RESEARCH

Population genomics reveals moderate genetic differentiation between populations of endangered Forest Musk Deer located in Shaanxi and Sichuan

Gang Liu[1*], Bao-Feng Zhang[2], Jiang Chang[3], Xiao-Long Hu[4], Chao Li[1], Tin-Tao Xu[5], Shu-Qiang Liu[2] and De-Fu Hu[2*]

Abstract

Background: Many endangered species exist in small, genetically depauperate, or inbred populations, hence promoting genetic differentiation and reducing long-term population viability. Forest Musk Deer (*Moschus berezovskii*) has been subject to illegal hunting for hundreds of years due to the medical and commercial values of musk, resulting in a significant decline in population size. However, it is still unclear to what extent the genetic exchange and inbreeding levels are between geographically isolated populations. By using whole-genome data, we reconstructed the demographic history, evaluated genetic diversity, and characterized the population genetic structure of Forest Musk Deer from one wild population in Sichuan Province and two captive populations from two ex-situ centers in Shaanxi Province.

Results: SNP calling by GATK resulted in a total of 44,008,662 SNPs. Principal component analysis (PCA), phylogenetic tree (NJ tree), ancestral component analysis (ADMIXTURE) and the ABBA-BABA test separated Sichuan and Shaanxi Forest Musk Deer as two genetic clusters, but no obvious genetic differentiation was observed between the two captive populations. The average pairwise F_{ST} value between the populations in Sichuan and Shaanxi ranged from 0.05–0.07, suggesting a low to moderate genetic differentiation. The mean heterozygous SNPs rate was 0.14% (0.11%—0.15%) for Forest Musk Deer at the genomic scale, and varied significantly among three populations (Chi-square = 1.22, $p < 0.05$, Kruskal–Wallis Test), with the Sichuan population having the lowest (0.11%). The nucleotide diversity of three populations varied significantly ($p < 0.05$, Kruskal–Wallis Test), with the Sichuan population having the lowest genetic θ_π (1.69×10^{-3}).

Conclusions: Genetic diversity of Forest Musk Deer was moderate at the genomic scale compared with other endangered species. Genetic differentiation between populations in Sichuan and Shaanxi may not only result from historical biogeographical factors but also be associated with contemporary human disturbances. Our findings provide

*Correspondence: gangl@caf.ac.cn; hudf@bjfu.edu.cn

[1] Institute of Wetland Research, Chinese Academy of Forestry, Beijing Key Laboratory of Wetland Services and Restoration, Beijing 100091, China
[2] College of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100085, China
Full list of author information is available at the end of the article

Liu *et al. BMC Genomics*   (2022) 23:668

Page 2 of 11

scientific aid for the conservation and management of Forest Musk Deer. They can extend the proposed measures at the genomic level to apply to other musk deer species worldwide.

Keywords: Forest Musk Deer, Population decline, Population genomics, Genetic diversity, Genetic differentiation, Gene flow

Background

Many endangered species exist in small, genetically depauperate, or inbred populations [1], hence promoting genetic differentiation and reducing long-term population viability [2, 3]. Genetic diversity is important to the breeding and long term persistence of endangered species [4, 5]. In theory, a species with a declining population size is challenged with low levels of genetic diversity and limited gene flow [6–8], inbreeding depression [9] and even high risks of extinction [10]. This means it is necessary to conduct population genetic diversity, and structure studies on endangered species; however, such evaluations are unfortunately inadequate [11–13], and genetic consequences following the disruption of gene flow still remain unknown in many cases.

Genetic diversity available for a species is determined by the genetic background of the ancestor populations, which in turn is heavily impacted by the demographic history and contemporary human disturbance [14]. Small effective population sizes due to historical bottleneck events could expose endangered species to inbreeding and loss of genetic variation [15]. Historical climatological events and geophysical barriers can alter range shifts, determine dispersal patterns and restrict gene flow, leading to significant changes in population size and genetic diversity [16]. Contemporary human-driven habitat loss and fragmentation are among the greatest threats to population decline and isolation, leading to genetic diversity loss [17, 18].

All musk deer species (*Moschus spp.*) are assessed to be globally threatened due to poaching and habitat loss [19–21], with six being ranked as endangered and one as vulnerable (*M. moschiferus*) according to The IUCN Red List of Threatened Species v. 2015 [22]. The musk deer is characterized by the musk produced by the musk pod on the abdomen of adult males, and the musk is valuable for use in traditional Chinese medicine and perfumes [23]. Consequently, musk deer species have been subject to intensive hunting for hundreds of years, resulting in severe population declines worldwide. The Forest Musk Deer (*M. berezovskii*) was once widely distributed in China [24]. However, its wild populations have been reduced dramatically due to illegal hunting and deforestation since the 1950s. Captive breeding of Forest Musk Deer has been practiced for several decades, with the aim to provide sources for the reintroduction project and help to invert population declines and hunting pressure in the wild. The population growth of Forest Musk Deer was quite slow before 2000 [25], but the last two decades have seen a relatively rapid increasing, with more than 11,340 individuals in captivity in Shaanxi Province, China (https://china.huanqiu.com/article/9CaKrnJWMYT). During the breeding, Forest Musk Deer is challenged with disease severity and immunity reduction, which may be exacerbated by inbreeding and genetic diversity declines in Forest Musk Deer [26, 27].

The development of next-generation sequencing technologies brought a new discipline "conservation genomics" into population studies of endangered species [28, 29]. Since genome-wide genetic diversity can be estimated with neutral and non-neutral genomic markers [30–33], this advantage generally leads to an unbiased estimation of genetic diversity [34]. Inbreeding estimates by the genomic data are usually less downwardly biased compared to that from traditional pedigree methods [2, 35, 36]. Genetic structure can be revealed at a relatively fine scale, even when the sign of population genetic differentiation is subtle [37]. Genomic data are also useful for reconstructing the species' demographic history, which is essential to guide conservation activities, especially when the population is in decline [31, 38]. In aiding the project of captive breeding, as well as the reintroduction, accurate information derived from genomic data is crucial for selecting source populations to minimize inbreeding [39–41]. Previous studies mainly focused on comparing genetic diversity among captive populations of Forest Musk Deer, and depended on traditional molecular markers, including mitochondrial DNA [42, 43], microsatellites [44, 45], MHC [46, 47] and RAD sequencing [48]. The current distribution of Forest Musk Deer is limited mainly to Shaanxi and Sichuan provinces in China. however, it remains unclear to what extent the level of gene flow is between geographically isolated populations at the genomic scale. It is also unknown how demographic history and human disturbance impact the population structure.

In this study, whole genome sequencing (WGS) data were used to reconstruct the demographic history, evaluate genetic diversity, characterize genetic structure, and reveal inbreeding pattern of Forest Musk Deer in one wild population in Sichuan Province and two captive populations in Shaanxi Province, China.

Liu *et al. BMC Genomics*   (2022) 23:668

Page 3 of 11

The current captive individuals are descendants of wild individuals captured from the natural habitats in the Qinling Mountains in Shaanxi. Therefore, we assume that the two captive populations can represent the wild population in the Qinling Mountains in Shaanxi, or at least the imprints of demographic history on the genetic background remain. We tested the hypothesis that genetic differentiation exists between populations in Sichuan and Shaanxi, and inferred the pattern of genetic composition.
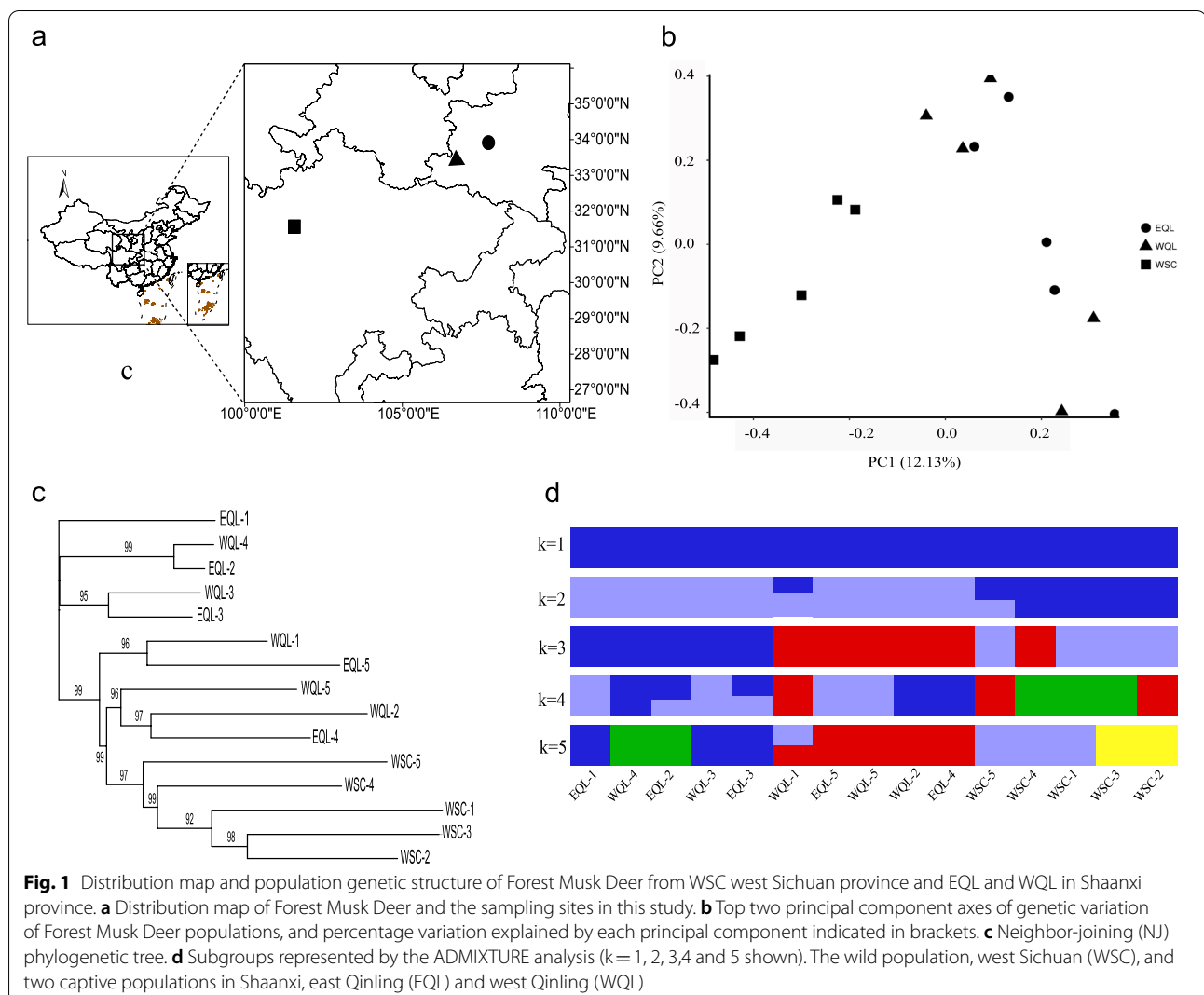
Results

Sequencing and mapping assessment
Error rate of sequencing data was lower than 0.04% for all 15 samples of Forest Musk Deer (Additional file 1: Table S1), and the average GC content was 43.32% (Additional file 1: Table S2). After mapping to the reference genome of Forest Musk Deer, it yielded an average sequencing depth of $25 \times$ (range: 16—30) and an average mapping percentage of 98.47% (range: 93.11%—99.55%) (Additional file 1: Table S3). The kinship coefficient was less than 0.177 for all individual pairs within each population (Additional file 1: Table S4), suggesting the genetic relatedness is not a major concern for the subsequent analysis. SNPs calling by GATK4 [49] resulted in a total of 44,008,662 SNPs after variant calling.

Population genetic structure
For the wild population, west Sichuan (WSC), and two captive populations in Shaanxi, east Qinling (EQL) and west Qinling (WQL) (Fig. 1a), the principal component analysis (PCA, Fig. 1b) was consistent with the

Fig. 1 Distribution map and population genetic structure of Forest Musk Deer from WSC west Sichuan province and EQL and WQL in Shaanxi province. **a** Distribution map of Forest Musk Deer and the sampling sites in this study. **b** Top two principal component axes of genetic variation of Forest Musk Deer populations, and percentage variation explained by each principal component indicated in brackets. **c** Neighbor-joining (NJ) phylogenetic tree. **d** Subgroups represented by the ADMIXTURE analysis (k = 1, 2, 3,4 and 5 shown). The wild population, west Sichuan (WSC), and two captive populations in Shaanxi, east Qinling (EQL) and west Qinling (WQL)

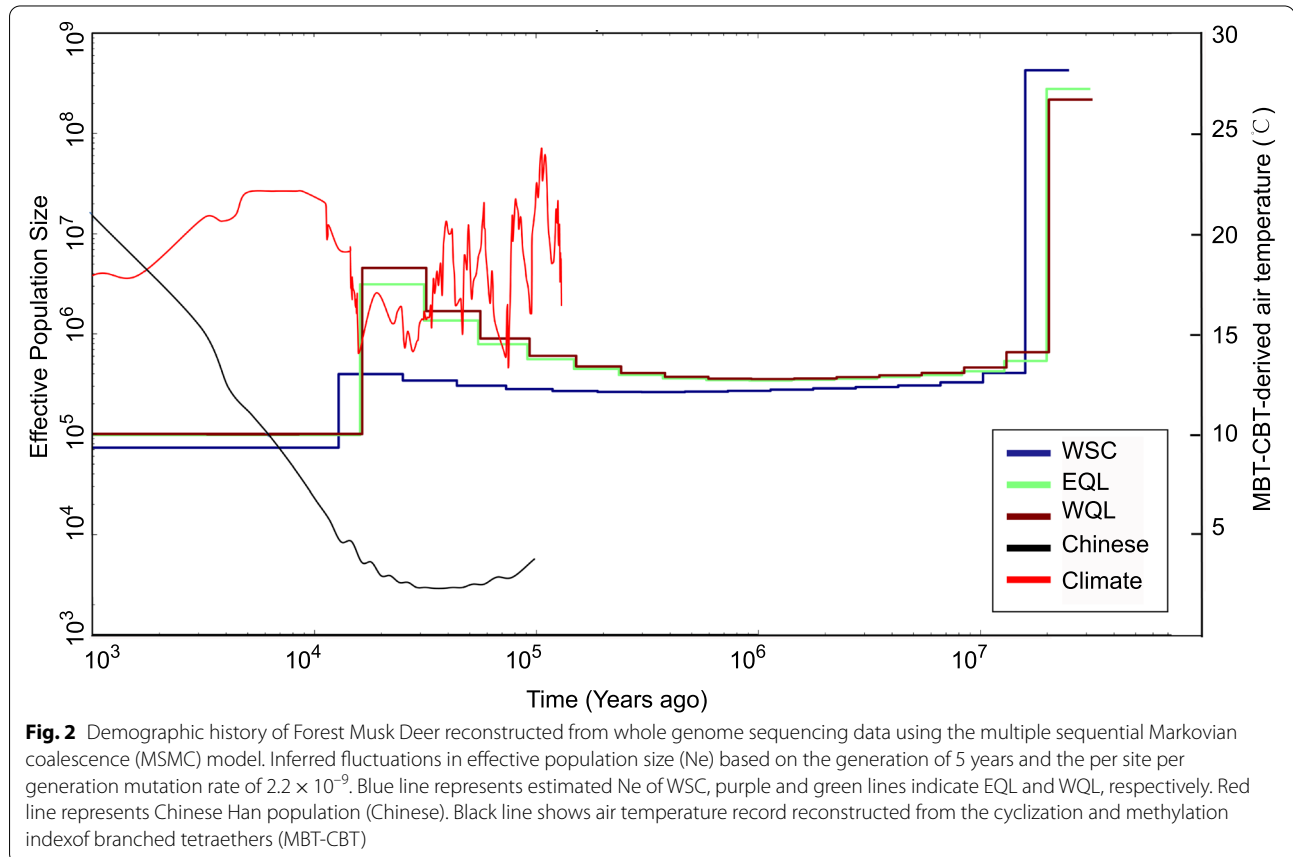Liu *et al. BMC Genomics*   (2022) 23:668

Page 4 of 11

results of the phylogeny relationships (Fig. 1c) and the admixture algorithm (Fig. 1d). The first principal component separated among populations, and the second one indicated some within population structure, which revealed that Sichuan and Shaanxi Forest Musk Deer could be discriminated by genomic data. The first two eigenvectors explained 12.13% genetic variation, suggesting that only a subset of genetic loci played a role driving the genetic differences between populations (Fig. 1b). The Neighbor-joining (NJ) phylogenetic tree based on pairwise SNP differences revealed separate genetic clusters between populations in Shaanxi and Sichuan, but no obvious genetic differentiation was observed between the two captive populations (EQL and WQL) (Figs. 1c). In contrast, the results from ADMIXTURE confidently discriminated between two genetically distinct populations under k = 2 (Fig. 1d), although the lowest CV-error was obtained with k = 1 (Additional file 2: Figure S1). The average pairwise F_{ST} value between WSC and EQL was 0.07, 0.02 between WQL and EQL and 0.05 between WSC and WQL population, suggesting a low to moderate genetic differentiation, which was more than that between WQL and EQL in Shaanxi.

Genetic diversity

By estimating the proportion of heterozygous SNPs per base pair, the mean heterozygous SNP rate was 0.14% (0.11%—0.15%) for Forest Musk Deer at the genomic scale, and varied significantly among three populations (Chi-square = 1.22, $p < 0.05$), with the wild WSC population having the lowest (0.11%). The nucleotide diversity of three populations varied significantly ($p < 0.05$), in contrast, WSC had the lowest genetic θ_π (1.69×10^{-3}) and θ_w (1.59×10^{-3}), while WQL had the highest θ_π (2.01×10^{-3}) and θ_w (2.00×10^{-3}) following by EQL (θ_π: 1.98×10^{-3} and θ_w: 1.94×10^{-3}). In terms of both heterozygous SNP rate and nucleotide diversity, it could be seen that the wild population (WSC) in Sichuan had lower genetic diversity than the two captive populations in Shaanxi (EQL and WQL).

Demographic history and gene flow

We reconstructed historical effective populations sizes (Ne) for all Forest Musk Deer populations using the MSMC [50]. All populations show a steady population increase followed by a dramatic expansion approximately 100,000 years ago, and then the population started to decline and showed a sharp

Fig. 2 Demographic history of Forest Musk Deer reconstructed from whole genome sequencing data using the multiple sequential Markovian coalescence (MSMC) model. Inferred fluctuations in effective population size (Ne) based on the generation of 5 years and the per site per generation mutation rate of 2.2×10^{-9}. Blue line represents estimated Ne of WSC, purple and green lines indicate EQL and WQL, respectively. Red line represents Chinese Han population (Chinese). Black line shows air temperature record reconstructed from the cyclization and methylation index of branched tetraethers (MBT-CBT)

population reduction around 10,000 years ago (Fig. 2). The effective population size of WSC was lower than both WQL and EQL, while WQL and EQL showed a similar pattern. The ABBA-BABA test [51] indicated more share-derived alleles between WQL and EQL ($D = 0.0033$, $Z = 8.257$, Fig. 3a and Additional file 1: Table. S5 and S6), which suggested that WSC was genetically far from both of WQL and EQL. The tree mix result (Fig. 3b) was consistent with the ABBA-BABA test.
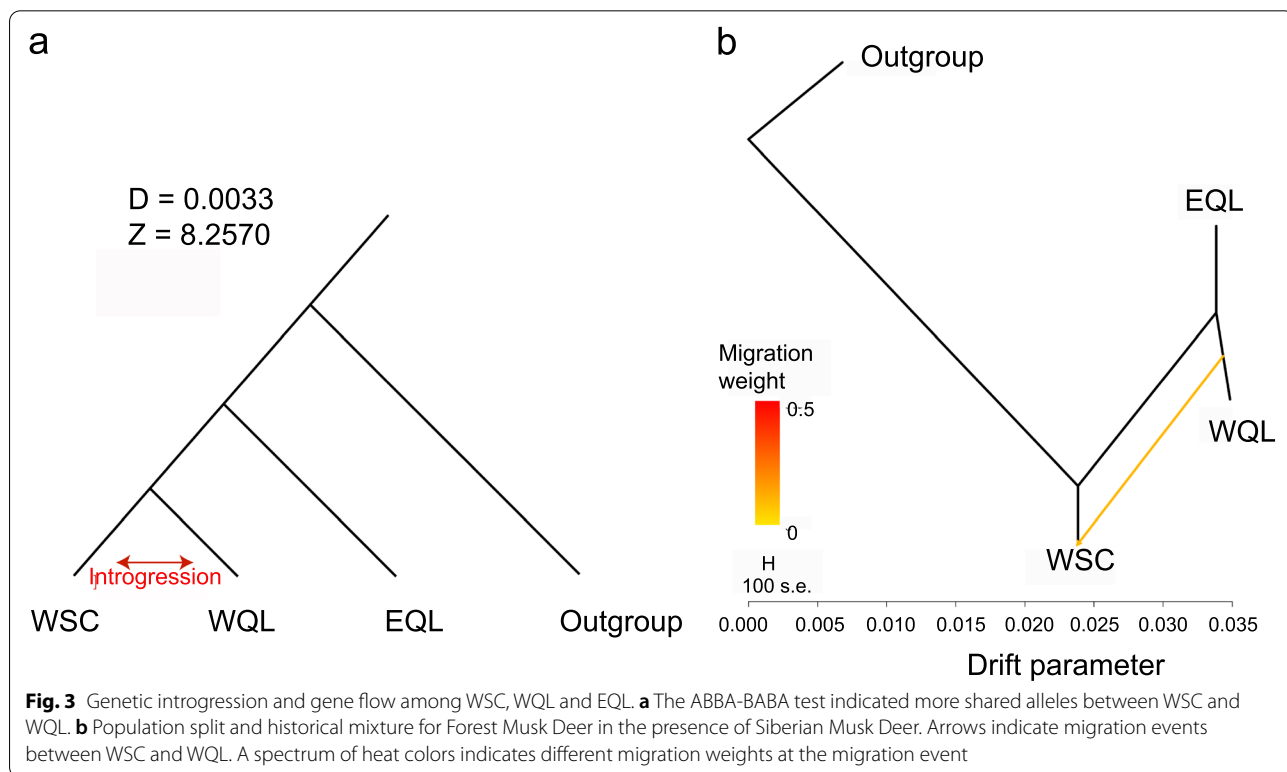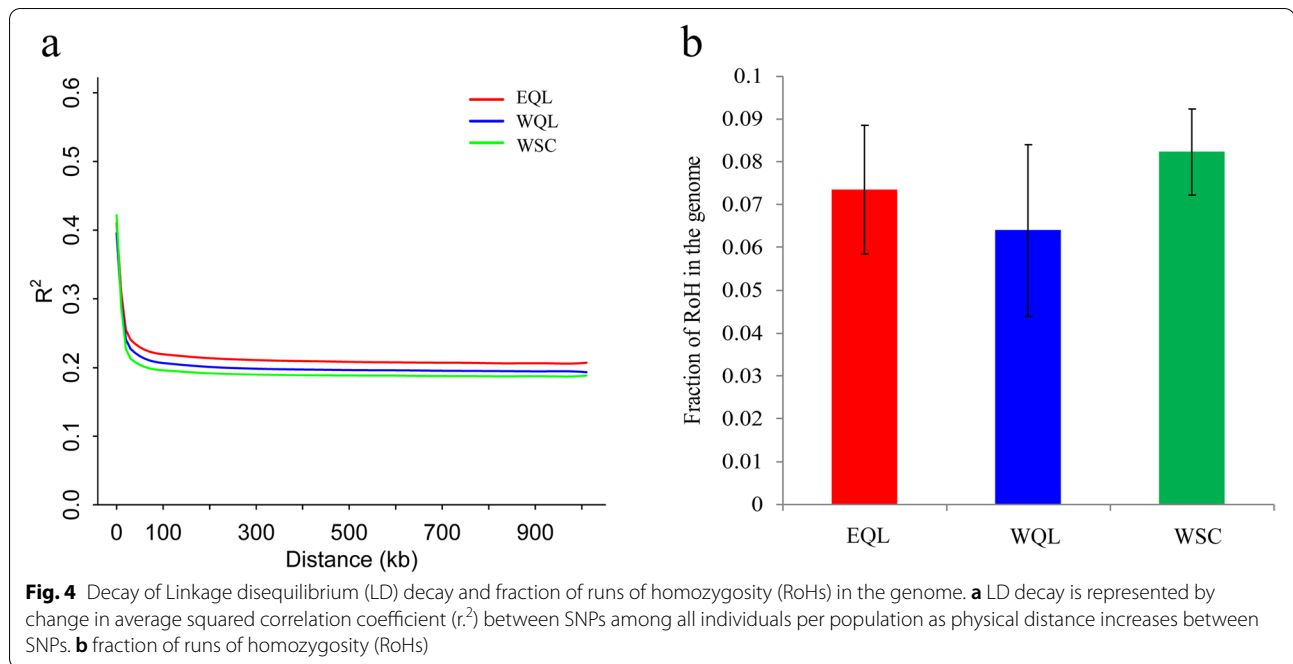
Inbreeding levels

Based on the linkage disequilibrium (LD) analysis, the extent of LD in both populations of WQL and EQL appeared consistently, but was quite different from the pattern in WSC. LD decayed to $r^2 < 0.2$ within 10 kb, but declined more slowly (Fig. 4a). Consistently with the LD results above, the two populations of WQL and EQL had lower numbers of RoHs above 200 kb than those of WSC (Fig. 4b), but no significant difference was detected (Chi-square $= 0.681$, $p > 0.05$).

Discussion

We found that the genetic diversity of the wild population in west Sichuan (WSC) was lower than that of both captive populations in Shaanxi (WQL and EQL), which was consistent with previous studies revealed by microsatellites [52], mitochondrial DNA [43], MHC [46], and RAD sequencing [48]. The highest level in Shaanxi is mainly due to abundant sources from different places, because the centers have a relatively short history and more widely distributed individuals are introduced into the breeding [52]. We compared the genome-wide heterozygosity of Forest Musk Deer with that of other mammal species (Additional file 2: Figure S2), and showed that it was higher than those in the Przewalski's Horse (0.039%) and Père David's Deer (0.054%), both of which went extinct in the wild in the early 1900s and were restored by limited founders in captivity [53]. However, the genome-wide heterozygosity of Forest Musk Deer was comparable to that of Giant panda, Chinese Pangolin and Sumatran Rhinoceros [54, 55] (Additional file 2: Figure S2). Meanwhile, the genome-wide heterozygosity of Forest Musk Deer was higher than the average (0.07%) for the mammals listed as EN in the IUCN Red List [22], suggesting that moderate genetic diversity for this species was observed. Such moderate levels of genetic diversity suggest that adaptive potential of this species still exists but is likely to decrease. The genomic consequences of long-term population declines include extreme reduction of genetic diversity and evidence of inbreeding depression [28]. However, dramatic short-term population declines need not necessarily result in major losses of genetic diversity [56].

Fig. 3 Genetic introgression and gene flow among WSC, WQL and EQL. **a** The ABBA-BABA test indicated more shared alleles between WSC and WQL. **b** Population split and historical mixture for Forest Musk Deer in the presence of Siberian Musk Deer. Arrows indicate migration events between WSC and WQL. A spectrum of heat colors indicates different migration weights at the migration event

Fig. 4 Decay of Linkage disequilibrium (LD) decay and fraction of runs of homozygosity (RoHs) in the genome. **a** LD decay is represented by change in average squared correlation coefficient (r^2) between SNPs among all individuals per population as physical distance increases between SNPs. **b** fraction of runs of homozygosity (RoHs)

The phylogeographic reconstruction, PCA and ADMIXTURE analyses provided evidence of moderate population genetic differentiation between populations of Forest Musk Deer located in Shaanxi and Sichuan, but weak genetic differentiation between two captive populations within Shaanxi. The distribution of wild Forest Musk Deer is divided into small populations by geographical barriers, such as Qinling Mountains and Minjiang River [57, 58], and this spatial separation may contribute to the genetic differentiation observed in Sichuan and Shaanxi. Past climate change is considered to be main drivers to shape the demographic history and genetic structure of terrestrial mammals [16]. For example, the geographical separation of the Qinling Mountains (Shaanxi) from the Minshan mountains (Sichuan) has caused morphological, behavioral and genetic distinctions of Giant Panda (*Ailuropoda melanoleuca*) [59, 60]. Cooling period during the Pleistocene may act on genetic differentiation of Forest Musk Deer similarly to Giant Panda by influencing gene flow between Sichuan and Shaanxi. A sudden rise in air temperature after the Last Glacial Maximum (LGM) about 12 kya might have change the forest habitats, and in turn might have driven the separation of Forest Musk Deer. In addition, the Sichuan Basin is not the refuge for Forest Musk Deer during the last glaciations, since Forest Musk Deer is a typical ruminant that prefers the high-altitude environment, instead of the warm and humid climate in the basin, which implies that the Sichuan Basin may also pose a geographical barrier to gene flow. Similar origin of

location and individual exchange between ex-situ centers within Shaanxi would help to explain the observed low genetic differentiation between the two captive populations (WQL and EQL).

It is important to note that most mammals contract during the most recent glacial period approximately 12.1 – 34.8 kya [59]. Forest Musk Deer population did not recover even after the temperature trajectory reversed, which suggested that, besides the biogeographical factors, human activities may have contributed to accelerating the population differentiation and decline. The distribution of wild Forest Musk Deer is characterized by scattered habitats being lost due to deforestation in the past and isolated by towns, roads, railways, and other infrastructure, which exacerbates the effects of habitat fragmentation and consequently restricts gene flow.

The demographic history based on genomic data further provided evidence that human disturbance was closely associated with the population declines of Forest Musk Deer. It could be inferred that the effective population size of Forest Musk Deer was impacted by human activities because the Chinese Han population continued to increase since 30 kya, whereas the effective population size experienced a steep decline in Forest Musk Deer. Forest Musk Deer, a species that adapts to high mountain forests and is characterized with shy and timid personality, is more vulnerable to the natural landscape changes and the destruction of hidden conditions [61]. Dramatic population decline due to human overexploitation, such as deforestation, poaching and harvest, can

Liu *et al. BMC Genomics*   (2022) 23:668

Page 7 of 11

influence reproduction and survival rates, and lead to extremely low population density (0.16–0.24 individuals per km^2 [62]).

Musk has been used for perfume and medicinal purposes for thousands of years Unfortunately for the endangered but economically important species, the species conservation and musk utilization are unbalanced. Current conservation efforts remain insufficient to offset the negative effects of population declines. The conservation priority is to enhance the patrol time to eliminate poaching [63, 64], which could eventually help restore the population and maintain the genetic diversity. In order to ensure population persistence, we propose additional conservation measures based on our findings. First, effective measures should be carried out to increase wild populations and restore suitable habitats, because targeted gene flow can allow purifying selection to function well and effectively to purge deleterious mutations. Targeted gene flow indicates translocating individuals with rare or favorable genes to areas. Second, the populations of Forest Musk Deer in Sichuan and Shaanxi are suggested to be considered as independent conservation management units. Third, artificial gene flow among different ex-situ centers should be prompted, and individuals with higher genetic diversity can be the candidate to be released into the wild.

The effects of recent severe demographic and genetic declines can be inferred by comparing the contemporary samples with the historical samples, or by running simulations under alternative historical scenarios based on the coalescent approach. This will be the research priority to evaluate the genetic consequences of wildlife commercial and medical exploitation within a comparative context. For captive populations, breeding management is an important factor in determining the genetic composition and population genetic structure, which means that pedigree records are strongly suggested to be done in order to avoid inbreeding in captivity. The proposed measures at the genomic level could be extended to apply to other musk deer species worldwide, since the knowledge from population genomics will contribute to the conservation of those endangered species.

Methods
DNA sample collection
Tissue or blood samples of 15 Forest Musk Deer individuals were collected from one wild population and two ex-situ populations. Specifically, five skin tissue samples were collected from the wild individuals captured in the west of Sichuan (WSC), five blood samples from the ex-situ center in Fengxian located in the west of Qinling Mountains (WQL) and five blood samples from the ex-situ center in Meixian located in the east of Qinling

Mountains (EQL) (Fig. 1a). The WQL was built in 1986, with a current population size of 258. The EQL was constructed in 2003 and has 196 individuals by the end of 2020. Both the WQL and EQL populations are composed of descendants of wild individuals from their natural habitats in the Qinling Mountains, and the number of founders is not known due to a lack of pedigree records. The exchange of captive-bred individuals occur less frequently among different ex-situ centers when captive-bred within Shaanxi Province. Blood samples were stored in vacuum blood collection tube (EDTA anticoagulation), and tissue samples were preserved using ethanol. All samples were stored at -80 ℃ (Additional file 1: Table S1).

Genome sequencing and SNP calling
Genomic DNA was extracted using QIAamp DNA Blood Mini Kit (Qiagen, Germany). The libraries were prepared as described in the commercial TruSeq kit (Illumina, the U.S.A). Whole-genome paired-end sequencing (PE150) was performed within the Illumina Novaseq 6000 platform using standard procedures (Novogen, China). The raw FASTQ data was de-multiplexed, and the adapters were trimmed using Trimmomatic v0.36.0 [64]. Clean sequencing reads were aligned against the reference genome of Forest Musk Deer downloaded from NCBI (BioSample ID: SAMN10822789, BioProject ID: PRJNA438286) using Burrows-Wheeler Alignment (BWA v0.7) tool [65]. Read alignments were sorted using SAMTools v1.11 [65], and duplicate reads were marked by Picard.

Variant calling was conducted using HaplotypeCaller and GenotypeGVCFs modules in GATK4 [49]. To ensure variant identification accuracy, GATK best practices including the density distribution of each parameter and the filtering criteria. SNPs were removed if any of the following parameter was met: missingness > 0.9, minor allele frequency < 0.05, QD < 2, MQ < 30, MQRankSum < −12.5, FS > 200, ReadPosRankSum < 20, QUAL < 30.0, AN < 40. The GVCF file from each sample was generated using HaplotypeCaller, and then GenotypeGVCFs was used to merge separate GVCF files with the aim of improving the sensitivity of variant detection. Before performing the subsequent population genomics analyses, genetic relatedness between individuals within each population was calculated using GCTA v1.94.0 [66]. Individual pairs were excluded from the subsequent analyses if their kinship coefficient was greater than 0.177, a threshold showing twins or first-degree relationships [67].

Genetic diversity estimation
In order to assess genetic diversity within and among Forest Musk Deer populations, the heterozygosity of each individual was calculated by dividing the total number

Liu *et al. BMC Genomics*   (2022) 23:668

Page 8 of 11

of heterozygous SNPs by genome size as previously described and averaged over each population [68]. Population genetics statistics θ_π was calculated using VCFtools v0.1.13 [69], and the Watterson's estimator θ_w was calculated using VariScan v2.0.3 [70].

Population structure analysis

The population structure analysis was performed using the commonly used methods [71], including phylogenetic tree (NJ tree), ancestral component analysis and principal component analysis (PCA). Based on the filtered SNPs, the GCTA program v1.94.0 [66] was applied to conduct PCA, by filtering the principal components through an algorithm of dimensionality reduction. The NJ tree was built using TreeBeST v1.9.2 (http://treesoft.sourceforge.net/) with the aim to visualize the genetic distance relationship between individuals, with a bootstrap value of 1,000. The standard VCF file was filtered and thinned according to user-specified command line options (–maf 0.05 –max-missing 0.9 –min-alleles 2 –max-alleles 2) by VCFtools v0.1.13 [69]. Additional filtering was handled for LD pruning (–indep-pairwise 50 5 0.2) using PLINK v.1.9 [72], and then population genetic structure was inferred using ADMIXTURE v1.3.0 [73]. The most likely value of K was identified from 20 independent runs for each value of K ranging from two to five, with 100 bootstraps. We used VCFtools v0.1.13 [69] to calculate pairwise Fst, a widely used genetic differentiation statistics, to examine the genetic differentiation between the three Forest Musk Deer populations (WSC, WQL, and EQL).

Demographic history reconstruction

Multiple sequential Markovian coalescence (MSMC) model can be used to reconstruct each population's demographic history based on genomic information [50]. MSMC is developed to overcome the limitation that the pairwise sequentially Markovian coalescent (PSMC) can only analyze one sample at a time [74]. In addition, MSMC can integrate and analyze the nearest common ancestor time between multiple allele sequences at the same time [75], thereby improving the accuracy and efficiency of effective population size (Ne) prediction. We performed demography inference using msmc2 v2.1.2 [75]. The parameters used were as follows: "-t 6 -p 1*2+15*1+1*2". A mutation rate (μ) of 2.2×10^{-9} mutations per site per year for Forest Musk Deer used in this study was according to the reference [17]. The generation time was estimated to be five years based on a report of mammal generation length [76].

Gene flow analysis

The Patterson's D-statistic, also called ABBA-BABA test, is the most widely used statistical approach that is applied to detect introgression across genomes between.

populations within a species or among species [51], and reveal the genomic footprint of post speciation introgression [77]. The appeal of D statistics is its simplicity, but the timing and direction of introgression can not be inferred [78]. We downloaded the whole genome re-sequencing data of Siberian Musk Deer (*M. moschiferus*) from NCBI (Project ID: PRJNA574937), and used it as the outgroup. We performed the test using angsd v0.935 [79], under the order set as WSC/WQL/EQL/outgroup or WSC/EQL/WQL/outgroup. The commands were as the following: angsd -doAbbababa2 1 -rmTrans 0 -blockSize 5,000,000 -doCounts 1 -enhance 1 -maxDepth 100. The D statistic was estimated using the formula [50], and z scores were calculated using all samples of each population. To test for significance, *p*-values were computed based on a two-tailed binomial test. Meanwhile, inference of gene flow was conducted based on genome-wide allele frequency data using the software TreeMix v1.12 [80].

Detection of inbreeding patterns

Detection of linkage disequilibrium (LD) patterns is necessary to infer historical changes of population and inbreeding patterns in the species demographic history [81, 82]. To assess the genomic extent of inbreeding in Forest Musk Deer, genome-wide LD was estimated for the three populations using PopLDdecay v1.0 [83]. The squared correlation coefficient (r^2) between any two loci from each population was calculated using VCFtools v0.1.13 [69], with the following parameters: "–ld -window -bp 500,000 -geno -r2", and average r^2 values were identified for pairwise SNPs by keeping the same distance.

To detect the genomic signatures of inbreeding in Forest Musk Deer, genome-wide runs of homozygosity (RoHs) was identified for each individual using the tool in PLINK v.1.9 [72], with the following parameters (-threshold 0.05 -snp 65 -kb 100 -missing 3 -gap 5,000 -density 5,000).

Liu *et al. BMC Genomics* (2022) 23:668

Page 9 of 11

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08896-9.

Additional file 1: Table S1. The sample information of 15 Forest Musk Deer individuals and the BioProject accession number to retrive raw data from the NCBI website is PRJNA765065. **Table S2.** Raw reads, sequencing error rate and GC content of 15 Forest Musk Deer samples. **Table S3.** Mapping and coverage statistics for 15 samples of Forest Musk Deer. **Table S4.** The kinship coefficient between individuals. The values below the diagonal refer to the kinship coefficient for all individual pairs. **Table S5.** The result of ABBA-BABA test results based on the Observe. Table S6. The result of ABBA-BABA test results based on the TransRem.

Additional file 2: Figure S1. Cross validation error (CV) plot from ADMIXTURE. **Figure S2.** Comparison of the genome-wide heterozygosity among species under different conservation status. (a) Crossoptilon mantchuricum exhibits lowest level of genetic diversity among bird species based on available estimates from genome-wide sequencing data. Genome-wide heterozygosity was a useful indicator of genetic diversity, which was measured as proportion of heterozygous SNPs per base pair and plotted in rank order for 91 mammal species including Forest musk deer. Colored bars indicate the IUCN Red List Categories and Criteria (https://www.iucnredlist.org/). References for each species were listed and provided in the supplementary dataset. Data of 89 mammal species except Forest musk deer and Père David's Deer were from the supplementary file in Hu JY, Hao ZQ, Frantz L, Wu SF, Chen W, Jiang YF, Wu H, Kuang WM, Li H, Zhang YP et al: Genomic consequences of population decline in critically endangered pangolins and their demographic histories. Natl Sci Rev 2020, 7(4):798-814.

Authors' contributions

D.F.H., and G.L., lead the project, and designed and conceived the study. B.F.Z., and X.L.H., collected and provided samples. G.L., B.F.Z., J.C., and C.L., analyzed and interpreted the data. T.T.X analyzed the data when in revision. G.L, S.Q.L., and D.F.H., drafted the manuscript. All authors revised and approved the final manuscript.

Availability of data and materials

The whole genome sequencing data for all 15 individuals generated in this study have been deposited at NCBI SRA Database with BioProject accession ID: PRJNA765065. The accession number for each sample of Forest musk deer can be seen in the Additional file 1: Table S1.

Declarations

Ethics approval and consent to participate

All experiments protocols were approved by Beijing Forestry University, which has an ethics committee (permit number: BJFU20181a09). All methods were carried out in accordance with relevant guidelines and regulations. All methods were reported in accordance with ARRIVE guidelines (https://arriveguidelines.org) for the reporting of animal experiments.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute of Wetland Research, Chinese Academy of Forestry, Beijing Key Laboratory of Wetland Services and Restoration, Beijing 100091, China. [2]College of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100085, China. [3]State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China. [4]College of Animal Science and Technology, Jiangxi Agricultural University, Nanchang 330022, China. [5]College of Plant Science, Jilin University, Changchun 130062, China.

References

1. Li H, Xiang-Yu J, Dai G, Gu Z, Ming C, Yang Z, Ryder OA, Li WH, Fu YX, Zhang YP. Large numbers of vertebrates began rapid population decline in the late 19th century. Proc Natl Acad Sci U S A. 2016;113(49):14079–84.
2. Wang W, Zheng Y, Zhao J, Yao M. Low genetic diversity in a critically endangered primate: shallow evolutionary history or recent population bottleneck? BMC Evol Biol. 2019;19(1):134.
3. Fraser DJ. Genetic diversity of small populations: Not always "doom and gloom"? Mol Ecol. 2017;26(23):6499–501.
4. Frankham R. Genetics and extinction. Biol Cons. 2005;126(2):131–40.
5. Baas P, van der Valk T, Vigilant L, Ngobobo U, Binyinyi E, Nishuli R, Caillaud D, Guschanski K. Population-level assessment of genetic diversity and habitat fragmentation in critically endangered Grauer's gorillas. Am J Phys Anthropol. 2018;165(3):565–75.
6. Casillas S, Barbadilla A. Molecular population genetics. Genetics. 2017;205(3):1003–35.
7. Charlesworth B, Charlesworth D. Population genetics from 1966 to 2016. Heredity (Edinb). 2017;118(1):2–9.
8. Lande R. Genetics and demography in biological conservation. Science. 1988;241(4872):1455–60.
9. Spielman D, Brook BW, Frankham R. Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci U S A. 2004;101(42):15261–4.
10. Grueber CE, Wallis GP, Jamieson IG. Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. Mol Ecol. 2008;17(18):3978–84.
11. Crow JF. Mid-century controversies in population genetics. Annu Rev Genet. 2008;42:1–16.
12. Primmer CR. From conservation genetics to conservation genomics. Ann N Y Acad Sci. 2009;1162:357–68.
13. Wakeley J. Recent trends in population genetics: more data! More math! Simple models? J Hered. 2004;95(5):397–405.
14. Fan H, Hu Y, Wu Q, Nie Y, Yan L, Wei F. Conservation genetics and genomics of threatened vertebrates in China. J Genet Genomics. 2018;45(11):593–601.
15. Ouborg NJ. Integrating population genetics and conservation biology in the era of genomics. Biol Lett. 2010;6(1):3–6.
16. Theodoridis S, Fordham DA, Brown SC, Li S, Rahbek C, Nogues-Bravo D. Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. Nat Commun. 2020;11(1):2557.
17. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364(6446):eaav6202.
18. Galarza JA, Sánchez-Fernández B, Fandos P, Soriguer R. Intensive management and natural genetic variation in red deer (cervus elaphus). J Hered. 2017;108(5):496–504.
19. Rao M, Zaw T, Htun S, Myint T. Hunting for a living: wildlife trade, rural livelihoods and declining wildlife in the Hkakaborazi National Park, north Myanmar. Environ Manage. 2011;48(1):158–67.
20. Singh PB, Saud P, Cram D, Mainali K, Thapa A, Chhetri NB, Poudyal LP, Baral HS, Jiang Z. Ecological correlates of Himalayan musk deer Moschus leucogaster. Ecol Evol. 2019;9(1):4–18.

Liu *et al. BMC Genomics*   (2022) 23:668

Page 10 of 11

21. Yi L, Dalai M, Su R, Lin W, Erdenedalai M, Luvsantseren B, Chimedtseren C, Wang Z, Hasi S. Whole-genome sequencing of wild Siberian musk deer (Moschus moschiferus) provides insights into its genetic features. BMC Genomics. 2020;21(1):108.
22. IUCN. The IUCN Red List of Threatened Species v. 2015. http://www.iucnredlistorg. Accessed 20 Dec 2021.
23. Jie H, Feng XL, Zhao GJ, Zeng DJ, Zhang CL, Chen Q. Research progress on musk secretion mechanism of forest musk deer. Zhongguo Zhong Yao Za Zhi. 2014;39(23):4522–5.
24. Su B, Wang YX, Lan H, Wang W, Zhang Y. Phylogenetic study of complete cytochrome b genes in musk deer (genus Moschus) using museum samples. Mol Phylogenet Evol. 1999;12(3):241–9.
25. Li L, Huang X, Liu G, Wang W, Wei N, Liu Y, Hu D, Meng M. The status of captive population of musk deer and analysis of its farming development in China. Sichuan J Zoology. 2012;31(3):492–6.
26. Li L, He L, Liu G, Liu S, Liu W, Meng M, Hu D. Discussion about relationship between biological characters and farming development of musk deer. Forestry Res Manage. 2012;2:26–9.
27. Hu XL, Liu G, Wang WX, Zhou R, Liu SQ, Li LH, Hu DF. Methods of preservation and flotation for the detection of nematode eggs and coccidian oocysts in faeces of the forest musk deer. J Helminthol. 2016;90(6):680–4.
28. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11(10):697–709.
29. Steiner CC, Putnam AS, Hoeck PE, Ryder OA. Conservation genomics of threatened animal species. Annu Rev Anim Biosci. 2013;1:261–81.
30. Wright BR, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. A demonstration of conservation genomics for threatened species management. Mol Ecol Resour. 2020;20(6):1526–41.
31. Hohenlohe PA, Funk WC, Rajora OP. Population genomics for wildlife conservation and management. Mol Ecol. 2021;30(1):62–82.
32. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27(9):489–96.
33. Angeloni F, Wagemaker N, Vergeer P, Ouborg J. Genomic toolboxes for conservation biologists. Evol Appl. 2012;5(2):130–43.
34. Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9(10):1205–18.
35. Ghoreishifar SM, Moradi-Shahrbabak H, Fallahi MH, Jalil Sarghale A, Moradi-Shahrbabak M, Abdollahi-Arpanahi R, Khansefid M. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, Bubalus bubalis. BMC Genet. 2020;21(1):16.
36. Foster Y, Dutoit L, Grosser S, Dussex N, Foster BJ, Dodds KG, et al. Genomic signatures of inbreeding in a critically endangered parrot, the kākāpō. G3 (Bethesda). 2021;11(11):jkab307.
37. Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. Genes (Basel). 2019;10(11):846.
38. Khan S, Nabi G, Ullah MW, Yousaf M, Manan S, Siddique R, Hou H. Overview on the role of advance genomics in conservation biology of endangered species. Int J Genomics. 2016;2016:3460416.
39. Gompert Z. Population genomics as a new tool for wildlife management. Mol Ecol. 2012;21(7):1542–4.
40. Latch EK. Integrating genomics into conservation management. Mol Ecol Resour. 2020;20(6):1455–7.
41. Bourgeois YXC, Warren BH. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. Mol Ecol. 2021;30(23):6036–71.
42. Peng H, Liu S, Zou F, Zeng B, Yue B. Genetic diversity of captive forest musk deer (Moschus berezovskii) inferred from the mitochondrial DNA control region. Anim Genet. 2009;40(1):65–72.
43. Feng H, Feng CL, Huang Y, Tang J. Structure of mitochondrial DNA control region and genetic diversity of Moschus berezovskii populations in Shaanxi Province. Genet Mol Res. 2016;15(2):gmr.15027578.
44. Cai Y, Yang J, Wang J, Yang Y, Fu W, Zheng C, Cheng J, Zeng Y, Zhang Y, Xu L, et al. Changes in the population genetic structure of captive forest musk deer (moschus berezovskii) with the increasing number of generation under closed breeding conditions. Animals (Basel). 2020;10(2):255.
45. Qi WH, Lu T, Zheng CL, Jiang XM, Jie H, Zhang XY, Yue BS, Zhao GJ. Distribution patterns of microsatellites and development of its marker in different genomic regions of forest musk deer genome based on high throughput sequencing. Aging (Albany NY). 2020;12(5):4445–62.
46. Yao G, Zhu Y, Wan QH, Fang SG. Major histocompatibility complex class II genetic variation in forest musk deer (Moschus berezovskii) in China. Anim Genet. 2015;46(5):535–43.
47. Cai R, Shafer AB, Laguardia A, Lin Z, Liu S, Hu D. Recombination and selection in the major histocompatibility complex of the endangered forest musk deer (Moschus berezovskii). Sci Rep. 2015;5:17285.
48. Fan J, Zheng X, Wang H, Qi H, Jiang B, Qiao M, Zhou J, Bu S. Analysis of genetic diversity and population structure in three forest musk deer captive populations with different origins. G3 (Bethesda). 2019;9(4):1037–44.
49. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
50. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014;46(8):919–25.
51. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol Biol Evol. 2015;32(1):244–57.
52. Wang D, Xu G, Wang H, He S, Pu S, Zheng X. Study on polymorphisms of microsatellites DNA of Chinese captive forest musk deer (Moschus berezovskii). Acta Theriologica Sinica. 2019;39(6):599–607.
53. Corlett RT. Restoration, Reintroduction, and Rewilding in a Changing World. Trends Ecol Evol. 2016;31(6):453–62.
54. Hu JY, Hao ZQ, Frantz L, Wu SF, Chen W, Jiang YF, Wu H, Kuang WM, Li H, Zhang YP, et al. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. Natl Sci Rev. 2020;7(4):798–814.
55. von Seth J, Dussex N, Díez-Del-Molino D, van der Valk T, Kutschera VE, Kierczak M, Steiner CC, Liu S, Gilbert MTP, Sinding MS, et al. Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. Nat Commun. 2021;12(1):2393.
56. Paijmans AJ, Stoffel MA, Bester MN, Cleary AC, De Bruyn PJN, Forcada J, Goebel ME, Goldsworthy SD, Guinet C, Lydersen C, et al. The genetic legacy of extreme exploitation in a polar vertebrate. Sci Rep. 2020;10(1):5089.
57. Xiao Y, Jiang H, Jiang W, Wang Y, Hu Z, Xu H. Evaluation of habitat fragmentation and landscape index for Moschus berezovskii in Fengxian county, Shaanxi Province. Journal of Zhejiang A&F University. 2008;25(3):331–5.
58. Xue W, Jiang H, Hu Z, Xu H. Changes of Moschus berezovskii distribution in Fengxian county of Shaanxi Province in last fifty years. Chin J Ecol. 2007;26(6):791–801.
59. Guang X, Lan T, Wan Q-H, Huang Y, Li H, Zhang M, Li R, Zhang Z, Lei Y, Zhang L, et al. Chromosome-scale genomes provide new insights into subspecies divergence and evolutionary characteristics of the giant panda. Science Bulletin. 2021;66(19):2002–13.
60. Liu X, Wang T, Wang T, Skidmore AK, Songer M. How do two giant panda populations adapt to their habitats in the Qinling and Qionglai Mountains. China Environ Sci Pollut Res Int. 2015;22(2):1175–85.
61. Yang S, Zhang T, Li Y, Xu S, Wronski T. Identifying personality traits and their potential application to the management of captive forest musk deer (Moschus berezovskii). Appl Anim Behav Sci. 2020;234(5): 105168.
62. Yao G. The evaluation of population density and the environment adaptive ability of forest musk deer[D]. Hangzhou: Zhejiang University; 2014.
63. Hoffmann M, Hilton-Taylor C, Angulo A, Böhm M, Brooks TM, Butchart SH, Carpenter KE, Chanson J, Collen B, Cox NA, et al. The impact of conservation on the status of the world's vertebrates. Science. 2010;330(6010):1503–9.
64. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
66. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
67. Wang P, Burley JT, Liu Y, Chang J, Chen D, Lu Q, Li SH, Zhou X, Edwards S, Zhang Z. Genomic consequences of long-term population decline in brown eared pheasant. Mol Biol Evol. 2021;38(1):263–73.
68. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim H-M, Jho S, Kim S, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun. 2013;4(1):2433.

Liu *et al. BMC Genomics*   (2022) 23:668

Page 11 of 11

69. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

70. Hutter S, Vilella AJ, Rozas J. Genome-wide DNA polymorphism analyses using VariScan. BMC Bioinformatics. 2006;7:409.

71. Zhang S, Li C, Li Y, Chen Q, Hu D, Cheng Z, Wang X, Shan Y, Bai J, Liu G. genetic differentiation of reintroduced père david's deer (elaphurus davidianus) based on population genomics analysis. Front Genet. 2021;12: 705337.

72. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

73. Alexander DH, Lange K. Enhancements to the Admixture algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:246.

74. Steinrücken M, Kamm J, Spence JP, Song YS. Inference of complex population histories using whole-genome sequences from multiple populations. Proc Natl Acad Sci U S A. 2019;116(34):17115–20.

75. Schiffels S, Wang K. MSMC and MSMC2: the multiple sequentially markovian coalescent. Methods Mol Biol. 2020;2090:147–66.

76. Pacifici M, Santini L, Marco MD, Baisero D, Rondinini C. Generation length for mammals. Nature Conservation. 2013;5(6025):87–94.

77. Hamlin JAP, Hibbins MS, Moyle LC. Assessing biological factors affecting postspeciation introgression. Evol Lett. 2020;4(2):137–54.

78. Martin SH, Amos W. Signatures of Introgression across the Allele Frequency Spectrum. Mol Biol Evol. 2021;38(2):716–26.

79. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 2014;15(1):356.

80. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8(11): e1002967.

81. Karimi K, Farid AH, Sargolzaei M, Myles S, Miar Y. Linkage disequilibrium, effective population size and genomic inbreeding rates in american mink using genotyping-by-sequencing data. Front Genet. 2020;11:223.

82. Chitneedi PK, Arranz JJ, Suarez-Vega A, García-Gámez E, Gutiérrez-Gil B. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Anim Genet. 2017;48(4):436–46.

83. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786–8.

Publisher's Note