

RESEARCH

Open Access



Genome collinearity analysis illuminates the evolution of donkey chromosome 1 and horse chromosome 5 in perissodactyls: A comparative study

Shaohua Li^{1,2,3}, Gaoping Zhao³, Hongmei Han⁴, Yunxia Li^{1,3}, Jun Li³, Jinfeng Wang², Guifang Cao⁵ and Xihe Li^{1,3*}

Abstract

Background: It is important to resolve the evolutionary history of species genomes as it has affected both genome organization and chromosomal architecture. The rapid innovation in sequencing technologies and the improvement in assembly algorithms have enabled the creation of highly contiguous genomes. DNA Zoo, a global organization dedicated to animal conservation, offers more than 150 chromosome-length genome assemblies. This database has great potential in the comparative genomics field.

Results: Using the donkey (*Equus asinus asinus*, EAS) genome provided by DNA Zoo as an example, the scaffold N50 length and Benchmarking Universal Single-Copy Ortholog score reached 95.5 Mb and 91.6%, respectively. We identified the cytogenetic nomenclature, corrected the direction of the chromosome-length sequence of the donkey genome, analyzed the genome-wide chromosomal rearrangements between the donkey and horse, and illustrated the evolution of the donkey chromosome 1 and horse chromosome 5 in perissodactyls.

Conclusions: The donkey genome provided by DNA Zoo has relatively good continuity and integrity. Sequence-based comparative genomic analyses are useful for chromosome evolution research. Several previously published chromosome painting results can be used to identify the cytogenetic nomenclature and correct the direction of the chromosome-length sequence of new assemblies. Compared with the horse genome, the donkey chromosomes 1, 4, 20, and X have several obvious inversions, consistent with the results of previous studies. A 4.8 Mb inverted structure was first discovered in the donkey chromosome 25 and plains zebra chromosome 11. We speculate that the inverted structure and the tandem fusion of horse chromosome 31 and 4 are common features of non-caballine equids, which supports the correctness of the existing *Equus* phylogeny to an extent.

Keywords: Chromosome-length genome assemblies, Comparative genomes, Chromosome evolution, DNA zoo

* Correspondence: lixh@imu.edu.cn

¹Research Center for Animal Genetic Resources of Mongolia Plateau, College of Life Sciences, Inner Mongolia University, Hohhot 010070, China

³Inner Mongolia Saikexing Institute of Breeding and Reproductive Biotechnology in Domestic Animal, Hohhot 011517, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Resolving the evolutionary history of species genomes is important for our understanding of genome organization and chromosomal architecture [1]. Karyotype or chromosomal evolution has traditionally been inferred using classical and molecular cytogenetic approaches, such as comparisons of G-banded chromosome patterns and fluorescence in situ hybridization (FISH), with limited resolution [2–8]. FISH-based methods do not have sufficient resolution to permit accurate identification of evolutionary breakpoint regions and fine-scale rearrangements [7, 8]. Owing to the lowering of the cost associated with sequencing, an increasing number of species genome scaffolds have reached chromosome length, although contig N50 still needs to be improved. This allows the inference of most of the inversions, translocations, fusions, and fissions that occurred during evolution by simple observational comparisons on a sequence-based whole-genome scale [9, 10]. Nevertheless, we should realize that the genome assembly is not completely correct [11, 12]. Only relatively conservative structural variations or analytical results supported by experimental evidence at the molecular or cellular level are credible [12]. In addition, comparative genomic analysis at the sequence level is beneficial in revealing the details of variation [12]. Therefore, when we demonstrate the use of genome assembly data for exploring the chromosomal evolution, the selected chromosomes have an extensive basis of molecular cytology research, and for highlighting detailed advantages, the selected structural variations are more complex or small. The perissodactyl karyotypes are slowly evolving in ceratomorphs and extremely rapidly evolving in equids [6]. The comparative genomic research of this group is relatively thorough, especially in equine animals; region-specific painting and bacterial artificial chromosome probes are used to determine the direction of evolutionarily conserved segments with respect to centromere positions [6, 13, 14]. At present, the National Center for Biotechnology Information (NCBI) database only includes the chromosome-level genome assembly of horse and donkey in perissodactyls, whereas DNA Zoo has chromosome-level genomes of plains zebra, white rhinoceros and Malayan tapir. Therefore, we will focus on some chromosomes of perissodactyls, especially the horse and the donkey.

The domestic donkey, *Equus asinus asinus* (EAS, $2n = 62$, NCBI Taxonomy ID: 83772), belongs to the Perissodactyla order, Equidae family, and *Equus* genus and descends from the African wild ass [15–17]. The order Perissodactyla includes three extant families: Tapiridae, Rhinocerotidae, and Equidae. The extant equid family is comprised of a single genus, *Equus*, which includes zebras, African and Asian asses, as well as horses. African

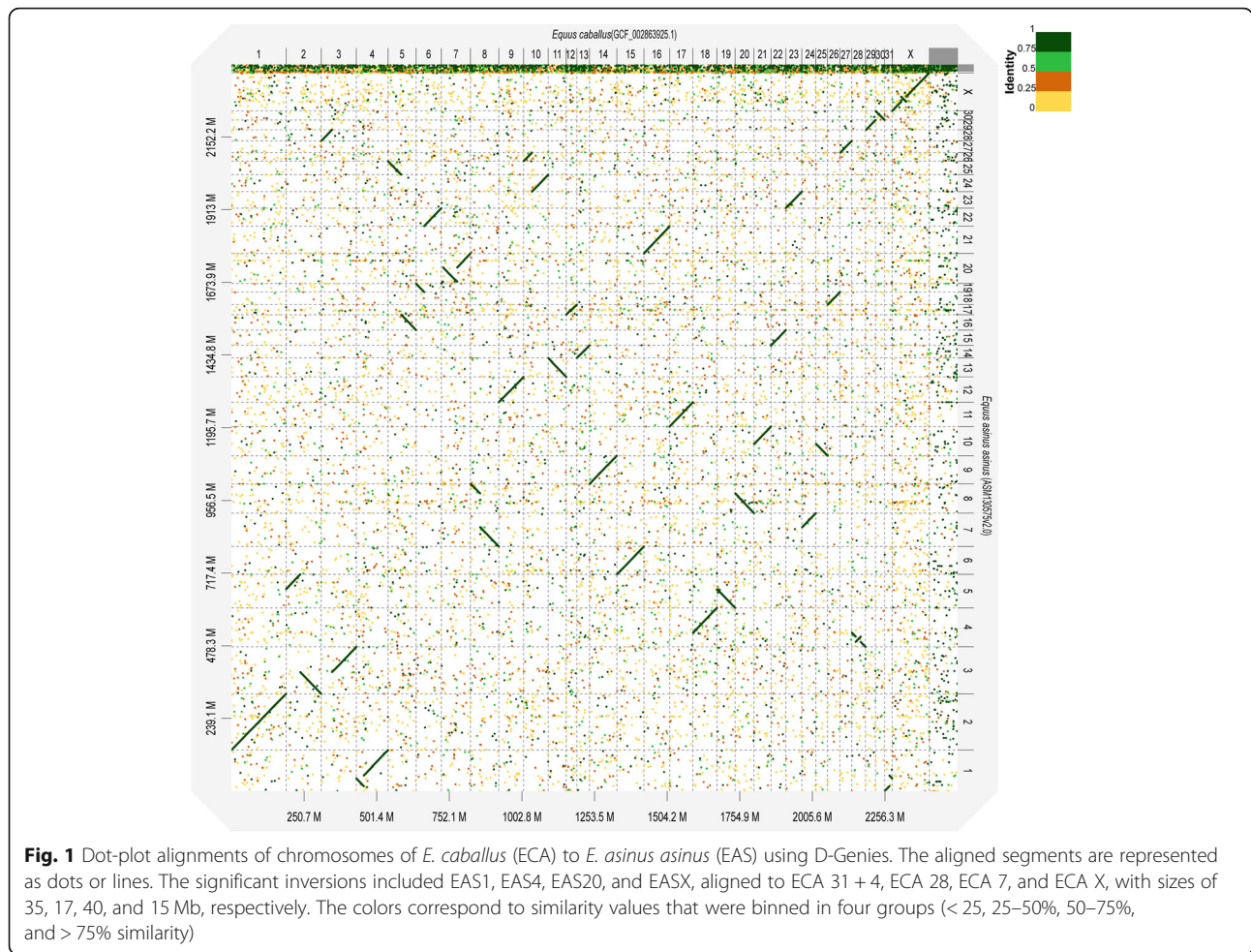
and Asian asses are sister groups, and asses and zebras are also known as non-caballine equids [18]. The early interest in donkeys was primarily triggered by their cross-breeding with horses (*Equus caballus*, ECA), mules or hinnies, with heterosis and fertility in a few cases [19–25]. However, the exploration of the above-listed issues must be based on a deep understanding of the parents. Unlike the large number of studies available for horses, our understanding of donkeys is still relatively poor. Although the NCBI database currently contains chromosome-length donkey genome, the synteny relationship in several places is inconsistent with the results of previous FISH-based methods [6, 13, 14, 26]. For example, it is mentioned in its supplementary data 1 that the donkey chromosome 2 assembly has a homologous relationship with the horse chromosome 1 and 28, which is obviously inconsistent with the results of previous studies on the independent correspondence between donkey chromosome 2 and horse chromosome 1 [6, 13, 14, 26]. As a result, we decided to use the donkey assembly provided by DNA Zoo.

DNA Zoo (<https://www.dnazoo.org/>), which is a global organization, dedicated to animal conservation, offers more than 150 chromosome-length genome assemblies of plants and animals [27, 28]. Here, we selected the donkey genome to prove the continuity and integrity of the assembly provided by this website. In addition, we identified the cytogenetic nomenclature, corrected the direction, and analyzed the genome-wide chromosomal rearrangements with reference to the results of the collinearity analysis performed in this study and previously published results. Finally, the donkey chromosome 1 and horse chromosome 5 were used as examples to explore their evolution in perissodactyls.

Results

Alignment to the horse genome and chromosome orientation

The present study identified an overall strong collinearity between the donkey and horse genomes (Figs. 1 and S1). Because the horse genome was constructed according to genetic physical mapping [11], we identified the cytogenetic nomenclature and adjusted the direction of the donkey chromosome-length sequence according to the subchromosomal painting results between donkey and horse [13, 29, 30]. Our findings showed the existence of 31 chromosome sequences numbered according to a cytogenetic standard, while 13 chromosome-length sequence directions were adjusted (including one scaffold internal adjustment). Finally, the modified donkey genome (ASM130575v2) was obtained and exhibited a scaffold N50 length of 95.5 Mb (Table 1). The comparison genome point diagram between the donkey and horse genomes is presented in Fig. 1. Compared with the



horse genome, we observed a significant number of inverted structures in chromosomes 1, 4, 20, and X of the donkey, with sizes of 35, 17, 40, and 15 Mb, respectively. The inverted structure located inside the X chromosome was the first validation at the sequence level.

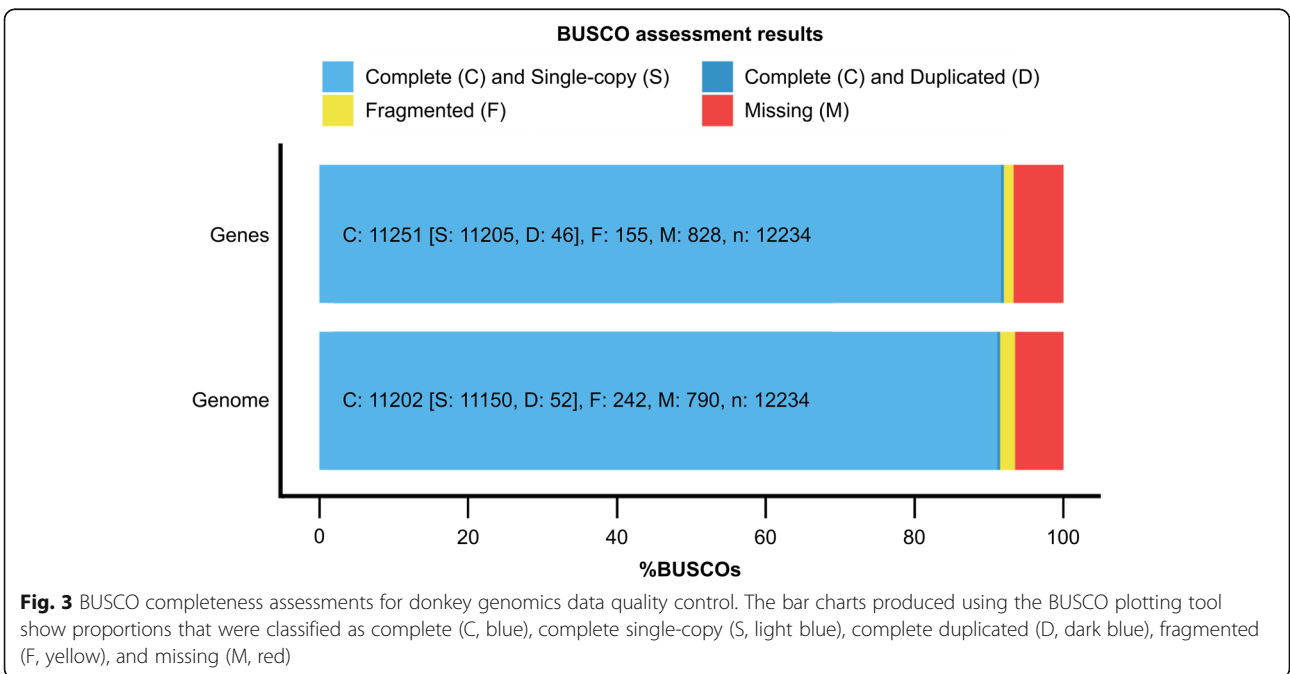
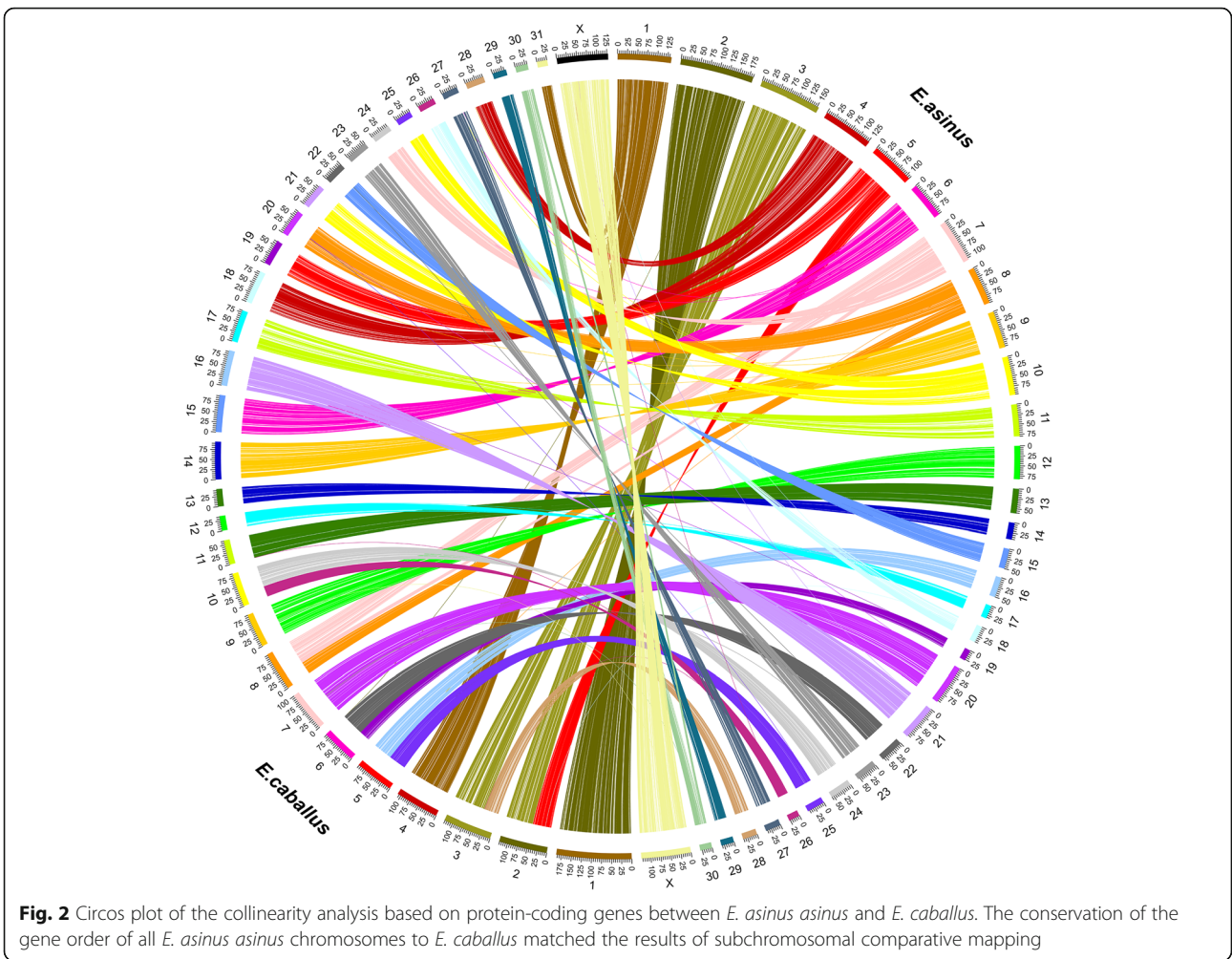
Gene annotation

We predicted a total of 18,937 protein-coding genes through homology-based approaches and then used the

annotations available for the horse to identify 16,691 donkey orthologs. A protein-based collinearity analysis showed the highly conserved order of genes for all chromosomes, as shown in Fig. 2. A Benchmarking Universal Single-Copy Ortholog (BUSCO) gene integrity analysis showed that among the lineage-specific profile (laurasiatheria_odb10.2019-11-20) which contains 12,234 markers, there were 11,205 single-copy orthologs and 46 duplicated orthologs, with complete genes accounting

Table 1 Quality metrics for the present assembly when compared with previous donkey genome assemblies

	This study (improved DNA Zoo, ASM130575v2)	Huang et al. [31] (ASM130575v1)	Renaud et al. [12]
N50 contigs	66.7 kb	66.7 kb	140.3 kb
N50 scaffolds	95.5 Mb	3.8 Mb	15.4 Mb
Total bases	2.391Gb	2.391 Gb	2.320 Gb
Largest scaffold	185.41 Mb	17.06 Mb	84.20 Mb
Unresolved bases per 100 kb	1396.07	1384.93	1121.61
Total number of predicted protein-coding genes	18,937	23,850	18,984



for 91.6% of these entities, which was indicative of good genomic integrity at the gene level (Fig. 3).

Collinearity analysis of EAS1 and ECA5 in perissodactyls and outgroups

The horse genome has been used as a reference for describing the structural variation in all other species. The results of collinearity analysis showed that the association between the orthologous regions of ECA31 and ECA4 found in EAS and plains zebra (*Equus burchellii*, EBU) did not exist in the white rhinoceros (*Ceratotherium simum*, CSI), Malayan tapir (*Tapirus indicus*, TIN), blue whale (*Balaenoptera musculus*, BMU), and humans (*Homo sapiens*, HSA) (Figs. 4 and 5). An inversion and a shift were noted in EAS1, orthologous to ECA31 (Figs. 4 and 5). Chromosome translocation refers to the transfer of a portion of one chromosome to another position on the same chromosome or a nonhomologous chromosome. When a chromosome translocation occurs involving a single chromosome, it is referred to as a shift. The similar inversion breakpoint in EAS1, orthologous to ECA31, did not exist in the orthologous regions of

EBU8, TIN20, BMU21, and HAS6 but was present in CSI37 (Figs. 4 and 5) [6, 13, 32]. We also detected an inversion and a shift in EAS1, CSI2, BMU19, and HAS7, orthologous to ECA4 (Figs. 4 and 5). The inversion breakpoint in EAS1, orthologous to ECA4, also occurred in BMU19, TIN5, and CSI2; however, because of the limited resolution of dot plots, it was not possible to determine the consistency of the inversion breakpoints or their evolution relationship. A collinearity analysis showed that the association between the short and long arms of ECA5, with the exception of EAS and EBU, was relatively conserved in perissodactyls; moreover, this conservation was also observed in humans among the order Primates and blue whales among the order Artiodactyla (Figs. 4 and 5). A 4.8 Mb inverted structure was first discovered in EAS25 and EBU11 (Figs. 1, 4 and 5).

We further used Mauve v2.4.0 for a more detailed analysis in ECA31 and 4 orthologous regions of perissodactyls. The results showed that the inverted breakpoints of EAS 1 and CSI37, orthologous to ECA31, were nearly located at the same position (Figures S2 and 5). Furthermore, the structure of the ECA4 orthologous regions

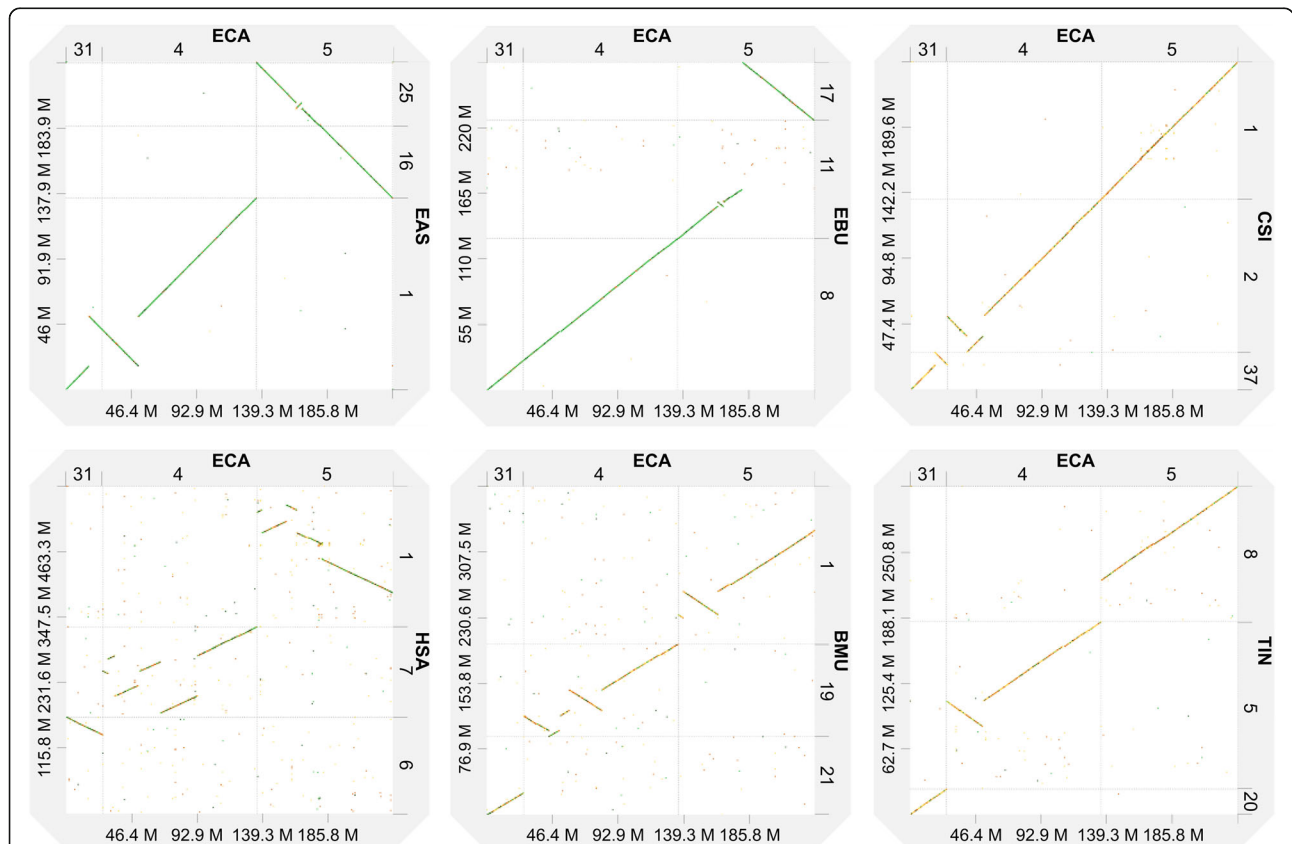
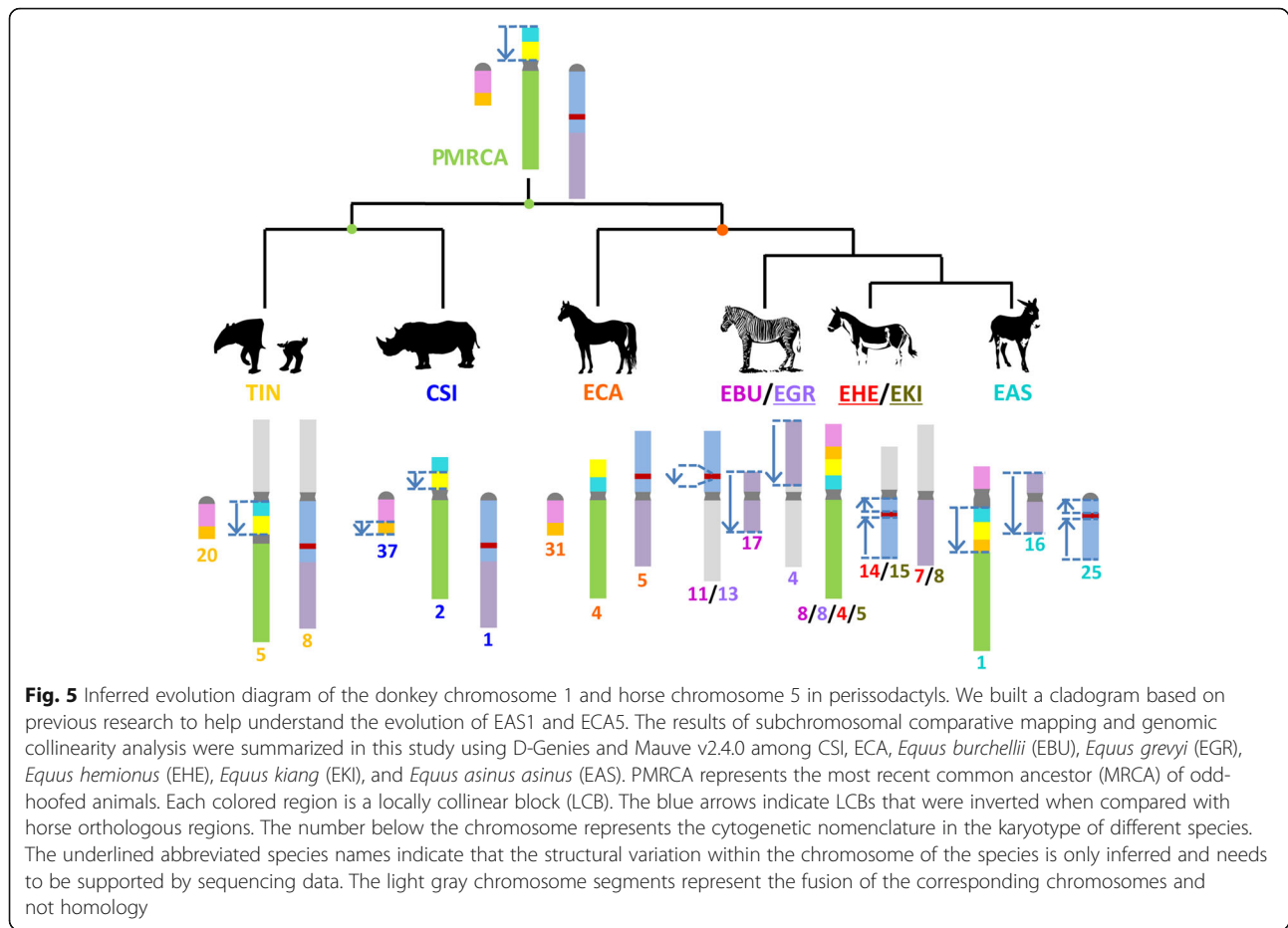


Fig. 4 Dot-plot alignments of EAS1 and ECA5 and their orthologs. Animals involved in comparative genomics include zebra (*Equus burchellii*, EBU), white rhinoceros (*Ceratotherium simum*, CSI), Malayan tapir (*Tapirus indicus*, TIN), blue whale (*Balaenoptera musculus*, BMU), and humans (*Homo sapiens*, HSA). Because the orientations of the CSI, TIN, and BMU chromosome-length scaffolds are unknown a priori, they were oriented using the strand that minimized the number and size of inversions with respect to the horse chromosomes



was different among the various perissodactyl species. In the Malayan tapir, the ECA4p orthologous region is inverted; in the white rhinoceros, the region is inverted and shifted.

Discussion

DNA zoo remains unexplored

Using the HiC technology [27, 28], DNA Zoo based on NCBI draft data generated a new genome assembly for the domestic donkey [31]. Its scaffold N50 length reached 95.5 Mb, six times larger than the best previously published donkey assembly [12]. There were 18,937 annotated protein-encoding genes, and the BUSCO gene integrity evaluation reached 91.6%, which was close to Renaud's genome assembly (18984) [12]. DNA Zoo offers chromosome-length genome assemblies of over 150 plants and animals. Considering the donkey genome alone as an example, the continuity and integrity of the assembly are relatively good and there is a good scope for exploration.

Alignment to the horse genome

We found an overall strong collinearity between the donkey and horse genomes, a result that was

consistent with those of a previous comparative genomic analysis performed using FISH [13, 14]. When compared with the ECA genome, we observed several distinct structural variations in chromosomes 1, 4, 20, and X of the donkey. Among them, EAS1, 20, and X were consistent with the results of a previous karyotype analysis and FISH [13, 33], whereas 1, 4, and 20 were consistent with the previous results of a subchromosomal genome assembly analysis [12]. The inverted structure detected within the X chromosome was the first validation at the sequence level. When compared with ECA5, a relatively small inverted structure (4.8 Mb) was found in EAS25 and EBU11. Such a small structural variation is difficult to screen using conventional molecular cytogenetic methods, such as karyotype analysis and FISH. Genetic or physical maps are relatively expensive and difficult to obtain. Therefore, based on the corresponding relationship between chromosome painting and gene mapping, we can refer to the genome assembly corrected by physical maps, such as the horse genome assembly, and orient chromosome-length sequences, eventually obtaining a chromosome-oriented assembly.

Evolution of EAS1 in perissodactyls

Together with the previously published results of comparative chromosome maps in the perissodactyl species [6, 13], our results helped in reconstructing the evolution of EAS1 orthologs in perissodactyl main lineages, especially the evolution of the internal inverted structure (Fig. 5). A collinearity analysis showed that EAS1 was orthologous to ECA31 and ECA4, which was consistent with the results of previous research at the molecular cytological level. Moreover, this tandem fusion occurs only in non-caballine equids [6, 13]. The structure of ECA4 orthologous regions was different among the various perissodactyl species. Based on the phylogenetic relationship [34, 35] and parsimony principles, we speculated on the evolution of EAS1 orthologous regions in different perissodactyl lineages. In the perissodactyl most recent common ancestor (MRCA), the orthologous regions of ECA31 and ECA4 correspond to independent and complete chromosomes, respectively. Besides, this independence is consistent in all the existing perissodactyl species, except the non-caballine equids. When compared with the MRCA of the tapir family, in Malayan tapirs, the ECA4 orthologous region was fused with the ECA8q orthologous region. When compared with the MRCA of Rhinocerotidae, in white rhinoceros, the corresponding orthologous blocks to ECA31 and ECA4 were inverted once in the distal end of CSI37 and the short arm of CSI2, and the position of the inverted breakpoint of CSI37 was the same as the orthologous region of EAS1, which may imply the existence of structure variation hotspots in the relevant areas. When compared with the perissodactyl MRCA, in the *Equus* MRCA, the orthologous region of ECA4p was inverted, similar to that observed in the extant horse. In the non-caballine MRCA, ECA31 and ECA4 were fused in tandem, forming structures similar to those found in EBU8. Therefore, in the process of speciation of the African wild ass, the orthologous region of distal block of ECA31 and the ECA4p block as a whole were inverted (Figs. 4, 5 and S2). We would like to emphasize that, because of the poor resolution of the genome assembly in the centromeric regions, none of the abovementioned or following analyses involved relevant regions. All analyses were based on chromosome painting and genome assemblies of individuals, rather than on groups, and may lack generality. The results of collinearity analysis based on genome assembly may not be completely correct unless it is a conservative variation that appears multiple times; therefore, FISH experiments are needed to verify their credibility [12, 26].

The direction of ECA5 evolution-fusion or fission

The long arm and the short arm of ECA5 are orthologous to chromosomes EPR23 (*Equus przewalskii*, EPR)

and EPR24, respectively [6, 36, 37]. Myka and Ahrens et al. [36, 38] used karyotype analysis and FISH to establish that the karyotype difference between ECA and EPR was only a Robertsonian translocation. However, when compared with ECA5, whether the separation of EPR23 and EPR24 was an ancestral structure or a derived structure has not been determined. Trifonov and Piras et al. [6, 37] used chromosome painting and multiple rounds of probe analysis to reveal that the association between ECA5p and ECA5q is relatively conserved in rhinoceros and tapirs. However, their research could only prove the homology of the whole chromosome (arms) or the conservative relationship of the order of some detection points. The work could not address the collinearity of the whole chromosome and the subtle structural variation. A collinearity analysis based on a chromosome-length genome assembly showed that the association between the short arm and long arm of ECA5, except for Asian and African asses, was relatively conserved in perissodactyls (Figs. 4 and 5). Furthermore, this conservation was observed in humans among Primates and blue whales among Artiodactyla (Fig. 4). Based on the phylogenetic relationship and parsimony principles [34, 35], we inferred that the association between the short arm and long arm of ECA5 was an ancestral characteristic of perissodactyl species and that EPR23 and EPR24 were generated via a Robertsonian fission event accompanied by the speciation of *Equus przewalskii*. In addition, an inverted structure with a length of 4.8 Mb was detected in the orthologous region of ECA5 in EAS25 and EBU11; this may be a characteristic chromosomal structural variation that happened during the occurrence of non-caballine equids, which is similar to the abovementioned association of ECA31 and ECA4 (Figs. 4 and 5).

Conclusions

Here we used the donkey genome as an example to prove the availability of DNA Zoo chromosome-level assembly data through the assessment of continuity and gene integrity and the collinearity analysis between horse and donkey genomes. By correcting the direction of chromosome-level sequence, we obtained a modified donkey assembly consistent with the previous results of FISH and sequence assembly.

When compared with the ECA genome, we observed several distinct inversions in donkey chromosomes 1, 4, 20, and X, which were also consistent with previous research results. Among them, to our knowledge, the inverted structure within the X chromosome was the first validation at the sequence level.

In addition, we reported the evolutionary details of EAS1 and ECA5 in perissodactyls using the chromosome-level genome assembly data provided by DNA Zoo and NCBI databases. ECA31 and 4 are the

orthologous chromosome of EAS1. In the perissodactyl MRCA, ECA31 and 4 are independent. This independence also exists in Tapiridae and Rhinocerotidae. We speculate that ECA4p was inverted in the Equus MRCA, and then the tandem fusion of ECA31 and 4 occurred in non-caballine equids. The tandem fusion of ECA31 and 4 is a possible feature of MRCA of non-caballine equids at the chromosome level.

Moreover, the short arm of ECA5 is orthologous to EAS25, and the long arm is orthologous to EAS16. The association between the short and long arms of ECA5 is found in both Tapiridae and Rhinocerotidae. We speculate this association to be present in MRCA of perissodactyls and equines too. In non-caballine equids, the long and short arms of ECA5 are broken to form independent chromosomes. This independence was retained in African asses, while different forms of chromosomal fusion occurred in zebras and Asian asses.

We discovered a 4.8 Mb inverted structure for the first time within the EAS25 and EBU11 compared to ECA. We speculate that the inverted structure is a common feature of non-caballine equids, proving the correctness of the existing equine phylogeny from the chromosome level [34, 35]; this also shows that the detailed analysis of sequence-based karyotype evolution has certain significance in predicting phylogenetic relationships.

With the lowering of the sequencing costs, an increasing number of chromosome-level genomes will be published. Combined with chromosome painting and gene mapping, the comparative genome based on chromosome-length genome assembly is bound to yield more interesting findings in the study of chromosome evolution.

Methods

Sampling of species genome assemblies

In the NCBI Genome database, among equids, the horse genome assembly alone has currently reached chromosome level. We collected the chromosome-length genome assemblies of plains zebra, white rhinoceros (*Ceratotherium simum*, CSI), Malayan tapir (*Tapirus indicus*, TIN), blue whale (*Balaenoptera musculus*, BMU), and humans (*Homo sapiens*, HSA) from the NCBI and DNA Zoo databases (Table S1) [27].

Chromosome orientation

The comparative genome point map between the horse and donkey genomes was generated using D-Genies (default parameters) [39]. Subsequently, BEDTools was used to modify the sequence direction of chromosomes according to the horse chromosome sequence direction and chromosome painting results [40].

Genome quality metrics

Various quality metrics were computed to evaluate the modified genome continuity using QUAST V5.0.2 (default parameters), including N50 and other technical metrics [41]. Based on the evolutionarily informed expectations of the gene content of near-universal single-copy orthologs, the BUSCO v4.0.6 metric is complementary to technical metrics, such as N50. The laurasiatheria_odb10 lineage-specific profile, which contains 12,234 BUSCO markers, was tested against assemblies of the donkey using the option “-m geno” [42].

Repeat masking for gene annotation

Before gene annotation, repeat masking was performed. RepeatMasker v4.0.9 was used to identify the transposable elements and tandem repeats that matched the entries in the Dfam_3.1 and RepBase-20,181,026 repeat library and were masked out in a soft-masked manner using the option “-parallel 20 -species equus -xsmall” [43].

Gene annotation

The homologous annotation approach based on the comparative genome was used to annotate the modified donkey genome. Cactus v1.1.1 [44] was used to generate the comparison genome file, CESAR v2.0 [45] was used as the annotation software, and the NCBI database annotation file “GCF_001305755.1_ASM130575v1_genomic.gff.gz” was used as the reference annotation. The guide tree was “(ECA:0.0033,(ASM130575v1:0.00023,ASM130575v2:0.00023):0.00303).”

The resulting gene set was used to identify the corresponding homologs in the horse genome. Orthologs in the horse genome were obtained using OrthoFinder v2.4.0 [46] with default parameters. The parsing of the output obtained in OrthoFinder was performed in-house using custom scripts. The software instructions suggested the presence of sufficient species sampling for inferring a good phylogenetic tree. Accordingly, six species with relatively complete genome annotations were selected to identify and analyze of the orthologs. Exception for ECA and EAS, the other species were used as outgroups (cattle, pigs, mice, and humans), as detailed in Table S2.

Circos v0.69–8 was used to visualize the collinearity relationships of orthologs between ECA and EAS.

Comparative genomic analysis

We used D-Genies (default parameters) to generate Dot-plot alignments between the horse and donkey genomes [39]. In addition, Mauve V2.4.0 (default parameters) was used for the visualization and detailed analysis of the study results [47].

Abbreviations

BMU: *Balaenoptera musculus*; BUSCO: Benchmarking Universal Single-Copy Ortholog; CSI: *Ceratotherium simum*; EAS: *Equus asinus asinus*; EBU: *Equus burchellii*; ECA: *Equus caballus*; EGR: *Equus grevyi*; EHE: *Equus hemionus*; EKl: *Equus kiang*; EPR: *Equus przewalskii*; FISH: Fluorescence in situ hybridization; HAS: *Homo sapiens*; LCB: Locally collinear block; MRCA: Most recent common ancestor; TIN: *Tapirus indicus*

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07984-6>.

Additional file 1.

Acknowledgements

We thank Professor Ning Zemin for the information on chromosome-level genome assembly in the DNA Zoo website. We also express our gratitude to Professor Yang Fengtang for his advice on chromosome evolution analysis. We are indebted to Professor Lu Ying for his advice on sequence analysis.

Authors' contributions

Conceptualization: Xihe Li. Data curation: Yunxia Li, Jun Li, Jinfeng Wang. Formal analysis: Shaohua Li, Gaoping Zhao, Jun Li, Jinfeng Wang, Xihe Li. Funding acquisition: Xihe Li. Investigation: Shaohua Li, Xihe Li. Methodology: Yunxia Li, Jinfeng Wang. Project administration: Xihe Li. Resources: Xihe Li. Software: Shaohua Li. Supervision: NA. Validation: Shaohua Li, Gaoping Zhao. Visualization: Shaohua Li. Roles/Writing - original draft: Shaohua Li, Hongmei Han, Guifang Cao. Writing - review & editing: Shaohua Li, Hongmei Han, Guifang Cao. The authors read and approved the final manuscript.

Funding

This research was funded by The Project of Inner Mongolia Autonomous Region Science and Technology Plan of China (Grant No.2020ZD0007).

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the website (DOI: <https://doi.org/10.6084/m9.figshare.14387126>).

Declarations

Ethics approval and consent to participate

Not applicable (the data used in this study is public).

Consent for publication

Not applicable.

Competing interests

All authors have no conflict of interest to declare.

Author details

¹Research Center for Animal Genetic Resources of Mongolia Plateau, College of Life Sciences, Inner Mongolia University, Hohhot 010070, China. ²College of Basic Medicine, Inner Mongolia Medical University, Hohhot 010110, China. ³Inner Mongolia Saikexing Institute of Breeding and Reproductive Biotechnology in Domestic Animal, Hohhot 011517, China. ⁴Department of Physical Education, Hohhot Minzu College, Hohhot 010051, China. ⁵College of Veterinary Science, Inner Mongolia Agricultural University, Hohhot 010018, China.

Received: 15 December 2020 Accepted: 6 September 2021

Published online: 15 September 2021

References

- Murphy WJ, Stanyon R, O'Brien SJ. Evolution of mammalian genome organization inferred from comparative gene mapping. *Genome Biol.* 2001; 2:reviews0005.1. <https://doi.org/10.1186/gb-2001-2-6-reviews0005>.
- Ryder OA, Epel NC, Benirschke K. Chromosome banding studies of the Equidae. *Cytogenet Cell Genet.* 1978;20:332–50. <https://doi.org/10.1159/000130862>.
- Yunis JJ, Sawyer JR, Dunham K. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science.* 1980;208:1145–8. <https://doi.org/10.1126/science.7375922>.
- Wienberg J, Jauch A, Stanyon R, Cremer T. Molecular cytogenetics of primates by chromosomal in situ suppression hybridization. *Genomics.* 1990;8:347–50. [https://doi.org/10.1016/0888-7543\(90\)90292-3](https://doi.org/10.1016/0888-7543(90)90292-3).
- Kearney L. Molecular cytogenetics. *Best Pract Res Clin Haematol.* 2001;14: 645–68. <https://doi.org/10.1053/beha.2001.0159>.
- Trifonov VA, Stanyon R, Nesterenko AI, Fu B, Perelman PL, O'Brien PCM, et al. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. *Chromosom Res.* 2008;16:89–107. <https://doi.org/10.1007/s10577-007-1201-7>.
- Bishop R. Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. *Biosci Horizons Int J Student Res.* 2010;3:85–95. <https://doi.org/10.1093/biohorizons/hzq009>.
- Savola S, Nardi F, Scotlandi K, Picci P, Knuutila S. Microdeletions in 9p21.3 induce false negative results in *CDKN2A* FISH analysis of Ewing sarcoma. *Cytogenet Genome Res.* 2007;119:21–6. <https://doi.org/10.1159/000109614>.
- Kim J, Farré M, Auvil L, Capitanu B, Larkin DM, Ma J, et al. Reconstruction and evolutionary history of eutherian chromosomes. *Proc Natl Acad Sci.* 2017;114:E5379–88. <https://doi.org/10.1073/pnas.1702012114>.
- Damas J, Kim J, Farré M, Griffin DK, Larkin DM. Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. *Genome Biol.* 2018;19:155. <https://doi.org/10.1186/s13059-018-1544-8>.
- Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol.* 2018;1:197. <https://doi.org/10.1038/s42003-018-0199-z>.
- Renaud G, Petersen B, Seguin-Orlando A, Bertelsen MF, Waller A, Newton R, et al. Improved de novo genomic assembly for the domestic donkey. *Sci Adv.* 2018;4:eaq0392. <https://doi.org/10.1126/sciadv.aq0392>.
- Musilova P, Kubickova S, Vahala J, Rubes J. Subchromosomal karyotype evolution in Equidae. *Chromosom Res.* 2013;21:175–87. <https://doi.org/10.1007/s10577-013-9346-z>.
- Yang F, Fu B, O'Brien PCM, Nie W, Ryder OA, Ferguson-Smith MA. Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. *Chromosom Res.* 2004;12:65–76. <https://doi.org/10.1023/B:CHRO.000009298.02689.8a>.
- Beja-Pereira A, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, et al. African origins of the domestic donkey. *Science.* 2004;304:1781. <https://doi.org/10.1126/science.1096008>.
- Rossel S, Marshall F, Peters J, Pilgram T, Adams MD, O'Connor D. Domestication of the donkey: timing, processes, and indicators. *Proc Natl Acad Sci U S A.* 2008;105:3715–20. <https://doi.org/10.1073/pnas.0709692105>.
- Raudsepp T, Christensen K, Chowdhary BP. Cytogenetics of donkey chromosomes: nomenclature proposal based on GTG-banded chromosomes and depiction of NORs and telomeric sites. *Chromosom Res.* 2000;8:659–70. <https://doi.org/10.1023/A:1026707002538>.
- Moehlan PD, Kebede F, Yohannes H. *Equus africanus*. The IUCN red list of threatened species 2015: e.T7949A45170994; 2015. <https://doi.org/10.2305/IUCN.UK.2015-2.RLTS.T7949A45170994.en>.
- Xu W, Miller DC, Clark AG, Antczak DF. Random X inactivation in the mule and horse placenta. *Genome Res.* 2012;22:1855–63. <https://doi.org/10.1101/gr.138487.112>.
- Xu W, Miller DC, Harman R, Antczak DF, Clark AG. Paternally expressed genes predominate in the placenta. *Proc Natl Acad Sci U S A.* 2013;110: 10705–10. <https://doi.org/10.1073/pnas.1308998110>.
- Wang Y, Gao S, Zhao Y, Chen WH, Shao JJ, Wang NN, et al. Allele-specific expression and alternative splicing in horse x donkey and cattle x yak hybrids. *Zool Res.* 2019;40:293–305.
- Ryder OA, Chernick LG, Bowling AT, Benirschke K. Male mule foal qualifies as the offspring of a female mule and jack donkey. *J Hered.* 1985;76:379–81. <https://doi.org/10.1093/oxfordjournals/jhered.a110118>.
- Henry M, Gastal EL, Pinheiro LEL, Guimarmes SEF. Mating pattern and chromosome analysis of a mule and her offspring. *Biol Reprod.* 1995;52: 273–9. https://doi.org/10.1093/biolreprod/52.monograph_series1.273.

24. Zhao CJ, Qin YH, Lee XH, Wu C. Molecular and cytogenetic paternity testing of a male offspring of a hinny. *J Anim Breed Genet.* 2006;123:403–5. <https://doi.org/10.1111/j.1439-0388.2006.00615.x>.
25. Zong E, Fan G. The variety of sterility and gradual progression to fertility in hybrids of the horse and donkey. *Heredity.* 1989;62(Pt 3):393–406. <https://doi.org/10.1038/hdy.1989.54>.
26. Wang C, Li H, Guo Y, Huang J, Sun Y, Min J, et al. Donkey genomes provide new insights into domestication and selection for coat color. *Nat Commun.* 2020;11:6014. <https://doi.org/10.1038/s41467-020-19813-7>.
27. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. *Science.* 2017;356:92–5. <https://doi.org/10.1126/science.aal3327>.
28. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, et al. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv.* 2018:254797. <https://doi.org/10.1101/254797>.
29. Di Meo GP, Perucatti A, Peretti V, Incarnato D, Ciotola F, Liotta L, et al. The 450-band resolution G- and R-banded standard karyotype of the donkey (*Equus asinus*, 2n = 62). *Cytogenet Genome Res.* 2009;125:266–71. <https://doi.org/10.1159/000235932>.
30. Bowling AT, Breen M, Chowdhary BP, Hirota K, Lear T, Millon LV, et al. International system for cytogenetic nomenclature of the domestic horse: report of the third International Committee for the Standardization of the domestic horse karyotype, Davis, CA, USA, 1996. *Chromosom Res.* 1997;5:433–43. <https://doi.org/10.1023/A:1018408811881>.
31. Huang J, Zhao Y, Bai D, Shiraigol W, Li B, Yang L, et al. Donkey genome and insight into the imprinting of fast karyotype evolution. *Sci Rep.* 2015;5:14106. <https://doi.org/10.1038/srep14106>.
32. Kulemzina AI, Proskuryakova AA, Beklemisheva VR, Lemskaya NA, Perelman PL, Graphodatsky AS. Comparative Chromosome Map and Heterochromatin Features of the Gray Whale Karyotype (Cetacea). *Cytogenet Genome Res.* 2016;148:25–34. <https://doi.org/10.1159/000445459>.
33. Raudsepp T, Lear TL, Chowdhary BP. Comparative mapping in equids: the asine X chromosome is rearranged compared to horse and Hartmann's mountain zebra. *Cytogenet Genome Res.* 2002;96:206–9. <https://doi.org/10.1159/000063050>.
34. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A.* 2014;111:18655–60. <https://doi.org/10.1073/pnas.1412627111>.
35. Welker F, Collins MJ, Thomas JA, Wadsley M, Brace S, Cappellini E, et al. Ancient proteins resolve the evolutionary history of Darwin's south American ungulates. *Nature.* 2015;522:81–4. <https://doi.org/10.1038/nature14249>.
36. Ahrens E, Stranzinger G. Comparative chromosomal studies of *E. caballus* (ECA) and *E. przewalskii* (EPR) in a female F1 hybrid. *J Anim Breed Genet.* 2005;122:97–102. <https://doi.org/10.1111/j.1439-0388.2005.00494.x>.
37. Piras FM, Nergadze SG, Poletto V, Cerutti F, Ryder OA, Leeb T, et al. Phylogeny of horse chromosome 5q in the genus *Equus* and centromere repositioning. *Cytogenet Genome Res.* 2009;126:165–72.
38. 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*). *Cytogenet Genome Res.* 2003;102:222–5.
39. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 2018;6:e4958. <https://doi.org/10.7717/peerj.4958>.
40. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
41. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–5. <https://doi.org/10.1093/bioinformatics/btt086>.
42. Seppely M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. *Gene prediction: methods and protocols.* New York: Springer New York; 2019. p. 227–45. https://doi.org/10.1007/978-1-4939-9173-0_14.
43. A.F.A, Smit R, Hubley P. Green RepeatMasker. <http://repeatmasker.org>. 2019.
44. Tsin YH. A simple 3-edge-connected component algorithm. *Theory Comput Syst.* 2007;40:125–42. <https://doi.org/10.1007/s00224-005-1269-4>.
45. Sharma V, Schwede P, Hiller M. CESAR 2.0 substantially improves speed and accuracy of comparative gene annotation. *Bioinformatics.* 2017;33:3985–7. <https://doi.org/10.1093/bioinformatics/btx527>.
46. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
47. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14:1394–403. <https://doi.org/10.1101/gr.2289704>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

