

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Deep Ancestral Introgressions between Ovine Species Shape Sheep Genomes via Argali-Mediated Gene Flow.

### Permalink

<https://escholarship.org/uc/item/8qc8h2b6>

### Journal

Molecular Biology and Evolution, 41(11)

### Authors

Lv, Feng-Hua

Wang, Dong-Feng

Zhao, Si-Yi

et al.

### Publication Date








2024-11-01

### DOI

10.1093/molbev/msae212

Peer reviewed

# Deep Ancestral Introgressions between Ovine Species Shape Sheep Genomes via Argali-Mediated Gene Flow

Feng-Hua Lv <sup>1</sup>, Dong-Feng Wang <sup>2,3</sup>, Si-Yi Zhao <sup>1</sup>, Xiao-Yang Lv <sup>4</sup>, Wei Sun <sup>4</sup>, Rasmus Nielsen <sup>5,6,7,\*</sup>, Meng-Hua Li <sup>1,\*</sup>

<sup>1</sup>Frontiers Science Center for Molecular Design Breeding (MOE), State Key Laboratory of Animal Biotech Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

<sup>2</sup>CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing, China

<sup>3</sup>College of Life Sciences, University of Chinese Academy of Sciences (UCAS), Beijing, China

<sup>4</sup>International Joint Research Laboratory, Universities of Jiangsu Province of China for Domestic Animal Germplasm Resources and Genetic Improvement, Yangzhou 225001, China

<sup>5</sup>Department of Integrative Biology, University of California at Berkeley, Berkeley, CA 94720, USA

<sup>6</sup>Department of Statistics, UC Berkeley, Berkeley, CA 94707, USA

<sup>7</sup>Globe Institute, University of Copenhagen, Copenhagen 1350, Denmark

\*Corresponding authors: E-mails: menghua.li@cau.edu.cn; rasmus\_nielsen@berkeley.edu.

Associate editor: Daniel Falush

## Abstract

Previous studies revealed extensive genetic introgression between *Ovis* species, which affects genetic adaptation and morphological traits. However, the exact evolutionary scenarios underlying the hybridization between sheep and allopatric wild relatives remain unknown. To address this problem, we here integrate the reference genomes of several ovine and caprine species: domestic sheep, argali, bighorn sheep, snow sheep, and domestic goats. Additionally, we use 856 whole genomes representing 169 domestic sheep populations and their six wild relatives: Asiatic mouflon, urial, argali, snow sheep, thinhorn sheep, and bighorn sheep. We implement a comprehensive set of analyses to test introgression among these species. We infer that the argali lineage originated ~3.08 to 3.35 Mya and hybridized with the ancestor of Pachyceriforms (e.g. bighorn sheep and snow sheep) at ~1.56 Mya. Previous studies showed apparent introgression from North American Pachyceriforms into the Bashibai sheep, a Chinese native sheep breed, despite of their wide geographic separation. We show here that, in fact, the apparent introgression from the Pachyceriforms into Bashibai can be explained by the old introgression from Pachyceriforms into argali and subsequent recent introgression from argali into Bashibai. Our results illustrate the challenges of estimating complex introgression histories and provide an example of how indirect and direct introgression can be distinguished.

**Key words:** *Ovis* genus, argali, sheep, ancestral introgression, genomes, demographic history.

## Introduction

Introgression or hybridization has been an important process for introducing both deleterious (e.g. Petr et al. 2019) and adaptive (e.g. Jones et al. 2018) genetic variation in many species. Recent studies have shown that introgression commonly affects speciation, adaptation (Oziolor et al. 2019), morphological trait evolution (Zou et al. 2022), and phylogenies (Nelson et al. 2021). Introgression between domestic animals and wild relatives has been reported in many previous studies (supplementary table S1, Supplementary Material online). When the introgression occurs from wild species into domestic species, it is often considered a source of adaptive variation in the response

to environmental change or as a source of genetic variation for animal or plant breeding (Frantz et al. 2013; Ai et al. 2015; Chen et al. 2018; Wu et al. 2018; Liu et al. 2019; Lawal et al. 2020; Zhou et al. 2020; Yu et al. 2021). Hybridization between domestic and wild relatives, however, may also lead to a loss of genetic variation in wild animals and consequently reduce their resilience to disease and climate change (Redford and Dudley 2018).

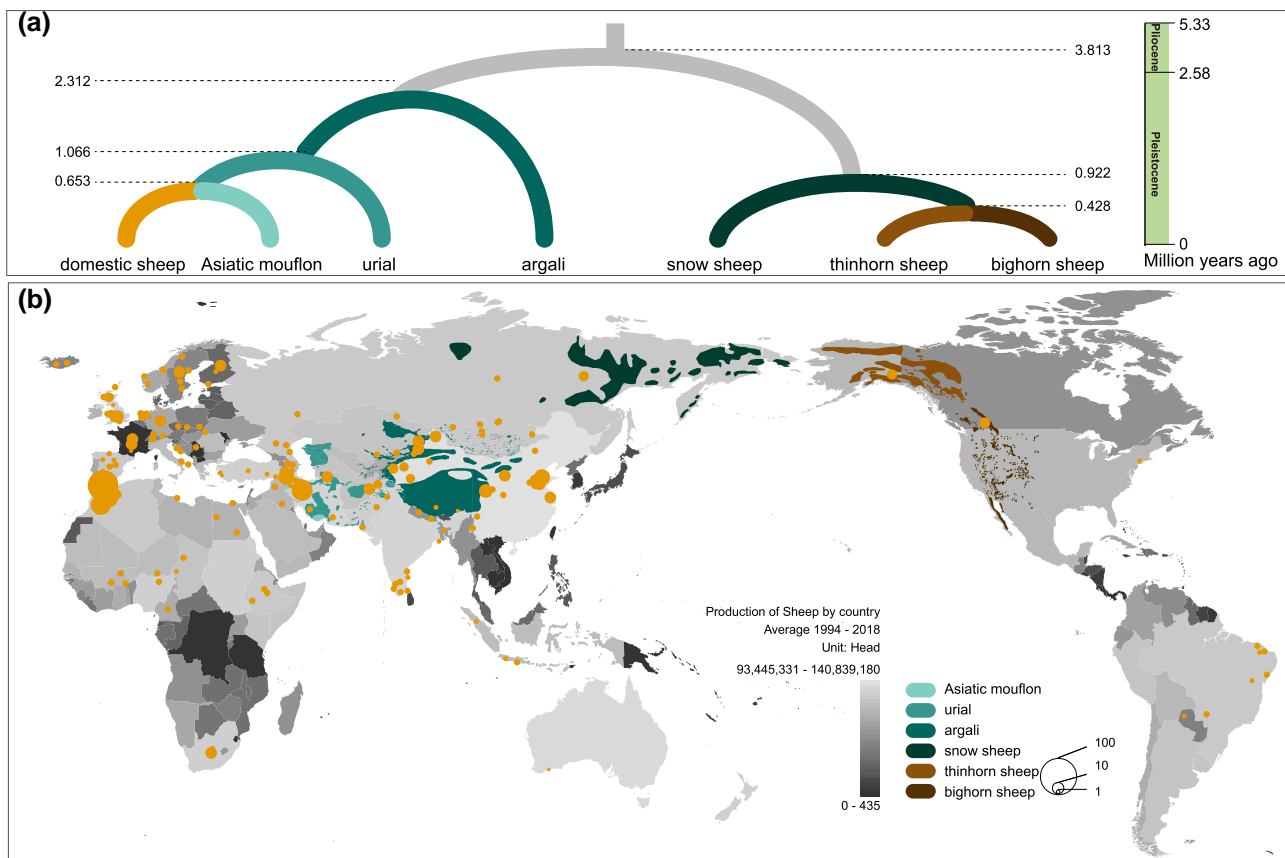
The evolutionary history of the *Ovis* genus, which includes domestic sheep and six wild relatives, has been significantly influenced by introgression (Chen et al. 2021). These species inhabit diverse ecological and climate conditions (see Fig. 1) and show a variety of adaptive traits

Received: March 10, 2024. Revised: September 12, 2024. Accepted: October 04, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Open Access



**Fig. 1.** Phylogeny, geographic distribution of *Ovis* species, and sampling locations. a) Phylogenetic tree of the genus *Ovis* with divergence times inferred by Upadhyay et al. (2021). b) Geographic distribution of six wild species (IUCN, [www.iucnredlist.org](http://www.iucnredlist.org)) and production of sheep by country and sample location ( $n = 856$ ), including six wild sheep and 169 domestic sheep populations. The dots indicate the sampling locations of domestic and wild sheep, and the pie charts show the sample numbers.

(Rezaei et al. 2010; Santos et al. 2021; Upadhyay et al. 2021), including insulating properties of fur and fat reserves of snow sheep and coat color pattern of Stone's sheep. Previous investigations showed that different species in the genus *Ovis* can hybridize and produce viable hybrid offspring (Nadler et al. 1973; Valdez et al. 1978; Li et al. 2022). Thus, species in the *Ovis* genus provide an excellent model for investigating the prevalence of historical interspecific hybridization and its impacts on species evolution. *Ovis* consists of 2 phylogenetic groups, the Arealiforforms/Moufloniforforms and Pachyceriforforms, and represents a complex phylogeny shaped by interspecific hybridization and introgression (Valdez et al. 1978; Chen et al. 2021). The Arealiforforms/Moufloniforforms comprises four species (domestic sheep [*Ovis aries*], Asiatic mouflon [*Ovis gmelini*], urial [*Ovis vignei*], and argali [*Ovis ammon*]), whose common ancestor can be traced to  $\sim 2.3$  Mya (Upadhyay et al. 2021), while the Pachyceriforforms consist of snow sheep (*Ovis nivicola*), thornhorn (*Ovis dalli*), and bighorn (*Ovis canadensis*) sheep. Whole-genome sequencing (WGS) analyses have revealed that North American wild sheep (thornhorn and bighorn sheep) are monophyletic sister species to snow sheep (Santos et al. 2021; Upadhyay et al. 2021), supporting Cowan's hypothesis that North American wild sheep diverged from a snow sheep-like

species (Cowan 1940). Also, previous genomic studies show that ancient hybridization may facilitate the formation of subspecies (*O. dalli stonei*; Santos et al. 2021) and adaptation of snow sheep to the subarctic climate (Upadhyay et al. 2021). Recent analyses have documented introgression from wild relatives to domestic sheep populations, affecting morphological (e.g. coat color and horns) and adaptive characteristics (e.g. high altitude and pneumonia) (Hu et al. 2019; Deng et al. 2020; Cao et al. 2021; Cheng et al. 2023).

Lv et al. (2022) detected apparent introgression from the North American thornhorn/bighorn sheep into Chinese Bashibai sheep. However, the Bering Strait is a physical barrier for terrestrial animals between Eurasia and the North American continent, impeding inter/intraspecific hybridization or gene flow events affecting species on the two different continents around 11 ka (Hu et al. 2010; Pacheco et al. 2022; Salis et al. 2022), the approximate time of sheep domestication (Sánchez-Villagra 2022). This conflict between observations of genomic introgression and phylogeographic patterns of the *Ovis* genus has inspired our interest in exploring interspecies introgression and the demographic history of the Pachyceriforforms species. To this end, we performed a comprehensive integrated analysis. We included the reference genomes of 5 species—

domestic sheep, argali, bighorn sheep, snow sheep, and goats—and compiled a large genomic data set consisting of 856 individuals from 169 domestic populations and six wild species (Fig. 1b; supplementary table S2, Supplementary Material online). We analyzed signals of introgression using multiple statistical methods with a focus on introgression between allopatric species, such as domestic sheep in Eurasia and wild sheep in Northern America.

## Results

### Variants and Single-Copy Orthologous Genes

We collected WGS reads from 856 samples, including 70 wild sheep from 6 species and 786 domestic sheep from 169 populations (supplementary table S2, Supplementary Material online). Approximately 97.79% of the reads were aligned to the sheep reference genome *Oar\_rambouillet\_v1.0*. The sequencing depth ranged from 6.63 to 28.01× in domestic sheep and from 9.45 to 30.11× in wild sheep. After single nucleotide polymorphism (SNP) filtering, 134.8 million SNPs were identified. Additionally, we aligned 5 reference genomes—domestic sheep, argali, bighorn sheep, snow sheep, and goats—and obtained 3,878 single-copy orthologous genes shared among domestic sheep, argali, bighorn sheep, and goats; 2,614 shared among domestic sheep, argali, bighorn sheep, and goats; and 3,140 shared among domestic sheep, argali, snow sheep, and goats, each with sequences exceeding 1,000 bp.

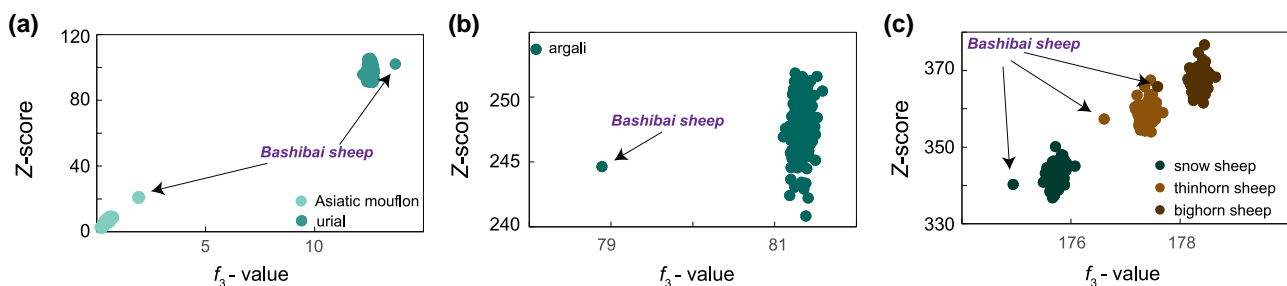
### Genetic Relatedness between Wild and Domestic Sheep Species

We first used the outgroup  $f_3$ -statistics to measure allele sharing. The Asiatic mouflon, being the direct ancestor of domestic sheep, unsurprisingly shares many alleles with them. We found that Asiatic mouflon had similar genetic divergence ( $f_3 = 0.200$  to  $0.733$ ) from sheep populations of different geographic origins, except for Bashibai sheep ( $f_3 = 1.93$ ;  $Z$ -scores =  $20.86$ ) (Fig. 2a; supplementary fig. S1a and table S3, Supplementary Material online). Compared to the argali, the urial showed closer genetic relatedness to domestic sheep, as reflected by  $f_3$  values, which

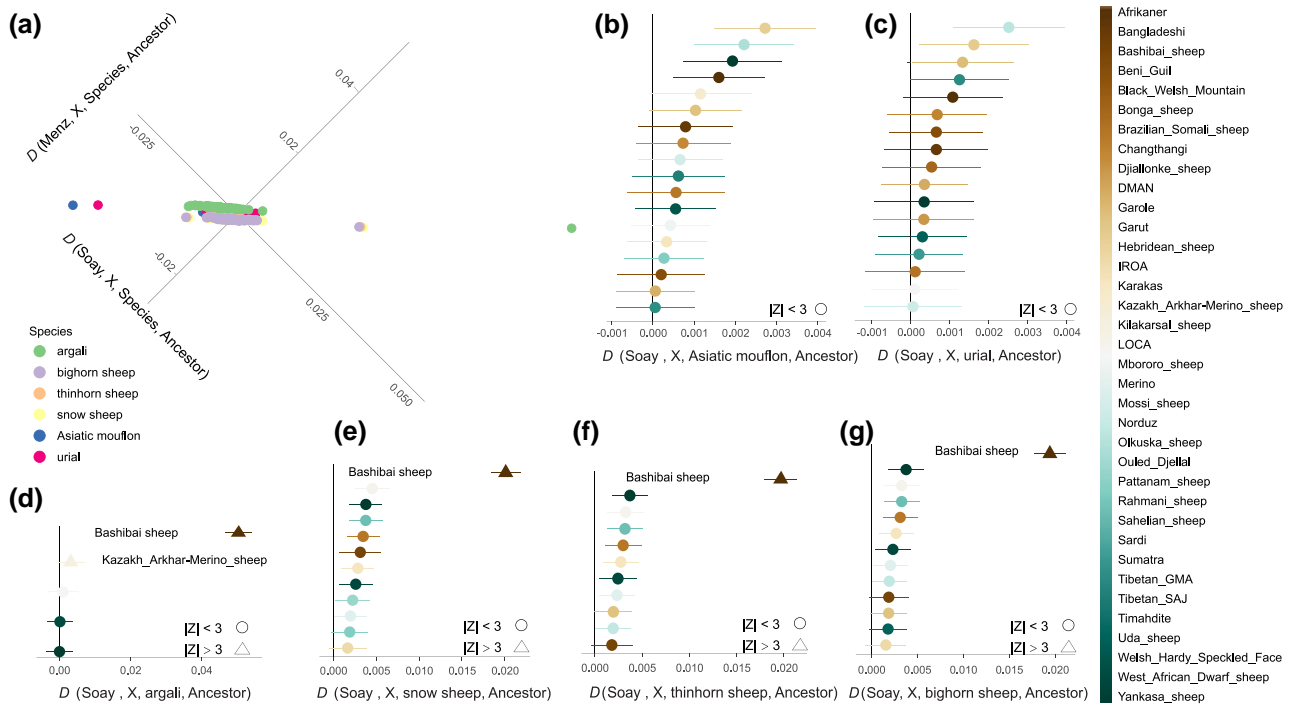
indicated similar pattern of allele sharing to that of Asiatic mouflon at the global geographic scale (Fig. 2a and b; supplementary fig. S1 and tables S4 and S5, Supplementary Material online). The three species of Pachyceriforms showed similarly low levels of allele sharing with domestic sheep populations (Fig. 2c; supplementary fig. S1 and tables S6 to S8, Supplementary Material online). In general, the  $f_3$ -statistics did not support a correlation of geographic distance and genetic distance between wild sheep and domesticated sheep populations, and we did not find increased allele sharing between wild relatives and their sympatric modern domestic sheep populations (Fig. 2; supplementary fig. S1 and tables S3 to S8, Supplementary Material online). Furthermore, we calculated the fixation index ( $F_{ST}$ ) between wild relatives and sheep populations (supplementary figs. S2 to S7, Supplementary Material online), and found similar results to those obtained with  $f_3$ -statistics (supplementary fig. S8, Supplementary Material online). We also observed a specific population of Bashibai sheep, whose  $f_3$  values significantly differed from the other domestic sheep populations ( $t$ -test,  $P < 0.01$ ; Fig. 2a to c; supplementary tables S3 to S8, Supplementary Material online) when compared to all the six species. Compared to the other sheep populations, Bashibai sheep showed larger genetic divergence to Asiatic mouflon and urial, but exhibited closer genetic affinity with argali and the Pachyceriforms species (Fig. 2a to c; supplementary tables S3 to S8, Supplementary Material online).

### Low-Frequency Introgressions between Wild and Modern Sheep

To further verify and quantify introgression from wild sheep into modern populations, we calculated allele sharing between wild relatives and domestic sheep populations using  $D$ -statistics (Patterson et al. 2012). Initially, Menz sheep and Soay sheep were used as proxies for unadmixed populations. As the  $D$ -statistics results were consistent using either of these two breeds (Fig. 3a; supplementary tables S9 to S14, Supplementary Material online), we subsequently report only analyses with the Soay sheep as a putatively unadmixed population. We



**Fig. 2.** Genetic differences between wild sheep and modern domestic sheep populations. a)  $f_3$ -statistics measure shared alleles between Asiatic mouflon/urial and modern domestic sheep populations. b)  $f_3$ -statistics measure shared alleles between argali and modern domestic sheep populations. c)  $f_3$ -statistics measure shared alleles between each species of the Pachyceriforms (snow sheep, thinhorn sheep, and bighorn sheep) and modern domestic sheep populations.



**Fig. 3.** Differential affinities of wild sheep to modern domestic sheep populations. a)  $D$ -statistics for the differential affinities of six wild sheep to present-day domestic sheep populations. Menz ( $y$  axis) and Soay sheep ( $x$  axis) were chosen as unadmixed populations. b) Differential affinities of Asiatic mouflon to modern domestic sheep populations. c) Differential affinities of urial to modern domestic sheep populations. d) Differential affinities of argali to modern domestic sheep populations. e) Differential affinities of snow sheep to modern domestic sheep populations. f) Differential affinities of thinhorn sheep to modern domestic sheep populations. g) Differential affinities of bighorn sheep to modern domestic sheep populations. Only the populations with positive  $D$  values are listed. The horizontal lines indicate three standard errors.

found 36/501 positive  $D$ -statistics and 5/36 were marginally significant ( $|Z| > 3$ ) (Fig. 3b to g). In addition to a significant signal of gene flow between argali and Kazakh Arkhar–Merino sheep (Fig. 3d), common introgression signals from argali, snow sheep, thinhorn sheep, and bighorn sheep into Bashibai were observed with  $Z$ -values as large as 26.60 (Fig. 3d to g).

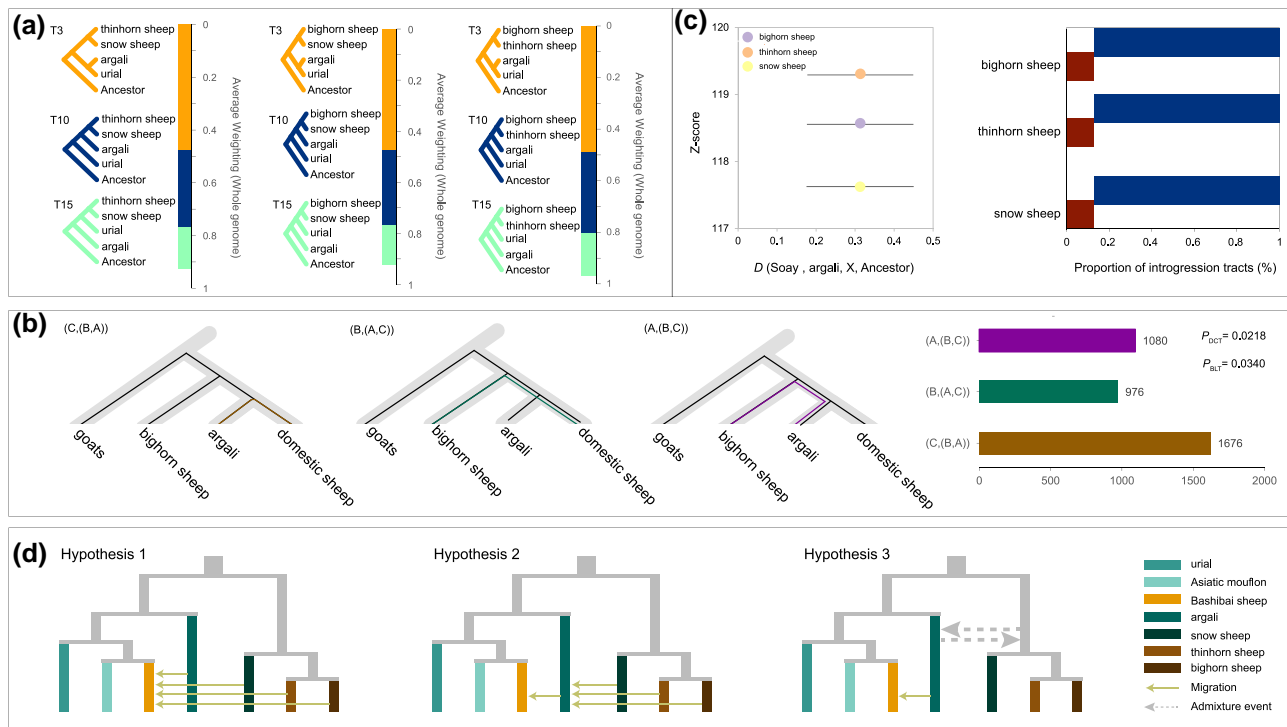
### Deep Ancient Introgression

The  $D$ -statistics analysis indicated introgression of wild sheep such as snow, thinhorn, and bighorn sheep into Bashibai sheep (Fig. 3e to g). However, while this introgression pattern previously has been reported (Lv et al. 2022), it showed conflicts with the fact that there is a substantial geographic barrier between the North American sheep and the Chinese Bashibai breed. A potential explanation is that introgression of wild sheep, such as snow, thinhorn, and bighorn sheep into Bashibai sheep, is a argali-mediated introgression, in which recent introgression occurred between Bashibai sheep and argali (Fig. 3d) and deep ancient introgression occurred between argali and the Pachyceriforms. To explore whether the apparent similarity between Bashibai sheep and the Pachyceriforms can be explained with gene flow through argali, we calculated  $D$ -statistics of the form (Bashibai sheep, argali, Pachyceriforms [snow sheep, thinhorn sheep, and bighorn sheep], and ancestor) to test the genetic relatedness among the Pachyceriforms, argali, and Bashibai sheep. The result showed significant affinity between the Pachyceriforms

and argali relative to the Pachyceriforms and Bashibai sheep (supplementary fig. S9, Supplementary Material online). To further elucidate this problem, we used several alternative methods for investigating the history of introgression between argali and the Pachyceriforms, including TreeMix (Pickrell and Pritchard 2012), phylogenetic weighting (Martin and Van Belleghem 2017), and the discordant-count test (DCT) + branch-length test (BLT) (Suvorov et al. 2022).

TreeMix analyses showed apparent introgression of the Pachyceriforms into argali (supplementary figs. S10 and S11, Supplementary Material online). For phylogenetic weighting analysis, we analyzed three data sets that each included five taxa: (i) urial, argali, thinhorn sheep, snow sheep, and the outgroup ancestor; (ii) urial, argali, bighorn sheep, snow sheep, and the outgroup ancestor; and (iii) urial, argali, thinhorn sheep, bighorn sheep, and the outgroup ancestor. In the three analyses, the 15 possible topologies that describe the relationship between Argaliforms, Moufloniforms, and Pachyceriforms were interrogated at the whole-genome level (supplementary figs. S12 to S14, Supplementary Material online). The three phylogenetic weighting analyses revealed similar topologies (Fig. 4a; supplementary figs. S12 to S14, Supplementary Material online). The 3 most common topologies were T3, T10, and T15 (Fig. 4a). T3 is consistent with the expected species tree, in which the three species of Pachyceriforms form a monophyletic group and the two Argaliforms and Moufloniforms form a monophyletic taxon (Rezaei et al. 2010; Chen et al. 2021), whereas T10 is consistent with





**Fig. 4.** Overview of the phylogenomic methods used to detect the introgression between genus *Ovis*. a) The three top weighting taxon topologies for these four taxa. Urrial, argali, snow sheep, and thinhorn sheep on the left; urrial, argali, snow sheep, and bighorn sheep on the middle; urrial, argali, snow sheep, and bighorn sheep on the right. b) The DCT + BLT was used to identify introgression from bighorn sheep into argali. c) Similar introgression patterns from the three species of the Pachyceriforms into argali.  $D$ -statistics estimate introgression on the left;  $f_d$ -statistics measure the proportion of introgression tracts on the right. d) Three alternative models with different introgressed events across the genus *Ovis*.

interspecific introgression or ILS in the ancestor of Argaliforms/Moufloniforms. T15 also agrees with the hypothesis of interspecific introgression or ILS in the ancestor of Argaliforms/Moufloniforms. However, under ILS, T10 and T15 should be equally common in the genome. The fact that T10 is more common than T15 (Fig. 4a) (Mann–Whitney  $U$  test,  $P = 1.715e-14$ ; see [Materials and Methods](#)) suggests the presence of introgression, for example, between argali and the Pachyceriforms.

To further investigate the pattern observed in Fig. 4a, we investigated the topologies of trees of bighorn sheep, argali, and domestic sheep, in which bighorn sheep, argali, and domestic sheep represent lineage of the Pachyceriforms, the Argaliforms, and the Moufloniforms, respectively, using BUSCO single-copy orthologous genes. We used two recently proposed tests of introgression, DCT and BLT (Suvorov et al. 2022), based on gene tree topology counts with a  $\chi^2$  goodness-of-fit test and branch lengths using a Mann–Whitney  $U$  test, respectively. A total of 3,878 gene trees were inferred from BUSCO single-copy orthologous genes, and among these, we identified 1,080 discordant gene trees of the form ([bighorn sheep, argali], domestic sheep) and 976 discordant gene trees of the form ([bighorn sheep, domestic sheep], argali), respectively, which are both different from the species tree (Fig. 4b). We observed significantly more trees with topology ([bighorn sheep, argali], domestic sheep) than with topology ([bighorn sheep, domestic sheep], argali) ( $\chi^2$  goodness-of-fit test,  $P_{DCT} = 0.0218$ ) and

corresponding significant deviations from the expected null distribution of  $d_{T1} = d_{T2}$  under ILS alone (Mann–Whitney  $U$  test,  $P_{BLT} = 0.0340$ ) (Fig. 4b). We conducted analogous tests on snow sheep, argali, and domestic sheep and found no significant excess of trees the topology ([snow sheep, argali], domestic sheep) over ([snow sheep, domestic sheep], argali) ( $\chi^2$  goodness-of-fit test,  $P_{DCT} = 0.1854$ ) (supplementary fig. S15, Supplementary Material online). The average genetic distance was lower for the topology ([snow sheep, argali], domestic sheep) ( $d_{T2} = 0.3359$ ) compared to ([snow sheep, domestic sheep], argali) ( $d_{T1} = 0.3489$ ), although this difference was not statistically significant (Mann–Whitney  $U$  test,  $P_{BLT} = 0.1435$ ) (supplementary fig. S15, Supplementary Material online). The excess of gene trees with the ([bighorn sheep, argali], domestic sheep) topology, the inequality  $d_{T1} \neq d_{T2}$ , and lower average genetic distance for the topology ([snow sheep, argali], domestic sheep) support our previous results.

We next used  $D$ -statistics and  $f_d$ -statistics to quantify the amount of introgression between argali and the Pachyceriforms. We found three positive  $D$ -statistics that were marginally significant ( $|Z| > 3$ ), and introgression from snow sheep, thinhorn sheep, and bighorn sheep into argali was observed with similar proportions of introgression tracts (12.87% to 13.05%) with each other (Fig. 4c). Furthermore, our observations indicated a significant overlap of  $\sim 67.55\%$  among the introgressed regions inferred to be derived snow sheep, thinhorn sheep, and bighorn sheep in the genome of Bashibai (supplementary fig. S16,

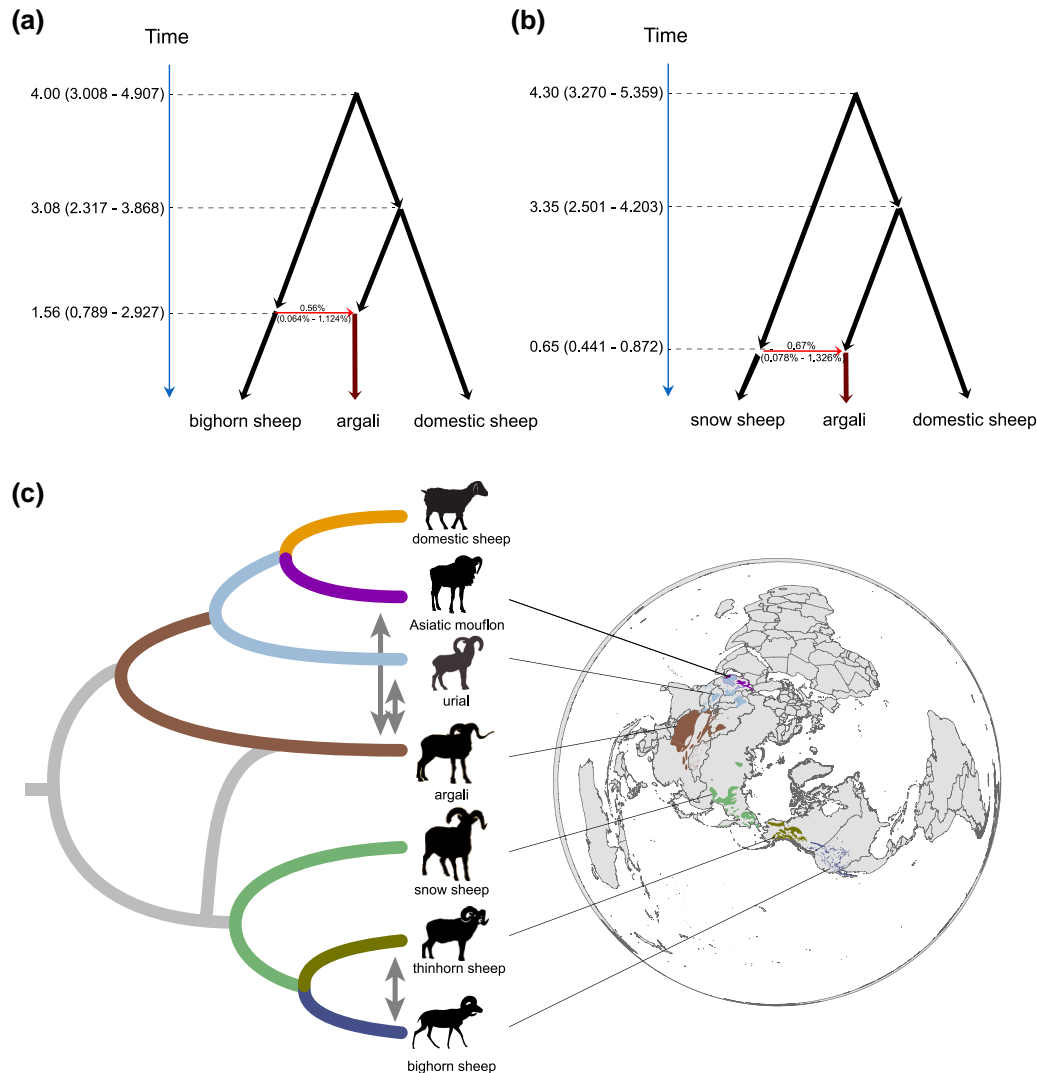
Supplementary Material online). These results suggest a deep ancient introgression in argali from the ancestor of the Pachyceriforms although more recent gene flow among multiple species cannot be ruled out. In Fig. 4d, we outline three possible scenarios that may be compatible with the observations of excess allele sharing reported so far.

To further distinguish between these hypotheses (Fig. 4d), we used analysis of the multidimensional site frequency spectra (SFS) using a composite likelihood method (Excoffier et al. 2013), which is well suited for investigating complex scenarios using large genomic data sets in this study. In accordance with our previous results, the best supported model with the highest average value of estimated  $\log_{10}(\text{likelihood})$  is hypothesis 3, which assumes ancestral introgression between argali and the Pachyceriforms (Fig. 4d; supplementary fig. S17, Supplementary Material online). We also used a Bayesian approach that models the multispecies-coalescent-with-introgression model (MSci) (Flouri et al. 2020) to estimate the timing and levels of ancestral hybridization between argali and

the species of Pachyceriforms. Using bighorn sheep, we obtained estimates of hybridization between argali occurring  $\sim 1.56$  Mya (95% HPD: 0.789 to 2.927 Mya) with an admixture fraction of  $\sim 0.56\%$  (95% HPD: 0.064% to 1.124%) in the argali (Fig. 5a). When using snow sheep, we estimate an admixture time of  $\sim 0.65$  Mya (95% HPD: 0.441 to 0.872 Mya) and an admixture fraction of  $\sim 0.67\%$  (95% HPD: 0.078% to 1.326%) in the argali from snow sheep (Fig. 5b). We notice that the HPDs from these analyses are overlapping and that the results may be somewhat sensitive to assumptions regarding effective population sizes through time, and therefore in itself, do not provide evidence for multiple introgression events. Nonetheless, the results are compatible with hypothesis of old introgression into the argali.

### The Distribution of Introgressed Segments in the Bashibai Sheep Genome

At the population level, we found significant introgression signals from argali into Bashibai sheep in this and previous



**Fig. 5.** Overview of phylogenetic scenario of the *Ovis* genus. Demographic coalescent-based model evaluated using the MSci BPP program: bighorn sheep, argali, and domestic sheep a) and snow sheep, argali, and domestic sheep b). A new simple model of the phylogeny scenario of the *Ovis* genus is supported by this study c).

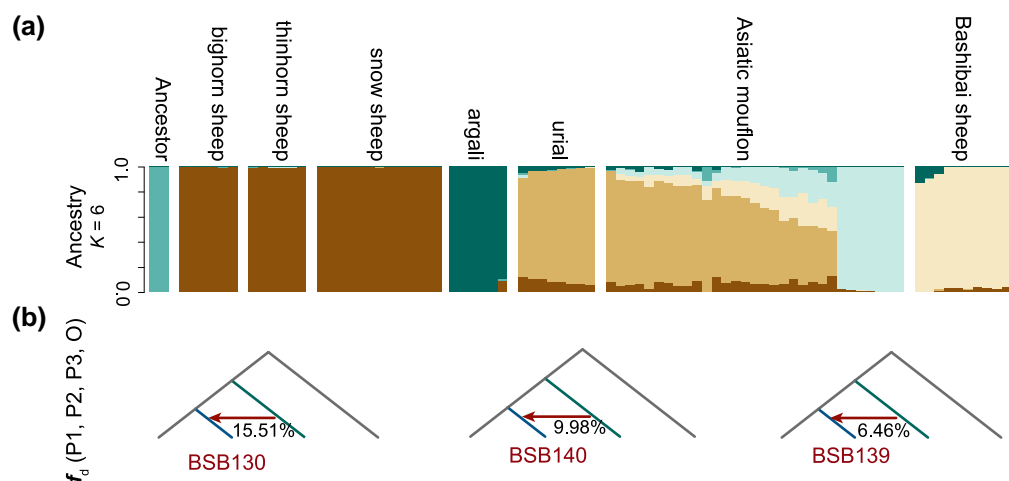
studies (Lv et al. 2022), but the degree of introgression varied considerably between individuals of Bashibai sheep.  $D$ -statistics revealed gene flow in only 3 out of the 10 individuals (BSB130, BSB140, and BSBS 139) from argali ( $|Z| > 3$ ; [supplementary fig. S18, Supplementary Material](#) online). We examined the genomic characteristics of the introgressed regions in terms of proportion, gene density, and distribution. The genome proportions of argali in BSB130, BSB140, and BSBS 139 were estimated to be 15.51%, 9.98%, and 6.46%, respectively ([Fig. 6a and b; supplementary fig. S19, Supplementary Material](#) online); correspondingly, 1,256, 834, and 565 blocks with introgression signals, respectively, were identified based on the  $f_d$  values ([supplementary table S15, Supplementary Material](#) online). This is consistent with artificial hybridization between argali and Bashibai sheep in Xinjiang (Joken-Anwax et al., 2007; Joken-Anwax, 2010), in which introgression regions may have heterogenized in the population. We further investigated the distribution of argali introgressed tracts in Bashibai sheep and in the Pachyceriforms. We found 1,846 introgression tracts from the Pachyceriforms into argali and among these ([supplementary fig. S20, Supplementary Material](#) online); 537 (24.55%) are also found as introgression tracts from argali into Bashibai ([supplementary fig. S21, Supplementary Material](#) online).

We annotated 857 functional genes within 1,846 introgression tracts ([supplementary table S16, Supplementary Material](#) online) and identified enrichment in 7 GO/KEGG terms ( $P_{adj} < 0.01$ ), including G protein-coupled receptor signaling pathway (GO:0007186) and olfactory transduction (oas04740) ([supplementary table S17, Supplementary Material](#) online). To explore potential evolutionary characteristics of the introgressed regions, we compared GC content, nucleotide diversity ( $H_e$ ), transposons elements (TEs), and  $dN/dS$  ratios with nonintrogressed regions. The analysis revealed no significant differences between introgressed and nonintrogressed tracts, except for some variation in  $H_e$  ([Supplementary Figs. 22 to 24, Supplementary Material](#) online).

## Discussion

Previous whole-genome analyses uncover signals of genetic introgression among the *Ovis* genus, in particular between domestic sheep and North American wild sheep such as bighorn sheep and thinhorn sheep, and identified introgressed genes associated with adaptation and morphological traits, such as *RXFP2* and *HBB* (Hu et al. 2019; Santos et al. 2021; Upadhyay et al. 2021). However, allopatric interspecific introgression events between domestic sheep and the Pachyceriforms species have never been documented and seem unlikely given their geographic separation. The significant introgression between Bashibai sheep and North American wild sheep (thinhorn and bighorn sheep) conflicts with the presence of a significant geographic barrier between them in the form of the Bering Strait, which separated Asia and North America beginning in the Holocene (Jakobsson et al. 2017), earlier than the domestication time of sheep (Colledge et al. 2005; Zeder 2008). In addition, no evidence has documented migrations of wild and domestic sheep between East Asia and America (Dunmire 2013; Spangler et al. 2017). However, occasional human-mediated hybridization remains a possibility.

We estimated the relative genomic proportions of the 15 different rooted topologies for the phylogeny urial, argali, snow sheep, thinhorn, and bighorn sheep. The commonly accepted species topology was the most common, but two other topologies, T10 and T15, were also commonly observed ([Fig. 4a; supplementary figs. S12 to S14, Supplementary Material](#) online). Under ILS, these topologies should be equally common, but the fact that T10 is more common than T15 ([Fig. 4a](#)) suggests the presence of introgression between argali and the Pachyceriforms, in accordance with hypothesis 3 from [Fig. 4d](#). Further analyses using a data set of single-copy ortholog genes including domestic sheep, argali, bighorn sheep, and goats, using branch length and gene tree count-based statistical approaches (Suvorov et al. 2022), again suggested introgression into the argali ([Fig. 4b](#)) and substantial argali introgression into Bashibai. Many of the introgression fragments in the argali were shared between Bashibai and



**Fig. 6.** Characteristics of introgressed tracts in Bashibai sheep from argali. a) Admixture proportions estimated by admixture analysis ( $K=6$ ). b) The proportion of introgressed tracts measured by  $f_d$ -statistics.



the Pachyceriforms. In fact, 537 of 1,846 introgressed tracts from the Pachyceriforms into argali were also found in the Bashibai sheep individuals ([supplementary fig. S21, Supplementary Material](#) online), conforming the hypothesis of ancestral hybridization into the argali later transmitted from the argali into the Bashibai ([Fig. 4d](#)) and contradicting the hypothesis of differential introgression of different genomic regions by different donor species ([Zheng et al. 2020](#)).

The fossil record from Yukon in the Middle Pleistocene Gold Run tephra ( $735 \pm 88$  kya) ([Osborn 1912](#); [Frick 1937](#); [Guthrie 1968](#)) supports the hypothesis of North America sheep deriving from a *nivicola*-like ancestor ([Cowan 1940](#)). Furthermore, the Bering Land Bridge connected East Asia and North America during several periods of the Pleistocene, thus providing the opportunity for the contact between argali/Asiatic mouflon/urial and Pachyceriforms ([Hu et al. 2010](#); [Batchelor et al. 2019](#)), which may have led to hybridization events between Argaliforms/Moufloniforms and Pachyceriforms, such as thinhorn sheep and Asiatic mouflon ([Chen et al. 2021](#)). As the argali are indigenous to the area in which Bashibai sheep are being bred, direct introgression between these 2 species is a likely scenario.

Glacial-interglacial cycles over the past million years have been associated with the formation and loss of the Bering Land Bridge, which can connect East Asia and North America. This formation of the Beringian land bridge provides the opportunity for introgression between otherwise genetically isolated species. Our results provide evidence of ancestral introgressions between Argaliforms/Moufloniforms and Pachyceriforms. Similar patterns of introgression have been documented in bears and horses ([Vershinina et al. 2021](#); [Lan et al. 2022](#)). Such introgression may also be a vehicle for the exchange of adaptive alleles. For instance, gene flow from the ancestors of extant brown bears into polar bears around 150 kya may have introduced new genetic variations that facilitated their adaptation to changing environments ([Lan et al. 2022](#)). Additionally, Eurasian horses diverged from North American populations  $\sim 1.0$  to  $0.8$  Mya, with bidirectional gene flows occurring between  $\sim 875$  to  $652$  kya and  $\sim 200$  to  $50$  kya ([Vershinina et al. 2021](#)). Our study, together with other recent studies, provides evidence that interglacial cycles dominate the alternating history of introgression and divergence in mammalian species in Eurasia and North America.

In conclusion, we identified apparent genetic introgression between domestic sheep in East Asia and wild sheep in North America and explored possible explanations. We uncovered a detailed evolutionary history of the *Ovis* genus, in which the argali originated  $\sim 3.08$  to  $3.35$  Mya and experienced an ancient hybridization event with the ancestor of the Pachyceriforms and much later introgression with domesticated sheep. Argali-mediated introgression instead of direct hybridization between domestic sheep and other wild relatives (snow sheep, thinhorn sheep, and bighorn sheep) shapes the domestic sheep genome. Our study provides an example of the difficulties in identifying the exact history of introgression between species. In particular, it can be very difficult to

distinguish between direct and indirect gene flow and simple statistics, such as Patterson's *D*, may not be sufficient to fully unravel the introgression history. However, by dating introgression events and by analyzing the degree of shared introgression tracts among species, it is possible to distinguish different introgression scenarios. In this case, we were able to demonstrate that the previously claimed introgression from the Pachyceriforms into Bashibai, in fact, can be explained by ancient introgression from Pachyceriforms into argali and subsequent recent introgression from argali into Bashibai.

## Materials and Methods

### Genome Data Collection, Filtering, and SNP Calling

WGS data of domestic and wild sheep were retrieved from the National Center for Biotechnology Information database (<https://www.ncbi.nlm.nih.gov/>) and [Lv et al. \(2022\)](#), which included a total of 856 individuals ([supplementary table S2, Supplementary Material](#) online). These individuals represented 6 wild sheep species (8 urial, 6 argali, 13 snow sheep, 6 thinhorn sheep, 6 bighorn sheep, and 31 Asiatic mouflon) and a worldwide collection of domestic sheep (194 individuals from 24 African populations, 317 individuals from 70 Asian populations, 259 individuals from 69 European populations, and 16 individuals from 6 North American populations) ([Fig. 1](#); [supplementary table S2, Supplementary Material](#) online). The samples encompass nearly the entire geographic distribution of domestic sheep across 6 genetic groups, as well as all wild relatives, representing a broad range of genetic diversity, geographic regions, and phylogenetic lineages.

All the raw reads were trimmed to remove Illumina adapters, and the low-quality reads or bases were filtered out using Trimmomatic v0.36 ([Bolger et al. 2014](#)) with default parameters (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2: True LEADING:3 TRAILING:3 MINLEN:36). All the passed quality-filtered reads were mapped to the sheep reference genome assembly *Oar\_rambouillet\_v1.0* (GenBank accession no. GCF\_002742125.1) with the Burrows–Wheeler Aligner (BWA) tool using the BWA-MEM algorithm ([Li and Durbin 2009](#)). Duplicate reads were removed using the *MarkDuplicates* module in GATK v 4.1.2.0 ([McKenna et al. 2010](#)). Genome-wide SNPs and INDELS were called from the BAM files via the GATK *HaplotypeCaller* module following the GATK best-practice recommendations ([McKenna et al. 2010](#)). Raw SNPs were selected by the GATK *SelectVariants* module and filtered by the GATK *VariantFiltration* module following the filter expression “QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.” All the variants (SNPs) that passed the hard filter threshold were further annotated with SnpEff v4.3t ([Cingolani et al. 2012](#)).

### Inference of Ancestral Alleles

The ancestral allelic state for each SNP was inferred using the reference genome assemblies of goats (*ARS1*), cattle

(ARS-UCD1.2), and sheep (*Oar\_rambouillet\_v1.0*). The assemblies of ARS1 and ARS-UCD1.2 were split into continuous regions of 1,000 bp bins in the output format of *fastq* files, which was subsequently mapped to the sheep reference genome assembly *Oar\_rambouillet\_v1.0*. The above analyses were performed using LAST v983 (<http://last.cbrc.jp>, last accessed June 15, 2022). Then, we called SNPs using BCftools v1.9 (<http://samtools.github.io/bcftools>, last accessed July 5, 2022) with the default setting and merged the 2 sets of SNPs (the SNP data of 856 wild and domestic sheep and the SNP data of goats and cattle). The ancestral allele of each SNP was defined if the genotypes of the cattle and goats were homozygous and identical.

### Relatedness of Wild Species to Domestic Populations

We quantified the genetic drift shared between wild species and domestic populations using the outgroup  $f_3$ -statistics test (Patterson et al. 2012). We implemented outgroup  $f_3$ -statistics of the form  $f_3$  (ancestral allele; wild sheep, domestic population) using the ancestral allele as the outgroup, where the wild sheep were Asiatic mouflon, urial, argali, snow sheep, bighorn sheep, or thinhorn sheep. All the analyses were conducted using the program *qp3Pop* with the default parameters. Standard errors and Z-scores were computed using a weighted block jackknife over 5 Mb blocks (Patterson et al. 2012).

### Admixture Tests

We applied Patterson's  $D$ -statistics (Patterson et al. 2012) to test if some domestic sheep populations shared derived alleles with wild sheep by genetic introgression. For the 4 populations  $H_1$ ,  $H_2$ ,  $H_3$ , and  $H_4$ , we calculated  $D$ -statistics as

$$D = \frac{\sum_i (P_{iH_1} - P_{iH_2})(P_{iH_3} - P_{iH_4})}{\sum_i (P_{iH_1} + P_{iH_2} - 2P_{iH_1}P_{iH_2})(P_{iH_3} + P_{iH_4} - 2P_{iH_3}P_{iH_4})}$$

where  $P_{iH_j}$  is the allele frequency of derived allele in SNP  $i$  in population  $H_j$ , and the sum in the numerator and denominator is over all SNPs in the genome.

Under the null hypothesis corresponding to topology ( $H_1$ ,  $H_2$ ;  $H_3$ ,  $H_4$ ), the value of  $D$  is expected to be 0. To understand the degree of introgression between domestic sheep and wild sheep, we calculated  $D$ -statistics (non-admixed population,  $X$ ; wild sheep, ancestral alleles), in which the ancestral alleles and  $X$  represented the outgroup and each population/breed of domestic sheep, respectively. We also used Soay sheep and Menz sheep as proxies for "nonadmixed" domestic populations, as both breeds exhibit pure genetic characteristics and have not undergone hybridization or introgression with other sheep populations (Chessa et al. 2009; Kijas et al. 2012; Lv et al. 2022). Six  $D$ -statistics were calculated as  $D$  (nonadmixed population,  $X$ ; wild species, ancestral allele), in which the wild species represented one of the wild sheep species such as Asiatic mouflon, urial, argali, snow sheep, thinhorn sheep, or bighorn sheep.

All analyses were implemented using the program *qpDstat* with the default parameters (Patterson et al. 2012). We computed  $D$ -statistics using the allele frequency-based method, and the standard errors and Z-scores estimated with a weighted block jackknife procedure over 5 Mb blocks were applied to assess significance. Absolute values of Z-scores  $\geq 3$  indicated that the  $D$ -statistics significantly deviated from 0 and the null hypothesis was rejected.

### Ancient Introgression Analyses between Argali and the Common Ancestor of the Pachyceriforms Species

We used 3 methods to explore the relationship between argali and the 3 species of Pachyceriforms, including Phylogenetic weighting (Martin and Van Belleghem 2017),  $D$ -statistics (Patterson et al. 2012), and the DCT-BLT, which combines counts of loci supporting either discordant topology and the branch-length distributions of the gene trees supporting them, to test an introgression-free null model (Suvorov et al. 2022). We tested for introgression in both argali and the Pachyceriforms using 3 species such as domestic sheep, argali, and bighorn sheep and using goats as outgroup. We adopted the phylogenetic weighting method to evaluate alternative genealogical relationships caused by introgression or ILS between argali and Pachyceriforms. Autosomal genotypes for all the wild sheep samples, BSB130, BSB139, BSB140, and ancestral alleles, were extracted and filtered with the setting "FORMAT/DP < 10 | FORMAT/GQ < 30." The genotypes of all the above samples were phased using BEAGLE.r1399 (Browning et al. 2021) with default settings, except for using 10,000 windows and 1,000 overlapping steps. Neighbor-joining trees were inferred with sliding windows of 50 SNPs and the GTR model. The exact weights for 3 distinct evolutionary scenarios were computed for all the inferred genealogies using the program *Twisst* (<http://github.com/simonhmartin/twisst>, last accessed May 22, 2022). Equality of topological proportions was tested using a paired Mann-Whitney  $U$  test applied to the weights (proportions) of topologies inferred on each chromosome.

We also calculated  $D$ -statistics (Patterson et al. 2012) to explore introgression patterns between argali and the 3 species of Pachyceriforms as  $D$  (nonadmixed population, argali;  $X$ , ancestral allele), in which the ancestral allele represents the outgroup. Furthermore, we used the  $f_4$  ratio to quantify the genome-wide fraction of introgression (Patterson et al. 2012). We calculated the  $f_4$  ratio using Menz sheep or Soay sheep as  $P_1$ , argali as  $P_2$ , 1 of the Pachyceriforms species as  $P_3$ , and ancestral alleles as the outgroup.  $P_3$  was randomly divided into 2 subgroups,  $P_{3a}$  and  $P_{3b}$ , to provide a pair of admixed populations.

$$f_4 - \text{ratio} = \frac{D(P_1, P_2, P_3, O)}{D(P_1, P_{3a}, P_{3b}, O)}$$

We further used TreeMix v.1.12 program (Pickrell and Pritchard 2012) to investigate the introgression between

the Pachyceriforms and argali. ML trees were first constructed using blocks of 1,000 SNPs. Then, we inferred trees with mixture events ( $M$ ) ranging from 1 to 6 and quantified the residual fit of each mixture event based on covariance (Pickrell and Pritchard 2012).

Additionally, we used the DCT-BLT approach, another gene-tree-based method (Suvorov et al. 2022), to identify phylogenetic asymmetry as an evidence of introgression from annotated genes. For the 5 species of domestic sheep, argali, bighorn sheep, snow sheep, and goats, single-copy orthologous genes were generated from the reference genome assemblies *GCF\_002742125.1*, *GCA\_003121645.1*, *GCA\_004026945.1*, *GCA\_903231385.1*, and *GCF\_001704415.1* using the program BUSCO v 4.1.4 (Manni et al. 2021). The single-copy orthologous gene cluster was aligned using the program MAFFT v7.490 (Kuraku et al. 2013), and multiple sequence alignments (MSAs) were trimmed using the program trimal 1.2.rev59 (Capella-Gutiérrez et al. 2009) under an automated setting and concatenated into a Supermatrix using the python script `geneSticher.py` (<https://github.com/ballesterus/Utensils/blob/master>, last accessed May 21, 2023). We inferred the ML trees for each gene and species using IQTREE v1.6.5 (Nguyen et al. 2015) with the setting of 1,000 ultrafast bootstrap replicates (Hoang et al. 2018) and the GTR + I + G model (Tavaré 1986). We performed a DCT and BLT test on all the triplets (domestic sheep, argali, and bighorn sheep) and (domestic sheep, argali, and snow sheep) extracted from the gene trees. We used a  $\chi^2$  goodness-of-fit test to test for deviations from equal frequencies of gene tree counts among discordant gene trees and a Mann–Whitney  $U$  test to compare the distribution of  $d_{T1}$  and  $d_{T2}$ .  $d_{T1}$  and  $d_{T2}$  are genetic distances in the first and second discordant topology, which are calculated by averaging the external branch lengths leading to the 2 sister taxa under the topologies ([bighorn sheep, domestic sheep], argali) and ([argali, bighorn sheep], domestic sheep) and the topologies ([snow sheep, domestic sheep], argali) and ([argali, snow sheep], domestic sheep). The  $P$ -values were subsequently corrected for multiple testing using the Benjamini–Hochberg procedure with a false discovery rate cutoff of 0.05.

### Evolutionary Simulations of Ancestral Introgression between Argali and the Common Ancestor of Pachyceriforms

We reconstructed alternative demographic models of the *Ovis* genus and used fastsimcoal v2.6 (Excoffier et al. 2013) to assess the fit of alternative models by approximating the likelihood with coalescent simulations (Nielsen 2000; Adams and Hudson 2004). The method can handle large WGS data set and is well suited to the study of complex scenarios (Gravel et al. 2011; Tennessen et al. 2012; Excoffier et al. 2013; Sousa and Hey 2013). Forty-six samples (average depth = 17.14 $\times$ ) from Bashibai sheep ( $n = 3$ ), Asiatic mouflon ( $n = 10$ ), urial ( $n = 8$ ), argali ( $n = 6$ ), snow sheep ( $n = 7$ ), thinhorn sheep ( $n = 6$ ), and bighorn sheep ( $n = 6$ ) were selected. A total of 822,229 high-quality

autosomal SNPs were kept following the 5 criteria as follows: (i) within intergenic regions, (ii) outside the CpG islands as defined by the UCSC annotation ([https://hgdownload.soe.ucsc.edu/hubs/GCF/002/742/125/GCF\\_002742125.1/bbi/](https://hgdownload.soe.ucsc.edu/hubs/GCF/002/742/125/GCF_002742125.1/bbi/), last accessed July 20, 2022), (iii) non-missing genotypes in all the samples, (iv) depth > 20 $\times$  in all the samples, and (v) ancestral state of the SNPs. The multidimensional unfold-SFS file was generated using the Python script `vcf2sfs.py` (<https://github.com/marqueda/SFS-scripts>, last accessed March 7, 2022).

We aimed to investigate if there was ancestral introgression or hybridization between argali and the common ancestor of Pachyceriforms. We proposed 3 alternative models, i.e. hypotheses 1, 2, and 3 (Fig. 4d). To compare the alternative models, we estimated parameters for models I, II, and III within certain bounds as described in [supplementary table S18, Supplementary Material](#) online, and obtained the maximum likelihood value for each model. We calculated the maximum likelihood of each model with all the entries of SFS for the first 25 cycles (-l 25) and then optimized the maximum likelihood of each model for the remaining 40 cycles (-L 65). We performed 1,000,000 coalescent simulations and 100 optimization runs starting from random initial conditions to obtain the highest likelihood for a given set of parameters. All the analyses were implemented using fastsimcoal v2.6 with a constant mutation rate of 1.51e-8/gen/site (Chen et al. 2019) and a generation time of 3 years (Zhao et al. 2017).

After identifying the optimal model for the evolutionary history of *Ovis*, we estimated species divergence or hybridization times between argali and Pachyceriforms under the MSci (Flouri et al. 2020). We selected the single-copy orthologous genes identified using the program BUSCO v 4.1.4 (Manni et al. 2021) and filtered the data set via the pipeline described in the DCT and BLT analyses. In total, 2,614 single-copy orthologous genes among species of domestic sheep, argali, and bighorn and 3,140 single-copy orthologous genes among species of domestic sheep, argali, and snow sheep with more than 1,000 bp were used in the MSci analyses. The analyses were implemented in the program BP&P v 4.6.2 (Flouri et al. 2020) under the MSci model with an automatic adjustment of fine-tune parameters. We used module A00 to infer the species divergence or hybridization times ( $\tau$ ), the population size parameter ( $\theta$ ), and the introgression probability ( $\phi$ ). Three independent replicates were run through 1,200,000 generations each and sampled every 2 generations, with a burn-in of 10,000 generations. We used TRACER version 1.7.2 (Rambaut et al. 2018) to combine the results of the 3 independent runs and assess the convergence (effective sample size > 200). The species divergence or hybridization time was calibrated based on the most recent common ancestor for *Ovis*'s species, which had uniform value of 3.2 to 5.4 Mya (Wang et al. 2016).

### Testing Genomic Patterns of Recent Introgression from Argali to Bashibai Sheep

We first used a model-based least-squares optimization algorithm to infer individual admixture coefficients (Frichot



et al. 2014). The data set, which included all the wild sheep samples, BSB130, BSB139, BSB140, and the inferred ancestor, was generated and filtered with an MAF < 0.05, Hardy–Weinberg equilibrium < 0.001, and a proportion of missing genotypes > 10%. We pruned the data set with the PLINK option “–indep-pairwise 50 5 0.2.” Finally, 4,781,407 SNPs were retained to calculate the admixture coefficients. We used the program sNMF (Frichot et al. 2014) to infer individual admixture coefficients with the parameter  $K$  ranging from 2 to 15. The best run was determined by the smallest value of the cross-entropy criterion.

The  $f_d$ -statistic was used to quantify the genome-wide fraction of introgression (Martin et al. 2015) and detect introgressed tracts. We calculated the  $f_d$  using the populations of Menz sheep and Soay sheep as  $P_1$ ; the individuals of BSB130, BSB139, and BSB140 as  $P_2$ ; the argali as  $P_3$ ; and the ancestral alleles as the outgroup. We performed  $f_d$  analyses using the python script ABBABABAwindows.py ([https://github.com/simonhmartin/genomics\\_general/blob/master/](https://github.com/simonhmartin/genomics_general/blob/master/), last accessed September 20, 2022) with 25 kb windows across chromosomes and a minimum of 250 SNPs within each window. Proportion of ancestor inferred by sNMF was used to identified threshold of  $f_d$  values. The genomic regions with  $f_d$  values greater than 0.3532, 0.3679, and 0.3632, corresponding to 15.51%, 6.46%, and 9.98% genome proportions of argali, were determined to be the introgressed regions for samples BSB130, BSB139, and BSB140, respectively. We annotated the introgressed regions using the sheep reference assembly *Oar\_rambouillet\_v1.0* in snpEff v4.3t (Cingolani et al. 2012) and compared nucleotide diversity dN/dS ratios and GC content between the introgressed regions and nonintrogressed regions. Additionally, we assessed the presence of transposons elements (TEs) in these regions using the RepeatMasker database ([https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/742/125/GCF\\_002742125.1\\_Oar\\_rambouillet\\_v1.0/](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/742/125/GCF_002742125.1_Oar_rambouillet_v1.0/)). A summary of all software used in this study is provided in [supplementary table S19, Supplementary Material](#) online.

## Supplementary Material

[Supplementary material](#) is available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank the High-Performance Computing Platform of China Agricultural University for providing computing resources.

## Author Contributions

F.-H.L. designed the study with helpful discussions with M.-H.L. and R.N. F.-H.L. and D.-F.W. performed the data analyses. S.-Y.Z. provided help with the data collection. D.-F.W., X.-Y.L., and W.S. provided help with coding and discussion. F.-H.L. and M.-H.L. wrote the manuscript. R.N.

and M.-H.L. revised the paper. All the authors reviewed and approved the final version of the manuscript.

## Funding

This study was financially supported by grants from the National Natural Science Foundation of China (nos. U21A20246, 31972527, and 32061133010), the Open Project Program of International Joint Research Laboratory in Universities of Jiangsu Province of China for Domestic Animal Germplasm Resources and Genetic Improvement, and the Chinese Universities Scientific Fund (2024TC162).

## Conflict of Interest

The authors declare no competing interests.

## Data Availability

The whole genome re-sequence data used for the study is publicly available under the sample accession numbers listed in supplementary table S2.

## References

- Adams AM, Hudson RR. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics*. 2004;**168**(3): 1699–1712. <https://doi.org/10.1534/genetics.104.030171>
- Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet*. 2015;**47**(3):217–225. <https://doi.org/10.1038/ng.3199>
- Batchelor CL, Margold M, Krapp M, Murton DK, Dalton AS, Gibbard PL, Stokes CR, Murton JB, Manica A. The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nat Commun*. 2019;**10**(1):3713. <https://doi.org/10.1038/s41467-019-11601-2>
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet*. 2021;**108**(10): 1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>
- Cao Y-H, Xu S-S, Shen M, Chen Z-H, Gao L, Lv F-H, Xie X-L, Wang X-H, Yang H, Liu C-B, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol*. 2021;**38**(3):838–855. <https://doi.org/10.1093/molbev/msaa236>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;**25**(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- 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. <https://doi.org/10.1126/science.aav6202>
- Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, et al. Whole-genome resequencing reveals worldwide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun*. 2018;**9**(1):2337. <https://doi.org/10.1038/s41467-018-04737-0>

- Chen Z-H, Xu Y-X, Xie X-L, Wang D-F, Aguilar-Gómez D, Liu G-J, Li X, Esmailzadeh A, Rezaei V, Kantanen J, et al. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol*. 2021;**4**(1):1307. <https://doi.org/10.1038/s42003-021-02817-4>
- Cheng H, Zhang Z, Wen J, Lenstra JA, Heller R, Cai Y, Guo Y, Li M, Li R, Li W, et al. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. *PLoS Genet*. 2023;**19**(2):e1010615. <https://doi.org/10.1371/journal.pgen.1010615>
- Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, et al. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science*. **324**:532–536. <https://doi.org/10.1126/science.1170587>
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;**6**(2):80–92. <https://doi.org/10.4161/fly.19695>
- Colledge S, Conolly J, Shennan S. The evolution of Neolithic farming from SW Asian origins to NW European limits. *Eur J Archaeol*. 2005;**8**(2):137–156. <https://doi.org/10.1177/1461957105066937>
- Cowan IM. Distribution and variation in the native sheep of North America. *Am Midl Nat*. 1940;**24**(3):505–580. <https://doi.org/10.2307/2420858>
- Deng J, Xie XL, Wang DF, Zhao C, Lv FH, Li X, Yang J, Yu JL, Shen M, Gao L, et al. Paternal origins and migratory episodes of domestic sheep. *Curr Biol*. 2020;**30**(20):4085–4095. <https://doi.org/10.1016/j.cub.2020.07.077>
- Dunmire WW. New Mexico's Spanish livestock heritage: four centuries of animals, land, and people. Albuquerque: UNM Press; 2013.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013;**9**(10):e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Flouri T, Jiao XY, Rannala B, Yang ZH. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Mol Biol Evol*. 2020;**37**(4):1211–1223. <https://doi.org/10.1093/molbev/msz296>
- Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, Semiadi G, Meijaard E, Li N, Crooijmans RPMA, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol*. 2013;**14**(9):R107. <https://doi.org/10.1186/gb-2013-14-9-r107>
- Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014;**196**(4):973–983. <https://doi.org/10.1534/genetics.113.160572>
- Frick C. Horned ruminants of North America. *Bull Am Mus Nat Hist*. 1937;**69**:1883–1965.
- Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu FL, Gibbs RA, Bustamante CD, Project G. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci USA*. 2011;**108**(29):11983–11988. <https://doi.org/10.1073/pnas.1019276108>
- Guthrie RD. Paleoeology of a Late Pleistocene small mammal community from Interior Alaska. *Arctic*. 1968;**21**(4):223–244. <https://doi.org/10.14430/arctic3266>
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBboot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;**35**(2):518–522. <https://doi.org/10.1093/molbev/msx281>
- Hu A, Meehl GA, Otto-Bliesner BL, Waelbroeck C, Han W, Loutre M-F, Lambeck K, Mitrovica JX, Rosenbloom N. Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nat Geosci*. 2010;**3**(2):118–121. <https://doi.org/10.1038/ngeo729>
- Hu X-J, Yang J, Xie X-L, Lv F-H, Cao Y-H, Li W-R, Liu M-J, Wang Y-T, Li J-Q, Liu Y-G, et al. The genome landscape of Tibetan sheep reveals adaptive introgression from argali and the history of early human settlements on the Qinghai-Tibetan plateau. *Mol Biol Evol*. 2019;**36**(2):283–303. <https://doi.org/10.1093/molbev/msy208>
- Jakobsson M, Pearce C, Cronin TM, Backman J, Anderson LG, Barrientos N, Björk G, Coxall H, de Boer A, Mayer LA, et al. Post-glacial flooding of the Bering Land Bridge dated to 11 cal ka BP based on new geophysical and sediment records. *Clim Past*. 2017;**13**(8):991–1005. <https://doi.org/10.5194/cp-13-991-2017>
- Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*. 2018;**360**(6395):1355–1358. <https://doi.org/10.1126/science.aar5273>
- Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, et al. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. **10**:e1001258. <https://doi.org/10.1371/journal.pbio.1001258>
- Kuraku S, Zmasek CM, Nishimura O, Katoh K. Aleaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res*. 2013;**41**(W1):W22–W28. <https://doi.org/10.1093/nar/gkt389>
- Lan T, Leppälä K, Tomlin C, Talbot SL, Sage GK, Farley SD, Shideler RT, Bachmann L, Wiig Ø, Albert VA, et al. Insights into bear evolution from a Pleistocene polar bear genome. *Proc Natl Acad Sci USA*. 2022;**119**:e2200016119. <https://doi.org/10.1073/pnas.2200016119>
- Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, Aljumaah RS, Mwacharo JM, Wu D-D, Zhang Y-P, et al. The wild species genome ancestry of domestic chickens. *BMC Biol*. 2020;**18**(1):13. <https://doi.org/10.1186/s12915-020-0738-1>
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;**25**(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li X, He S-G, Li W-R, Luo L-Y, Yan Z, Mo D-X, Wan X, Lv F-H, Yang J, Xu Y-X, et al. Genomic analyses of Pamir argali, Tibetan sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation. *Genome Res*. 2022;**32**(9):1669–1684. <https://doi.org/10.1101/gr.276769.122>
- Liu L, Bosse M, Megens H-J, Frantz LAF, Lee Y-L, Irving-Pease EK, Narayan G, Groenen MAM, Madsen O. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat Commun*. 2019;**10**(1):1992. <https://doi.org/10.1038/s41467-019-10017-2>
- Lv F-H, Cao Y-H, Liu G-J, Luo L-Y, Lu R, Liu M-J, Li W-R, Zhou P, Wang X-H, Shen M, et al. Whole-Genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol*. 2022;**39**(2):msab353. <https://doi.org/10.1093/molbev/msab353>
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc*. 2021;**1**(12):e323. <https://doi.org/10.1002/cpz1.323>
- Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol*. 2015;**32**(1):244–257. <https://doi.org/10.1093/molbev/msu269>
- Martin SH, Van Belleghem SM. Exploring evolutionary relationships across the genome using topology weighting. *Genetics*. 2017;**206**(1):429–438. <https://doi.org/10.1534/genetics.116.194720>
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;**20**(9):1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Nadler CF, Korobitsina KV, Hoffmann RS, Vorontsov NN. Cytogenetic differentiation, geographic distribution, and domestication in Palearctic sheep (*Ovis*). *Z Saugtierkd*. 1973;**38**:109–125.
- Nelson TC, Stathos AM, Vanderpool DD, Finseth FR, Yuan Y-W, Fishman L. Ancient and recent introgression shape the



- evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet.* 2021;**17**(2):e1009095. <https://doi.org/10.1371/journal.pgen.1009095>
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;**32**(1):268–274. <https://doi.org/10.1093/molbev/msu300>
- Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics.* 2000;**154**(2):931–942. <https://doi.org/10.1093/genetics/154.2.931>
- Osborn HF. The age of mammals in Europe, Asia and North America. New York: The Macmillan Company; 1912.
- Ozioloz EM, Reid NM, Yair S, Lee KM, Guberman Verploeg S, Bruns PC, Shaw JR, Whitehead A, Matson CW. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science.* 2019;**364**(6439):455–457. <https://doi.org/10.1126/science.aav4155>
- Pacheco C, Stronen AV, Jędrzejewska B, Plis K, Okhlopkov IM, Mamaev NV, Drovetski SV, Godinho R. Demography and evolutionary history of grey wolf populations around the Bering Strait. *Mol Ecol.* 2022;**31**(18):4851–4865. <https://doi.org/10.1111/mec.16613>
- Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. *Genetics.* 2012;**192**(3):1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Petr M, Pääbo S, Kelso J, Vernot B. Limits of long-term selection against Neandertal introgression. *Proc Natl Acad Sci USA.* 2019;**116**(5):1639–1644. <https://doi.org/10.1073/pnas.1814338116>
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;**8**(11):e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol.* 2018;**67**(5):901–904. <https://doi.org/10.1093/sysbio/syy032>
- Redford KH, Dudley N. Why should we save the wild relatives of domesticated animals? *Oryx.* 2018;**52**(3):397–398. <https://doi.org/10.1017/S0030605318000601>
- Rezaei HR, Naderi S, Chintauan-Marquier IC, Taberlet P, Virk AT, Naghash HR, Rioux D, Kaboli M, Pompanon F. Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Mol Phylogenet Evol.* 2010;**54**(2):315–326. <https://doi.org/10.1016/j.ympev.2009.10.037>
- Salis AT, Bray SCE, Lee MSY, Heiniger H, Barnett R, Burns JA, Doronichev V, Fedje D, Golovanova L, Harington CR, et al. Lions and brown bears colonized North America in multiple synchronous waves of dispersal across the Bering Land Bridge. *Mol Ecol.* 2022;**31**(24):6407–6421. <https://doi.org/10.1111/mec.16267>
- Sánchez-Villagra MR. The process of animal domestication. New Jersey: Princeton University Press; 2022.
- Santos SHD, Peery RM, Miller JM, Dao A, Lyu F-H, Li X, Li M-H, Coltman DW. Ancient hybridization patterns between bighorn and thinhorn sheep. *Mol Ecol.* 2021;**30**(23):6273–6288. <https://doi.org/10.1111/mec.16136>
- Sousa V, Hey J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet.* 2013;**14**(6):404–414. <https://doi.org/10.1038/nrg3446>
- Spangler GL, Rosen BD, Ilori MB, Hanotte O, Kim ES, Sonstegard TS, Burke JM, Morgan JLM, Notter DR, Van Tassell CP. Whole genome structural analysis of Caribbean hair sheep reveals quantitative link to West African ancestry. *PLoS One.* 2017;**12**(6):e0179021. <https://doi.org/10.1371/journal.pone.0179021>
- Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol.* 2022;**32**(1):111–123. <https://doi.org/10.1016/j.cub.2021.10.052>
- Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci.* 57–86.
- Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science.* 2012;**337**(6090):64–69. <https://doi.org/10.1126/science.1219240>
- Upadhyay M, Kunz E, Sandoval-Castellanos E, Hauser A, Krebs S, Graf A, Blum H, Dotsev A, Okhlopkov I, Shakhin A, et al. Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Mol Ecol.* 2021;**30**(24):6701–6717. <https://doi.org/10.1111/mec.16184>
- Valdez R, Nadler CF, Bunch TD. Evolution of wild sheep in Iran. *Evolution.* 1978;**32**(1):56–72. <https://doi.org/10.2307/2407410>
- Vershinina AO, Heintzman PD, Froese DG, Zazula G, Cassatt-Johnstone M, Dalén L, Der Sarkissian C, Dunn SG, Ermini L, Gamba C, et al. Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. *Mol Ecol.* 2021;**30**(23):6144–6161. <https://doi.org/10.1111/mec.15977>
- Wang X, Li Q, Takeuchi GT. Out of Tibet: an early sheep from the Pliocene of Tibet, *Protovis himalayensis*, genus and species nov. (Bovidae, Caprini), and origin of Ice Age mountain sheep. *J Vertebr Paleontol.* 2016;**36**(5):e1169190. <https://doi.org/10.1080/02724634.2016.1169190>
- Wu DD, Ding XD, Wang S, Wojcik JM, Zhang Y, Tokarska M, Li Y, Wang MS, Faruque O, Nielsen R, et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol.* 2018;**2**(7):1139–1145. <https://doi.org/10.1038/s41559-018-0562-y>
- Yu H, Xing Y-T, Meng H, He B, Li W-J, Qi X-Z, Zhao J-Y, Zhuang Y, Xu X, Yamaguchi N, et al. Genomic evidence for the Chinese mountain cat as a wildcat conspecific (*Felis silvestris bieti*) and its introgression to domestic cats. *Sci Adv.* 2021;**7**(26):eabg0221. <https://doi.org/10.1126/sciadv.abg0221>
- Zeder MA. Domestication and early agriculture in the Mediterranean basin: origins, diffusion, and impact. *Proc Natl Acad Sci USA.* 2008;**105**(33):11597–11604. <https://doi.org/10.1073/pnas.0801317105>
- Zhao Y-X, Yang J, Lv F-H, Hu X-J, Xie X-L, Zhang M, Li W-R, Liu M-J, Wang Y-T, Li J-Q, et al. Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in East Asia. *Mol Biol Evol.* 2017;**34**(9):2380–2395. <https://doi.org/10.1093/molbev/msx181>
- Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, Pan X, Gong M, Zhang Y, Guo Y, et al. The origin of domestication genes in goats. *Sci Adv.* 2020;**6**(21):eaz5216. <https://doi.org/10.1126/sciadv.aaz5216>
- Zhou Y, Zhao X-B, Li Y-W, Xu J, Bi A-Y, Kang L-P, Xu D-X, Chen H-F, Wang Y, Wang Y-G, et al. Triticum population sequencing provides insights into wheat adaptation. *Nat Genet.* 2020;**52**(12):1412–1422. <https://doi.org/10.1038/s41588-020-00722-w>
- Zou T, Kuang W, Yin T, Frantz L, Zhang C, Liu J, Wu H, Yu L. Uncovering the enigmatic evolution of bears in greater depth: the hybrid origin of the Asiatic black bear. *Proc Natl Acad Sci USA.* 2022;**119**(31):e2120307119. <https://doi.org/10.1073/pnas.2120307119>