and times of the pairwise equine population splits. The demography parameters were estimated in 10-Mb nonoverlapping windows, and the Akaike information criterion (47) was used to determine the mode of the split.

In addition, the joint site frequency spectrum (SFS) for the African asses and the horses served as input for ∂a∂i (48). The ∂a∂i model selection was performed using a likelihood ratio test. The precision of the *P* value was assessed by sampling one site per nonoverlapping (100-kb) window across the genome and applying ∂a∂i on the sampled SFS (*SI Appendix*, section 10).

Chromosomal Structure and Gene Flow. Onager cytogenetic characterization was done at the Gluck Center, University of Kentucky, using FISH of horse BAC clones. Karyotype-variant regions (KVRs) were defined as minimal regions amid cytogenetic markers (17) enclosing a structural difference between the species. The local *D*(SOM,DON;GRE,TWI) statistics (1 Mb) were modeled with a linear random-effects model with a KVR random effect. The modeling was implemented with lme4 (49), and the significance of the KVR parameter was assessed with a parametric bootstrap (*SI Appendix*, section 11).

Data Availability. The sequencing data generated in this study for QUA, BOE, GRE, HAR, KIA, ONA, and SOM have been deposited in the European Nucleotide Archive (accession no. PRJEB7446).

ACKNOWLEDGMENTS. We thank T. Brand, P. Selmer Olsen, and the laboratory technicians at the Danish National High-Throughput DNA Sequencing Center for technical assistance; T. Mailund for help with HMMCoal analyses; J. N. MacLeod and T. Kalbfleisch for discussions about the horse reference genome assembly; R. M. Guðmundsdóttir for drawing the equid figures; J. Fronczek and M. Houck (San Diego Zoo) for help

accessing onager metaphasic chromosomes; the San Diego Zoological Society's Genetics Laboratory at the Beckman Institute for onager slides; and C. D. Sarkissian, C. Gamba, L. Ermini, and R. Fernandez for fruitful discussions. This work was supported by funding from the Danish Council for Independent Research, Natural Sciences; the Danish National Research Foundation (Grant DNFR 94); Marie-Curie Actions (Career Integration Grant FP7 CIG-293845); and the International Research Group Program (Project IRG14-08) of the Deanship of Scientific Research, King Saud University. A.G. was supported by a Marie-Curie Intra-European Fellowship (FP7 IEF-299176). H.J. was supported by a Marie-Curie Initial Training Network EUROTAST Grant (FP7 ITN-290344). M.S. was supported by a Lundbeck Foundation Grant (R52-A5062). J.M. was supported by a Geoffrey C. Hughes Fellowship. This paper is published in connection with a project of the University of Kentucky Agricultural Experiment Station (paper no. 14-14-047).

1. Macfadden BJ (2005) Fossil horses—evidence for evolution. *Science* 307(5716): 1728–1730.
2. Orlando L, et al. (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78.
3. Moehlman PD, Shah N, Feh C (2008) *Equus hemionus*. Available at: www.iucnredlist.org. Accessed July 1, 2014.
4. Moehlman PD, Yohannes H, Teclai R, Kebede F (2008) *Equus africanus*. Available at: www.iucnredlist.org. Accessed July 1, 2014.
5. Moehlman PD, Rubenstein DI, Kebede F (2013) *Equus grevyi*. Available at: www.iucnredlist.org. Accessed July 1, 2014.
6. Weinstock J, et al. (2005) Evolution, systematics, and phylogeography of pleistocene horses in the new world: A molecular perspective. *PLoS Biol* 3(8):e241.
7. Orlando L, et al. (2009) Revising the recent evolutionary history of equids using ancient DNA. *Proc Natl Acad Sci USA* 106(51):21754–21759.
8. Wade CM, et al.; Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326(5954):865–867.
9. Andersson LS, et al. (2012) Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature* 488(7413):642–646.
10. Vilstrup JT, et al. (2013) Mitochondrial phylogenomics of modern and ancient equids. *PLoS ONE* 8(2):e55950.
11. Steiner CC, Ryder OA (2011) Molecular phylogeny and evolution of the Perissodactyla. *Zool J Linn Soc* 163:1289–1303.
12. McCue ME, et al. (2012) A high density SNP array for the domestic horse and extant Perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* 8(1):e1002451.
13. Lorenzen ED, Arctander P, Siegismund HR (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): Insights from mtDNA and microsatellites. *Mol Ecol* 17(12):2812–2824.
14. Kimura B, et al. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc Biol Sci* 278(1702):50–57.
15. Bowling AT, et al. (2003) Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. *Cytogenet Genome Res* 102(1-4):226–234.
16. Leonard JA, et al. (2005) A rapid loss of stripes: The evolutionary history of the extinct quagga. *Biol Lett* 1:291–5.
17. Musilova P, Kubickova S, Vahala J, Rubes J (2013) Subchromosomal karyotype evolution in Equidae. *Chromosome Res* 21(2):175–187.
18. Faria R, Navarro A (2010) Chromosomal speciation revisited: Rearranging theory with pieces of evidence. *Trends Ecol Evol* 25(11):660–669.
19. Stevison LS, Hoehn KB, Noor MA (2011) Effects of inversions on within- and between-species recombination and divergence. *Genome Biol Evol* 3:830–841.
20. Eisenmann V (1992) Origins, dispersals, and migrations of *Equus* (Mammalia, Perissodactyla). *Cour Forschungsinstitut Senckenb* 153:161–170.
21. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991):282–284.
22. Murphy WJ, Stanyon R, O'Brien SJ (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. *Genome Biol* 2(6):S0005.
23. Tomiyasu A, et al. (2011) Novel pathogenic mutations and copy number variations in the *VPS13A* gene in patients with chorea-acanthocytosis. *Am J Med Genet B Neuropsychiatr Genet* 156B(5):620–631.
24. Shimo H, et al. (2011) Comprehensive analysis of the genes responsible for neuroacanthocytosis in mood disorder and schizophrenia. *Neurosci Res* 69(3):196–202.
25. Fiore R, Rahim B, Christoffels VM, Moorman AFM, Püschel AW (2005) Inactivation of the *Sema5a* gene results in embryonic lethality and defective remodeling of the cranial vascular system. *Mol Cell Biol* 25(6):2310–2319.
26. Zhu B, et al. (2013) The *SEMA5A* gene is associated with hippocampal volume, and their interaction is associated with performance on Raven's progressive matrices. *Neuroimage* 88C:181–187.
27. Cheng Y, Quinn JF, Weiss LA (2013) An eQTL mapping approach reveals that rare variants in the *SEMA5A* regulatory network impact autism risk. *Hum Mol Genet* 22(14): 2960–2972.
28. Dutton A, et al. (2009) Phasing and amplitude of sea-level and climate change during the penultimate interglacial. *Nat Geosci* 2:355–359.
29. Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28(8):2239–2252.
30. Mailund T, et al. (2012) A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet* 8(12):e1003125.
31. Lindsay EH, Opdyke ND, Johnson NM (1980) Pliocene dispersal of the horse *Equus* and late Cenozoic mammalian dispersal events. *Nature* 287:135–138.
32. Steiner CC, Ryder OA (2013) Characterization of Prdm9 in equids and sterility in mules. *PLoS ONE* 8(4):e61746.
33. Cordingley JE, et al. (2009) Is the endangered Grevy's zebra threatened by hybridization? *Anim Conserv* 12:505–513.
34. Huang J, et al. (2014) Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci Rep* 4:4958.
35. McCoy RC, et al. (2014) Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly repetitive transposable elements. *PLoS ONE* 9(9):e106689.
36. Schubert M, et al. (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* 9(5):1056–1082.
37. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
38. Prüfer K, et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.
39. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
40. Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19(2):301–302.
41. Mayrose I, Barker MS, Otto SP (2010) Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst Biol* 59(2):132–144.
42. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
43. Sandve GK, Ferkingstad E, Nygård S (2011) Sequential Monte Carlo multiple testing. *Bioinformatics* 27(23):3235–3241.
44. Busing F, Meijer E, Van Der Leeden R (1999) Delete-m jackknife for unequal m. *Stat Comput* 9:2–7.
45. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8(11):e1002967.
46. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.
47. Akaike H (1974) A new look at the statistical model identification. *Autom Control IEEE Trans* 19:716–723.
48. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.
49. Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. arXiv:1406.5823.