Conserv Genet (2011) 12:921–931
DOI 10.1007/s10592-011-0195-z

RESEARCH ARTICLE

High connectivity among argali sheep from Afghanistan and adjacent countries: Inferences from neutral and candidate gene microsatellites

G. Luikart · S. J. Amish · J. Winnie · A. Beja-Pereira · R. Godinho · F. W. Allendorf · R. B. Harris

Received: 1 July 2010 / Accepted: 28 January 2011 / Published online: 6 April 2011

Abstract We quantified population connectivity and genetic variation in the Marco Polo subspecies of argali mountain sheep (*Ovis ammon polii*) by genotyping 9 neutral and 8 candidate gene microsatellite loci in 172 individuals noninvasively sampled across five study areas in Afghanistan, China, and Tajikistan. Heterozygosity and allelic richness were generally high (mean H = 0.67, mean A = 6.1), but were significantly lower in the China study area (H = 0.61, $P < 0.001$; A = 4.9, $P < 0.01$). One marker in an immune system gene (*TCRG4*) showed an excess of rare alleles compared to neutral expectations. Another immune system gene (*GLYCAM-1*) showed excessive differentiation (high F_{ST}) between study areas. Estimates of genetic differentiation were similar ($F_{ST} = 0.035$ vs. 0.033) with and without the two loci deviating from neutrality, suggesting that selection is not a primary driver of overall molecular variation, and that candidate gene loci can be used for connectivity monitoring, as long as selection tests are conducted to avoid biased gene flow estimates. Adequate protection of argali and maintenance of inter-population connectivity will require monitoring and international cooperation because argali exhibit high gene flow across international borders.

Keywords Bottlenecks · Habitat fragmentation · Gene flow · *Ovis ammon* · Pamir Mountains · Natural selection · Adaptation · Infectious disease · Noninvasive genetic monitoring · Mountain ungulate

G. Luikart (✉)
Fish and Wildlife Genomics Group, Flathead Lake Biological Station, University of Montana, Polson, MT 59860, USA
e-mail: gordon.luikart@mso.umt.edu

G. Luikart · S. J. Amish · F. W. Allendorf
Fish and Wildlife Genomics Group, Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

G. Luikart · A. Beja-Pereira · R. Godinho
CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

J. Winnie · R. B. Harris
Department of Ecosystem and Conservation Science, University of Montana, Missoula, MT 59812, USA

Introduction

Genetic assessments and monitoring are increasingly crucial for delineating population boundaries and movement corridors, and possibly for understanding adaptation to environmental change in extreme environments (Shackleton 1997; Schwartz et al. 2007). Availability of candidate adaptive gene markers (e.g. Kohn et al. 2006) along with neutral loci could make feasible the assessment of both adaptive and neutral connectivity, i.e. gene flow and adaptation to changing environments (Black et al. 2001).

Argali (*Ovis ammon*) are an ecologically and economically important species, but are increasingly threatened throughout their range. The Marco Polo subspecies of argali (*Ovis ammon polii*) is among the largest wild sheep and is perhaps the most charismatic wild animal in the Pamir Mountains of Tajikistan, China, Kyrgyzstan, and Afghanistan (Fedosenko and Blank 2005). The Marco Polo subspecies is categorized as Near Threatened on the IUCN Red List (IUCN 2009). Argali in the Pamir Mountains are important because of their role as a flagship species for the entire ecosystem (Schaller and Kang 2008). However, populations are susceptible to human pressures including poaching, displacement, competition, and disease

transmission from livestock, as well as possible habitat fragmentation (Shackleton 1997). Due to their innate habitat preferences, argali are generally distributed patchily, with areas of inappropriate habitat separating populations. Unlike the related mountain sheep (e.g., *O. canadensis*) of North America, argali are generally believed to be willing to traverse long distances, possibly across seemingly inhospitable terrain. Thus the degree to which populations are truly fragmented, either demographically or genetically, is often an open question.

Unfortunately, argali are among the most difficult of wild ungulates to study due to their wary nature, choice of remote and precipitous habitats, and low population density. Argali, unlike North American wild sheep, are cursorial and will move long-distances to escape predators or disturbance. Little is known about argali movements or migrations because they are difficult to capture, tag or collar, and track. In part because they are intolerant of human disturbance, argali typically live only where human access is difficult or infrequent.

Marco Polo argali in the Wakhan Corridor of Afghanistan are found only in a small section of the Big Pamir Mountains, in the eastern Little Pamirs, and in the Wakhjir Valley; thus, possible isolation among populations is a legitimate concern (Fig. 1). In contrast, it is likely that populations in Tajikistan are more contiguous in nature (Weinberg et al. 1997). The exact status of Marco Polo argali populations in China is uncertain, although they are known to exist in relatively high numbers in most of the Taxkorgan Nature Reserve in Kashi Prefecture, Xinjiang (Schaller et al. 1987; Gong et al. 2007; Schaller and Kang 2008).

Noninvasive and remote genetic sampling (Taberlet et al. 1999; Beja-Pereira et al. 2009) facilitates research on elusive species such as argali. Molecular genetic markers and metapopulation models allow assessment of migration rates over the recent past (dozens to hundreds of generations, assuming migration-drift equilibrium), and also current migration rates by identifying actual immigrants, e.g. using individual-based assignment tests (Cornuet et al. 1999; Paetkau et al. 2004) or by quantifying immigrant ancestry (Wilson and Rannala 2003). Molecular markers also allow detection of recent population size reductions or bottlenecks (e.g., Cornuet and Luikart 1996).

Detecting loci under selection is important because selection can bias estimates of population genetic parameters, e.g. F_{ST} (Luikart et al. 2003). Detecting selection signatures also can help infer if a population has experienced a recent selection or stress event such as environmental change or disease die-off, which could help infer the cause of population declines. For example, Simões et al. (2008) studied the genetic response to selection and detected both a reduced effective population size (increased drift

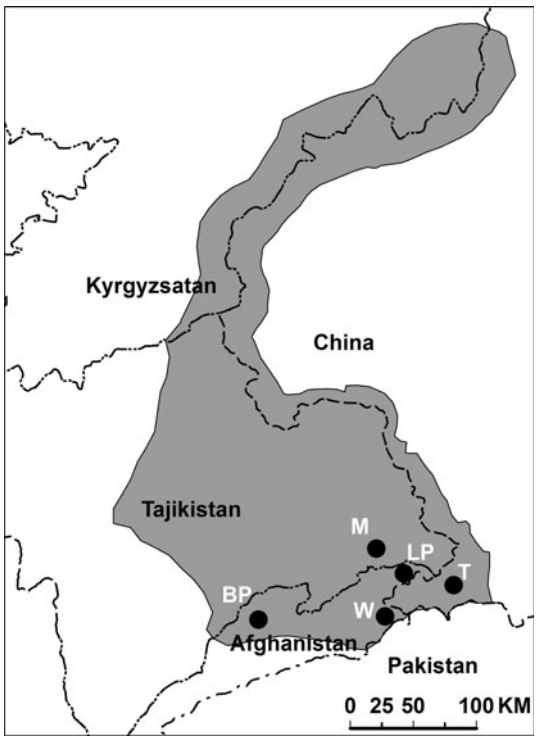

Fig. 1 Map of approximate distribution of the Marco Polo argali subspecies, *O. a. polli* (modified from Schaller and Kang 2008) showing national boundaries and the five study areas in the Pamir Mountains. The large *grey* area is only a coarse range depiction in that argali populations are not distributed continuously throughout the entire *shaded* area. *Black filled circles* show approximate locations of our Murghab (*M*) study area in southern Tajikistan, the Taxkorgan (*T*) study area in Xinjiang, China, three study areas in Afghanistan: the Big Pamir (*BP*), Little Pamir (*LP*), Wakhjir (Waghjir) Valley (*W*)

at multiple microsatellite loci) and directional selection (F_{ST}-outlier effects at a single microsatellite locus) during the adaptation to a new environment in populations of *Drosophila*. The authors suggest that selection at a single locus was associated with adaptive challenges that increased mortality, contributing to genome-wide drift and reduced effective population size.

Selection can be detected as extremely high (or low) genetic differentiation (F_{ST}) between populations at a single locus compared to neutral loci. Researchers have developed "F_{ST} outlier" tests (Beaumont and Nichols 1996; Antão et al. 2008) and shown that they have reasonable power (Beaumont and Balding 2004) to detect directional selection between populations. Selection can also be detected using neutrality tests within populations (Watterson 1978). For example, Paterson (1998) detected even allele frequencies at an *MHC* locus (Major Histocompatibility Complex includes more than 100 genes) in a population of Soay sheep (*Ovis aries*). Microsatellites in genes affected by selection will undergo genetic hitchhiking and show the selection signature of the gene under

selection. This approach has been implemented in large-scale genome scans of thousands of loci to identify genes or genome regions under selection (Payseur et al. 2002; Vasemägi et al. 2005).

We genotyped neutral and candidate adaptive gene microsatellites from fecal DNA sampled from five study areas within three countries with resident Marco Polo argali populations (Afghanistan, China, and Tajikistan) to assess the genetic and demographic status of argali across the region. Our specific objectives were to (1) quantify connectivity of argali among countries and among the three remaining areas with argali in Afghanistan, (2) test for reduced variation and bottleneck signatures within study areas, and (3) test for signatures of selection at immune system genes that might result from adaptive differentiation or stress events such as disease die-offs.

Methods

Study area

Although geographers have not agreed on precise boundaries, the Pamir Mountains are generally viewed as constituting the eastern portion of Tajikistan, the northern half of the Wakhan Corridor in Afghanistan, and the southwest corner of Xinjiang, China. This arid (annual precipitation typically $\sim$130 mm/year) and high elevation (highest peaks >7,000 m) range is typified by broad valleys and relatively gentle slopes ("pamir" refers to broad, grassy plateau-like topography).

We sampled from argali in five different locations within the Pamir Mountains; three within Wakhan District, Badakhshan Province, Islamic Republic of Afghanistan, one within Gorno-Bakakhshan, Tajikistan, and one within Taxkorgan County, Kashi Prefecture, Xinjiang Autonomous Region, People's Republic of China. We also collected samples from the Karichinai Valley in Khunjerab National Park, Pakistan, but because none yielded DNA these are not discussed further. We termed our five study areas (Fig. 1) the Big Pamir Mountains, the Little Pamir Mountains, and the Wakhjir Valley (all within Afghanistan), the Murghab study area (in southern Gorno-Badakhshan Province, Tajikistan), and the Taxkorgan study area (in Xinjiang, China). All five study areas were typified by rolling hills and rugged mountains at elevations of 3,900–5,300 m, vegetated by arid, steppe vegetation communities of grasses and low-lying shrubs.

The Big Pamir Mountains extend from approximately 37°N to 37°20′N latitude and 72°45′E to 73°30′E longitude, and are bordered on the south by the Wakhan River and on the north by the Panj (Amu Darya) River, which also forms the border between Afghanistan and Tajikistan. Within the Big Pamirs, we sampled only from approximately 1000 km^2 centered on 37°N latitude and 73°E longitude. The Little Pamir Mountains, located approximately 150 km east of the Big Pamir Mountains, are centered on 37°23′N latitude and 74°20′E longitude. The Wakhjir Valley, located approximately 37°03′N latitude and 74°30′E longitude, forms a small spur of Afghanistan that separates Xinjiang, China from Gilgit-Baltistan of Pakistan. Our Murghab study area in south-eastern Tajikistan extended along an east–west line of approximately 45 km located about 10–30 km north of the Afghan border. Our Taxkorgan study area was located within the Taxkorgan Nature Reserve in Taxkorgan County, Kashi Prefecture in Xinjiang, and was centered on approximately 37°23′N latitude and 75°20′E longitude.

Geographic distance between study areas ranged from a minimum of $\sim$15 km between Little Pamirs and Murghab to a maximum of 212 km between Taxkorgan and the Big Pamirs (Fig. 1). Distance within Afghanistan ranged from 38 km between the Wakhjir Valley and the Little Pamirs to 164 km between the Big Pamirs and the Little Pamirs.

Sampling

All field work was conducted either on foot, horseback, or "yak-back". Because argali move frequently through difficult terrain and our own movements were circumscribed by the valley systems separated by steep ridges, we made no attempt to impose a standardized geographic sampling regime. Instead, we attempted to survey for argali by walking to high vantage points to search for animals during early morning and late afternoon time periods.

Whenever we encountered fecal pellets we were certain had been freshly deposited by argali, we collected three fecal pellets from each pellet group (i.e. pellet pile). We only collected pellets adjacent to each other within the group, reducing to inconsequential the probability of >1 individual argali being represented within each individual sample. We avoided collecting from pellet groups that were scattered over more than approximately a 0.1 m^2 area, or that appeared to have been deposited while the animal was moving. We took GPS locations for each sample (unless samples were within a few paces of an existing GPS fix, in which case we recorded the same location), and noted the date, time, and name of the collector.

Fecal pellets were stored in sterile 30 cm centrifuge tubes with securely fitting screw-tops to which internal "sporks" were attached (which allowed individual handling of each sample without risk of contamination; Evergreen Scientific, Los Angeles, CA, USA). We placed three fecal pellets into approximately six parts of 95% ethyl alcohol (ETOH) for each part fecal material, and stored them at room temperature for 1–4 months before

extraction. We collected and extracted DNA from one pellet from each of 240 pellet groups.

DNA extraction, genotyping, and sexing

Genetic analyses were conducted in two laboratories. Initial work was undertaken at CTM/CIBIO (Centro de Testagem Molecular), Portugal, where fecal samples were extracted and eight microsatellite markers were co-amplified in three multiplex PCR reactions as described in Harris et al. (2010). All individual fecal samples were initially genotyped twice to quantify the quality of the nuclear DNA for producing genotype data. Samples with reliable amplifications (electropherogram peak height > 50 units and identical genotypes from the two replicate genotypings at 8 loci) in this first step were selected to continue the genotyping process. The samples with reliable amplifications were independently re-genotyped three to six times total for each of the 8 loci.

The remaining genetic analyses were conducted at the University of Montana Conservation Genetics Laboratory (MCGL), Missoula, Montana, USA. Six microsatellite loci (MAF36, FCB304, FCB266, ADC, MAF33, KRT2) were genotyped in both labs on a large subset of samples as a data quality check. Eleven additional loci were genotyped the MGCL for a total of 17 loci. Yet another locus (*MAF226*) was genotyped at MCGL in all five populations but was excluded due to allelic dropout and a strong deviation from Hardy–Weinberg proportions (mean $F_{IS} = 0.43$). All loci genotyped at MCGL were re-genotyped three to six times, as in the Portugal laboratory. Individuals with less than 14 (of 17) loci with a consensus genotype (from at least three successful genotypings per locus) were excluded from all analyses.

Among the 17 microsatellites, 9 were putatively neutral loci, and 8 were located in candidate (functional) genes, including 7 located in introns of genes (*KRT2, MHC2* (i.e., *OLADRBps*), *TCRG4, IFNG, MMP9, GLYCAM-1, LIF*), and one located a few hundred base pairs upstream from the candidate gene (*ADCYAP-1*). All candidate genes have some immune system function, except for KRT2 which codes for keratin, a molecule in horn and hair. All loci are described in Luikart et al. 2008a, b.

Multiplex and a single-locus PCRs were optimized and 10ul reactions were performed on MJR PTC200 thermocyclers using touch-down profiles (Table 1). Each reaction contained: 2.5 μl of template DNA, 4.5 μl of QIA multiplex mix (Qiagen), and either 1 μl of 10× primer mix, or 1 μl of 2 pM forward and reverse primers. Two different touch-down profiles with 35 cycles were used, one with an initial annealing temperature of 63°C stepping down to 58°C, and another starting at 58°C and stepping down to 53°C. Fluorescently labeled DNA fragments were visualized on an ABI 3130xl automated capillary sequencer (Applied Biosystems) in the Murdock DNA Sequencing Facility at the University of Montana. Allele sizes were determined using the ABI GS600LIZ ladder (Applied Biosystems). Chromatograms were viewed and analyzed using GeneMapper software v3.7 (Applied Biosystems).

Table 1 Characteristics of the 17 loci genotyped

Locus						
Name	*N*	*A*	SE	H_e	SE	F_{IS}
MAF36	34.2	6.0	0.71	0.74	0.01	−0.03
MAF48	33.2	7.6	0.24	0.81	0.01	0.02
MAF209	33.6	4.6	0.51	0.67	0.02	0.09
FCB304	33.8	6.0	0.84	0.69	0.05	−0.01
FCB266	34.4	6.2	0.86	0.76	0.03	0.05
HH62	33.8	7.8	0.73	0.72	0.02	0.10
MAF33	34.4	5.6	0.51	0.53	0.04	0.16
MAF65	33.6	8.6	0.81	0.81	0.02	0.07
ILST30	33.4	2.8	0.37	0.48	0.02	0.15
ADCYAP-1	34.4	6.8	0.58	0.79	0.01	−0.03
KRT2	34.2	9.8	0.20	0.84	0.02	−0.01
MHC2	29.2	6.6	0.68	0.76	0.04	0.06
TCRG4	32.2	5.4	0.93	0.53	0.09	0.07
IFNG	34.2	1.2	0.20	0.01	0.01	−0.02
MMP9	32.8	6.4	0.51	0.74	0.02	0.07
GLYCAM-1	30.0	9.2	1.16	0.79	0.04	−0.04
LIF	29.8	3.8	0.20	0.65	0.01	0.00

The top nine loci are presumed to be selectively neutral and are not near coding genes

N is the mean number of individuals genotyped among the five study areas, for each locus. *A* is the mean allelic richness among the five study areas

Consensus (i.e. most probable genotypes) genotypes were identified as in previous work (Luikart et al. 2008a; Harris et al. 2010). Consensus genotypes for microsatellite loci were based on three to six independent sample runs. Rules for determining genotypes were as follows: for a sample to be heterozygous at a locus, both alleles had to be observed twice; for a sample to be homozygous the single allele had to be observed in three independent (replicate) genotypings. In addition, ten percent of samples were randomly chosen, re-extracted, and repeat genotyped to monitor for errors. No genotype differences or errors were detected.

Sex was determined in the MCGL laboratory by PCR amplification of the amelogenin gene as in Pidancier et al. (2006). Two PCR products (~315 and 359 bp) were obtained for males but only the longer product for females. Due to the large size of the fragments at the amelogenin locus, consensus genotypes were determined as follows: heterozygotes (males) were only accepted only if a male

band was only observed twice in a heterozygous genotype or if the male band was observed three or more times; homozygotes where only the female band was observed less than three times (e.g. of three independent PCRs) were classified as of unknown gender. Sex was determined for 163 of the 172 samples.

Data analysis

The probability that two unrelated individuals (or two random siblings) would have identical genotypes (P_{ID}) was computed using DROPOUT (McKelvey and Schwartz 2005) and API-CALC (Ayres and Overall 2004).

Principal correspondent analysis (PCA) and multilocus genotype matching were conducted in GENALEX (Peakall and Smouse 2006) to identify outliers due to potential genotyping errors or non-argali samples, and to identify identical genotypes. Loci contributing significantly more unique individuals than expected were found with DROPOUT (McKelvey and Schwartz 2005). We estimated expected heterozygosity, tested for gametic (linkage) disequilibrium, and assessed departures from Hardy–Weinberg proportions using exact tests and a Markov chain as implemented in GENEPOP 3.4 (Raymond and Rousset 1995). Allelic richness estimates were corrected for sample sizes using rarefaction (Kalinowski 2005). We quantified genetic differentiation among study areas using exact tests for allele frequency differences and using GENEPOP 3.4.

We tested for reduced allelic richness and reduced heterozygosity (e.g. in study areas with low variation) using Wilcoxon's signed-ranks test. This is a nonparametric test for paired comparisons that is appropriate and powerful when homologous loci are examined in related populations. We tested for genetic signatures of recent population bottlenecks using heterozygosity excess (i.e., deficit of rare alleles) across multiple neutral loci (Cornuet and Luikart 1996, Luikart and Cornuet 1998). We used two mutation models (the stepwise model SMM; and two-phase model TPM with 80% SMM and 20% multi-step mutations with variance of 12) to cover the range of likely mutations models for microsatellite loci (Piry et al. 1999).

Connectivity among populations was assessed using indices of genetic differentiation (F_{ST}) as well as the number of migrants (Nm) estimated from both equilibrium models and assignment test approaches that do not assume equilibrium. Equilibrium migration rate models included the private alleles method (in GENEPOP 3.4), and the F_{ST} method assuming an island model of migration. Two nonequilibrium methods included a Bayesian assignment test approach (BAYESASS; Wilson and Rannala 2003) and an individual-based the assignment test of Rannala and Mountain (1997) coded in GENECLASS 2.0 (Piry et al. 2004).

We tested for locus-specific signatures of selection in two ways. First we tested for evenness of allele frequencies at an individual locus within populations using BOTTLENECK, which gives a probability for each locus being at mutation-drift equilibrium (Cornuet and Luikart 1996). We also plotted the probability values of each locus to help assess genome-wide patterns caused by demographic events (e.g. bottlenecks) that affect all loci similarly. Selection on individual loci can cause an excess of deficit of heterozygosity (i.e. rare alleles) compared to mutation-drift equilibrium expectations.

Second, we tested each locus for an excessively high or low F_{ST} compared to the mean observed F_{ST} by using F_{ST}-outlier tests (Beaumont and Balding 2004, implemented in Antão et al. 2008). An excessively high F_{ST} at a locus compared to thousands of simulated neutral loci indicates possible divergent selection; an excessively low F_{ST} suggests possible balancing selection.

Results

We identified 172 individuals from the five study areas. Among the 17 loci, eleven pairs of loci deviated significantly ($P < 0.01$) from gametic disequilibrium. However among our 1,360 tests (272 pairs of loci in each of 5 populations), approximately 13.6 deviations were expected by chance alone ($\alpha = 0.01$). No pair of loci deviated strongly from gametic disequilibrium in more than one population. The P_{ID} for identifying sibling pairs was estimated to be less than one in one hundred thousand (10^{-5}) in all study areas for the 17 loci, making power high for resolving between random individuals and sibs (Table 2).

Genetic variation

Heterozygosity ranged from a low of 0.61 in Taxkorgan (China) to a high of 0.69 in Murghab (Table 2). Allelic

Table 2 Genetic variation in each of the five study areas

Study area	N	H_e (SE)	A	F_{IS}	P_{IDsibs}
Big Pamir	63	0.68 (0.05)	6.40	0.010	8.6E-07
Taxkorgan	38	0.61 (0.05)	4.88	−0.027	6.1E-06
Little Pamir	29	0.65 (0.06)	6.65	0.050*	2.1E-06
Murghab (Tajik.)	24	0.69 (0.05)	7.31	0.028	8.2E-07
Wakhjir	18	0.69 (0.05)	5.94	0.017	1.0E-06
Mean	34.4	0.67	6.12	0.040	2.2E-6

N is the number of individuals sampled. H_e is mean expected heterozygosity. A is allelic richness (mean number of alleles per locus corrected for sample size N). P_{IDsibs} is the probability of identity among sibs (Waits et al. 2001)

* $P < 0.05$

richness ranged from 4.8 in Taxkorgan to 6.7 in Murghab. Taxkorgan had significantly lower heterozygosity and allelic richness than each other study areas ($P < 0.05$; Wilcoxon signed-ranks test). Little Pamir and Wakhjir had the second lowest heterozygosity and allelic richness, respectively. Genetic sex identification yielded a sex ratio of 44% males (72 males to 90 females) over all study areas. Within study areas, sex ratios ranged from 80% males in Wakhjir, 63% males in Taxkorgan, to 44% males in Big Pamir, 36% males in Murghab, to only 4% (one male) in Little Pamir.

Genetic structure and connectivity

Mean F_{ST} for the 17 loci was 0.035 among the five study areas. Mean F_{ST} for the nine putatively neutral loci (0.033) was similar to F_{ST} for the candidate loci (0.04) (Table 3), so most results below are reported for all 17 loci, unless stated otherwise. Pairwise mean F_{ST}'s ranged from 0.008 (between Murghab and Big Pamir) up to 0.055 (between Taxkorgan and Wakhjir). Taxkorgan had the highest pairwise F_{ST}'s ranging from a low of 0.033 (with Murghab) to 0.055 (with Wakhjir). Murghab had the lowest F_{ST}'s ranging from only 0.008 with Big Pamir, to the 0.033 with Taxkorgan.

We obtained estimates of 5.1 and 6.6 migrants per generation using the private alleles method and F_{ST}-based method (assuming an island model), respectively. The mean frequency of private alleles $p(1)$ was 0.026. The Bayesian approach (BAYESASS) for estimating the current number of migrants did not yield informative results because there was not enough information in the data given the relatively low F_{ST} (Faubet et al. 2007), despite our fairly large number of loci with high heterozygosity.

Six highly probable immigrant individuals ($P > 0.99$) were identified in four of the five populations using the individual-based assignment test of Rannala and Mountain (1997). The probable migrants included the following: two into Murghab (females from Big Pamir), one in Little Pamir (a male from Wakhjir), one in Big Pamir (a male from Murghab), and two in Wakhjir (a male from Big Pamir and a male from Little Pamir). The estimated probability of each of the six putative immigrants actually being an immigrant ranged from 99.90% for the immigrant in Little Pamir to 99.95% for the immigrant into Big Pamir. When we lowered the stringency criterion for identification of a migrant (from $P > 0.99$ to $P > 0.90$), five additional migrants were identified, including two into Murghab, two in Wakhjir, and one in Big Pamir.

No immigrants were identified in the China study area of Taxkorgan. In fact, only one individual of 38 from Taxkorgan could potentially be an immigrant, but the probability of that individual being a resident from Taxkorgan was still 11% (Fig. 2). When we lowered the criterion of certainty for the identification of an immigrant (from $P > 0.99$ to $P > 0.90$), Taxkorgan, unlike all other study areas, still showed no evidence of immigrants (e.g., Fig. 2).

Table 3 F_{ST} between all pairs of sampling areas

	Big Pamir	Taxkorgan	Little Pamir	Murghab	Wakhjir
Big Pamir	–	0.075	0.038	0.007	0.020
Taxkorgan	0.030	–	0.060	0.058	0.055
Little Pamir	0.040	0.048	–	0.029	0.016
Murghab	0.009	0.024	0.038	–	0.009
Wakhjir	0.016	0.051	0.044	0.011	–

The 9 putatively neutral loci are included below the diagonal. The eight candidate gene loci are included above the diagonal

Selection tests and F_{ST}

The *TCRG4* gene microsatellite had a significant excess ($P < 0.01$) of rare alleles (i.e., uneven allele frequency distribution compared to neutral expectations) in both the Murghab and the Little Pamir study areas. None of the nine neutral loci or the other seven candidate gene loci deviated from expected allele frequencies under mutation-drift equilibrium (Fig. 3).

Two candidate gene microsatellite loci had an F_{ST} value significantly different from neutral expectations. *GLYCAM-1* had a significantly higher F_{ST} ($F_{ST} = 0.068$; $P = 0.02$) and *ADCYAP-1* had a significantly lower F_{ST} ($F_{ST} = 0.002$, $P = 0.03$) than expected under neutrality. Neither F_{ST} deviation was significant at the 0.01 level. No neutral loci gave evidence of selection or deviated from mutation-drift equilibrium (Fig. 4).

Estimates of genetic differentiation were similar with and without the three outlier loci (*GLYCAM-1*, *ADCYAP-1*, and *TCRG4*): the mean F_{ST} decreased slightly from $F_{ST} = 0.035$ for all 17 loci to $F_{ST} = 0.033$ for the 14 loci with no selection signature. Pairwise F_{ST}, computed after removing outliers, changed most for the Taxkorgan area in China. For example, F_{ST} declined from 0.051 to 0.033 when removing the three outlier microsatellites in candidate genes.

Discussion

Our study of neutral and candidate adaptive genes in argali populations suggests relatively high variation within, and low differentiation among populations compared to other mountain sheep (e.g., Gutierrez-Espeleta et al. 2000; Worley et al. 2006; Epps et al. 2005; Hogg et al. 2006; Luikart et al. 2008a). This is similar to results obtained in

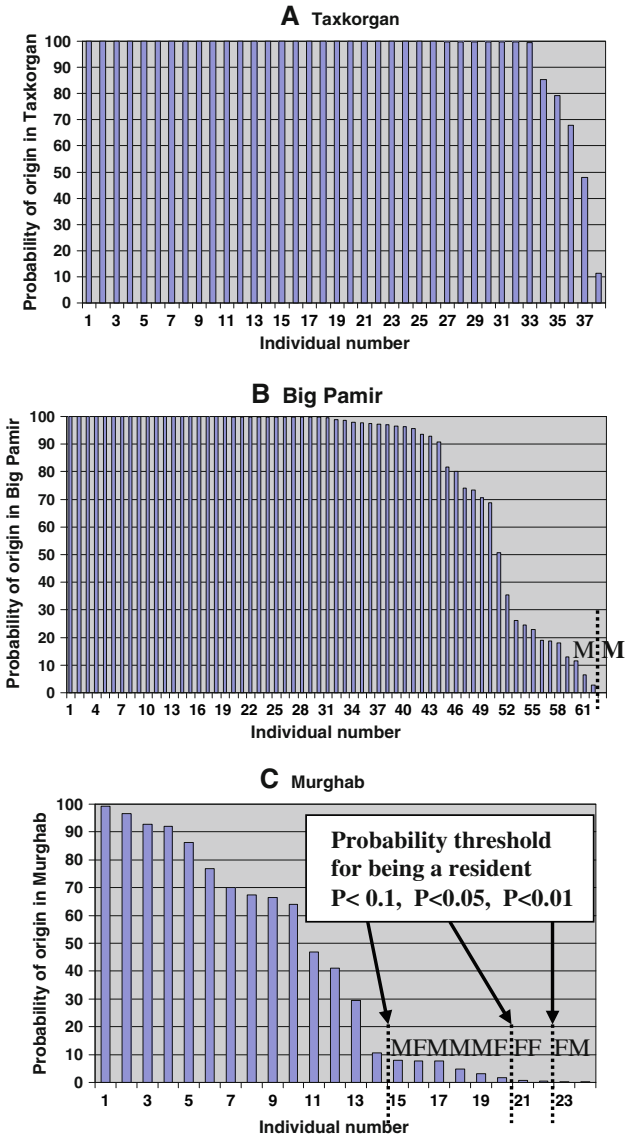

Fig. 2 Assignment test estimates of the probability of local origin of each individual in the study area from which it was sampled. In **a**, the individual (#38) least likely to originate from Taxkorgan (China) still had an estimated 11% probability of originating in Taxkorgan. In the Big Pamir **b**, one individual (#63) had a probability of only 0.04% of originating locally, and was therefore considered an immigrant. In **c** the Murghab (Tajikistan) study area, two individuals (#23 and #24) had a very low probability of local origins (<0.3%). The letters M and F designate male and female individuals with a reasonably low probability of being local residents

Mongolia using mtDNA of argali populations previously assumed to represent different subspecies (Tserenbataa et al. 2004).

Genetic variation and bottlenecks

Heterozygosity and allelic richness were high compared to many of the same loci genotyped in other wild sheep, in which mean heterozygosity is approximately $H = 0.60$ or

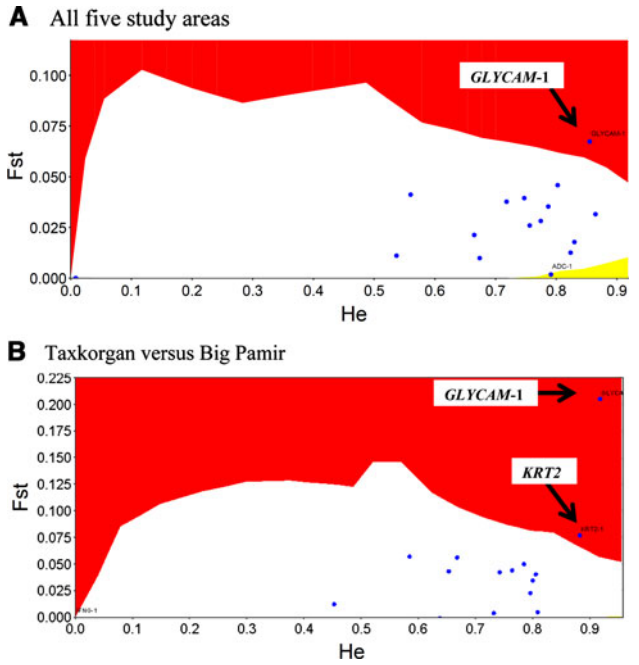

Fig. 3 F_{ST}-outlier test results showing: **a** the *GLYCAM*-1 locus with excessively high F_{ST} among **a** all five study areas, **b** between Taxkorgan (China) versus Big Pamir. *Dots* represent loci. *White area* with most *dots* represents the expected area for neutral loci (99% confidence area)

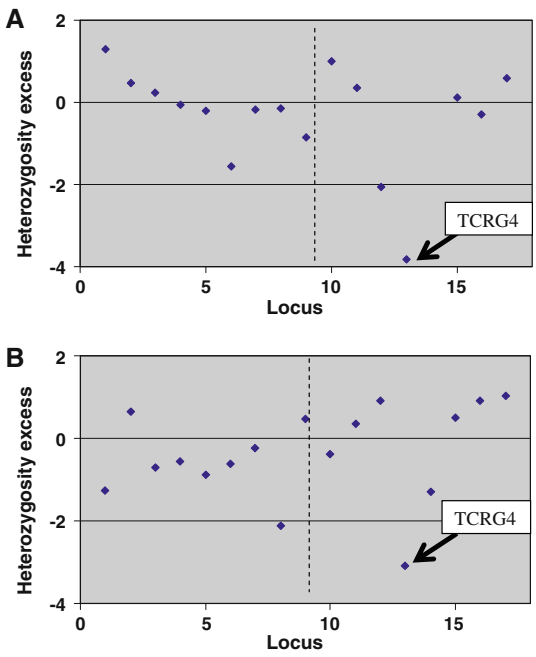

Fig. 4 One locus (TCRG4, see *arrow*) had a significant deviation from mutation-drift equilibrium, i.e., a deficit of heterozygosity (also called an excess of rare alleles), in two populations: **a** Little Pamir, and **b** Murghab (Tajikistan). Loci (*dots*) at mutation-drift equilibrium will have zero heterozygosity excess (*y*-axis). Loci are in the same order as listed in Table 1. *Vertical dashed line* separates the 9 neutral (1–9) and 8 candidate adaptive gene loci (10–17)

lower (Ozut 2001; Epps et al. 2005; Hogg et al. 2006; Luikart et al. 2008a, b), and even lower in other wild ungulate species (e.g., Gebremedhin et al. 2009 and papers cited therein). This suggests that Marco Polo argali populations have relatively high effective population sizes and that our study areas are not yet isolated or inbred, as has been feared (Shackleton 1997; Harris et al. 2010).

Nonetheless, the significantly lower heterozygosity and allelic richness in Taxkorgan (China) compared to our other study areas suggests this population is smaller (Schaller et al. 1987, 2008), and perhaps relatively more isolated than the other populations. The Taxkorgan population is near the southeastern edge of the range of argali and there is a long fence (350 km) near the Tajik-Chinese border (Schaller et al. 2008) that could potentially reduce connectivity of the Chinese argali with other populations in Murghab and the Afghan Pamirs; however it is uncertain if the fence is a barrier because, for example, it is not continuous (e.g. open on some high slopes) and argali might jump over it in some locations.

The absence of strong bottleneck signatures, even in Taxkorgan, and the reasonably high allelic richness suggest no evidence of recent or severe reductions in population size. Power for detecting severe reductions is reasonably high when using seventeen highly variable microsatellite loci and 38 individuals (Cornuet and Luikart 1996; Luikart and Cornuet 1998), as we have from Taxkorgan. Thus if the Taxkorgan population has become genetically bottlenecked or increasingly isolated, which seems likely, the signal might not detectable if the isolation was recent (e.g. <2–4 argali generations ago). Bottleneck signatures can take several generations to become detectable if the bottleneck effective size (N_e) remains fairly large (e.g. >50; Fig. 3 in Cornuet and Luikart 1996). Genetic bottleneck signatures also might be obscured by recent immigration.

Differentiation and connectivity

The genetic differentiation (F_{ST}) in argali is similar to or lower than other mountain sheep sampled at similar spatial scales in North America. For example, in desert bighorn sheep (*O. canadensis nelsoni*) from Arizona, F_{ST}'s ranged from 0.04 to 0.20 (Gutierrez-Espeleta et al. 2000). Furthermore, over a geographic distance of only 5 km, F_{ST} ranged from 0.046 to 0.113 in desert bighorn sheep populations without and with a barrier (e.g. road), respectively (Epps et al. 2005). Worly et al. (2006) found that genetic differentiation in thinhorn sheep populations (*Ovis dalli*) from western Canada was similar to that reported in desert bighorn sheep (Gutierrez-Espeleta et al. 2000). Rocky Mountain bighorn sheep (*O. canadensis canadensis*) also show F_{ST} of ~0.11 at this spatial scale (e.g. 40 km across Glacier National Park, Luikart et al. 2008a).

The lower differentiation among argali is consistent with their more cursorial nature compared to North American sheep. Argali tend to migrate and run for long distances following threat rather short sprints into steep escape terrain as bighorn sheep do. Argali are built for running, having longer legs than North American sheep. They also will move across large valleys, a behaviour which is less common in North American sheep.

The highest differentiation (F_{ST}) for Taxkorgan among our study areas is consistent with the significantly reduced genetic variation there (compared with other study areas), and suggests increased isolation or a smaller population size. We recommend additional studies of Taxkorgan and other argali on the Chinese side of international borders, along with monitoring of genetic variation to ensure early detection of population declines, isolation, or recruitment problems, which could potentially be developing. The lowest genetic differentiation in Murghab (F_{ST}'s 0.008–0.033) is consistent with it having the highest genetic variation and being centrally located in the heart of the distribution range of Marco Polo argali.

To estimate gene flow, mean F_{ST} values can give only very rough estimates of average number of migrants per generation, and only if populations are near mutation-drift equilibrium (Whitlock and McCauley 1999). Because most natural populations are seldom near equilibrium, and violate other assumptions, our estimates of ~5 or 6 migrants per generation must be interpreted with great caution; the actual number of migrants could be far higher, for example.

Current (contemporary) gene flow can be detected from the identification of actual migrants by using individual-based assignment tests (Paetkau et al. 2004). For example, Taxkorgan had no detectable immigrants (out of 38 individuals sampled) suggesting relatively low connectivity. The identification of putative immigrants in all other populations suggests they currently are not isolated. The threshold of 99% certainty for identification of a migrant could be viewed as overly stringent. With our sample size of 172 individuals (and $\alpha < 0.01$), we expect 1.7 migrants to be identified by chance alone (as false positives); whereas we identified 6 migrants ($P > 0.99$). Because we detected 6 probable migrants, it seems likely that several true migrants exist and that most populations, except perhaps Taxkorgan in China, have current migration rates greater than zero. The use of 95% certainty ($\alpha < 0.05$) for each individual assignment resulted in identification of 11 probable migrants when ~8 were expected by chance alone.

Selection and adaptation

We detected evidence for selection only at candidate gene loci, not at neutral loci, suggesting candidate gene

approaches can potentially identify loci under selection when using noninvasive sampling in wild sheep. The *GLYCAM-1* microsatellite showed higher F_{ST} than neutral expectations (Fig. 2). This could potentially result from selection at this locus or at other genes nearby such as *IFNg* which is less than 20 centimorgans away from *GLYCAM-1* in domestic sheep, and which has been associated with parasite load in sheep (Coltman et al. 2001) and other ungulates (Ezenwa et al. 2010). *GLYCAM-1* function involves mediating the trafficking of blood-born lymphocytes into secondary lymph nodes, and also is expressed in the mammary gland of lactating mammals (Hou et al. 2000; Rasmussen et al. 2002). Further studies and collection of parasite data are needed to assess if *GLYCAM-1* genotypes are associated with disease resistance in sheep.

The lower F_{ST} than neutral expectations at *ADCYAP-1* could reveal balancing selection for even and similar allele frequencies in multiple study areas (e.g., Paterson 1998). The *ADCYAP-1* gene (adenylate cyclase-activating polypeptide) is involved in regulating production of interleukin 6 that activates the production of T-helper cell 2 (Th2) cytokines involved in defense against helminths and other extracellular parasites (Mosmann and Sad 1996). *ADCYAP-1* was recently found to be associated with nematode parasite infection in domestic sheep (Crawford et al. 2006), and heterozygotes had lower parasite loads in wild bighorn sheep (Luikart et al. 2008b). If argali suffer significant mortalities from disease it is possible that parasites and disease in the Pamirs have lead to selection at *ADCYAP-1*. Future research is needed to assess potential effects of disease in argali and on *ADCYAP-1*.

Removal of the two F_{ST} outlier loci caused little change in mean F_{ST} among the five study areas (from 0.035 to 0.30, without *ADCYAP-1* and *GLYCAM-1*). Similarly, removal of the locus (TCRG4) with a heterozygosity-excess had little effect on our multilocus F_{ST} estimates among study areas.

Interestingly, removal of *GLYCAM-1* decreased F_{ST} between Taxkorgan (China) and other study areas. For example, F_{ST} changed from $\sim$0.05 to 0.04 when we removed *GLYCAM-1* when comparing Taxkorgan with Big Pamir or Taxkorgan to Wakhjir (Fig. 2b). Removal of *GLYCAM-1* did not substantially reduce F_{ST} between other study areas, suggesting this *GLYCAM-1* gene contributes substantially to the relatively high multi-locus F_{ST} observed between Taxkorgan and other study areas. The *MHC* locus also had the second highest F_{ST} between Taxkorgan and other study areas. These observations raise the speculative hypothesis that some disease-related selection differential exists between Taxkorgan and other study areas.

How could selection tests and genotyping of both neutral and candidate adaptive loci help advance conservation genetics studies? Many candidate adaptive loci will behave as neutral loci, and thus can be used to assess genetic variation, differentiation (F_{ST}), and demography (*Nm* and change in N_e). However, if a locus reliably shows a response to selection, it could be used to monitor or detect adaptive challenges (e.g. disease die-offs or environmental change) or to identify adaptively-differentiated populations that have exceptionally high F_{ST} only at candidate genes associated with selection gradients (e.g. disease or temperature). Future developments in genomics will allow noninvasive analyses of hundreds of neutral and candidate adaptive genes, which will not only help detect population declines but perhaps infer their cause; For example, if disease candidate genes show high F_{ST} then a disease-related die-off could be inferred as the cause of a population bottleneck (Simões et al. 2008).

Conclusions

Our study illustrates the potential usefulness of genotyping both neutral and candidate adaptive loci, which can allow inferences about both demographic status (migration and bottlenecks) and selection events such a disease epizootics and environmental change. Our study suggests that candidate gene loci can be used for connectivity monitoring as long as "outlier tests" are conducted to avoid using non-neutral loci when estimating parameters (e.g. F_{ST}) that can be biased by natural selection. Future noninvasive studies will include 100s of loci (e.g., SNPs) in candidate genes thanks to advances in genotyping technologies for partially-degraded DNA, such as RT-QPCR assays, which we are developing for use in Fluidigm SNP-chip dynamic arrays (Allendorf et al. 2010).

Argali populations appear to have high genetic variation and connectivity in the Pamirs within Wakhan District of Afghanistan, and Murghab (Tajikistan), but potentially are becoming isolated in Taxkorgan, China. We recommend additional studies, including genetic and demographic monitoring of connectivity, along with disease status, to help maintain connectivity and ensure persistence of argali populations. The establishment of international coordination involving Afghanistan, China, Tajikistan (as well as Pakistan, where a few argali remain), would help monitor connectivity and facilitate conservation of argali, their habitat, and other species in the region (Schaller 2007).

Acknowledgments Our study was part of the Afghanistan Biodiversity Conservation Program of the Wildlife Conservation Society (WCS), supported by the United States Agency for International Development (USAID). For field assistance we thank B. Habib, Z. Moheb, Sabir, and A. Khairzad. We also thank S. Kondratov, K.J. Zhang, S. Ostrowski, D. Bedunah, S. Nikzad, Z. Ejlasi, I. Farahmand, Q. Sahar, K. Sediqi, K. Sidiqi, G. Sediq, A. Ahamad, R. King, and L.

Yook. We thank A. Dehgan, P. Zahler, P. Smallwood, P. Bowles, and A. Simms for advice and administrative assistance. AB-P was supported by SFRH/BPD/17822/2004 RG by SFRH/BPD/36021/2007, and this work was supported by POCI/CVT/567558/2004 all from Fundacao para a Ciencia e Tecnologia (FCT), Portugal. G.L. and F.W.A. were supported in part by a grant from U.S. National Science Foundation (Grant DEB 074218). G.L. also was supported by grants from NSF (DEB 1067129), the Walton Family foundation, CIBIO, the University of Porto, Portugal, and a research grant PTDC/BIA-BDE/65625/2006 from the Portuguese Science Foundation.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation. Nature Rev Genet 11:697–709

Antão TA, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on an Fst-outlier method. BMC Bioinformatics 9:323

Ayres KL, Overall ADJ (2004) API-CALC 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives. Mol Ecol Notes 4:315–318

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B 263:1619–1626

Beja-Pereira A, Oliviera R, Alves PC, Schwartz M, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Resour 9:1279–1301

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. Annu Rev Entomol 46:441–469

Cornuet J-M, Luikart G (1996) Description and evaluation of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes for selecting or excluding populations as origins of individuals. Genetics 153:1989–2000

Crawford AM et al (2006) Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. BMC Genom 7:178

Epps CW, Palsbøll PJ, Wehausen JD, Roderick GK, Ramey RR II, McCullough DR (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. Ecol Lett 8:1029–1038

Ezenwa VO, Etienne RS, Luikart G, Beja-Pereira A, Jolles AE (2010) Hidden consequences of living in a wormy world: nematode-induced immune-suppression facilitates tuberculosis invasion in African buffalo. Am Naturalist 176:613–624

Faubet P, Waples RW, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Mol Ecol 16:1149–1166

Fedosenko AK, Blank DA (2005) Ovis ammon. Mammalian Species 773:1–15

Gebremedhin B et al (2009) Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian walia ibex. Anim Conserv 12:89–100

Gong MH, Dai ZG, Zeng ZH, Zhang Q, Song YL (2007) A preliminary survey of population size and habitats of Marco Polo sheep (*Ovis ammon polii*) in Taxkorgan Nature Reserve, Xinjiang, China. Acta Theriologica Sinica 27:317–324 (In Chinese with English abstract)

Gutierrez-Espeleta GA, Kalinowski ST, Boyce WM, Hedrick PW (2000) Genetic variation and population structure in desert bighorn sheep: implications for conservation. Conserv Genet 1:3–15

Harris RB, Winnie J Jr, Amish SJ, Beja-Pereira A, Godinho R, Costa V, Luikart G (2010) Population estimation of argali (*Ovis ammon*) in the Afghan Pamir using capture-recapture modeling from fecal DNA. J Wildlife Manag 74:668–677

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. Proc Royal Soc 273: 1491–1499

Hou Z, Bailey JP, Vomachka AJ, Matsuda M, Lockefeer JA, Horseman ND (2000) Glycosylation-dependent cell adhesion molecule 1 (*GLYCAM 1*) is induced by prolactin and suppressed by progesterone in mammary epithelium. Endochronology 141:4278–4283

IUCN (2009) IUCN Red List of Threatened Species; section by Harris RB and Reading R. IUCN Species Survival Commission, Gland Switzerland. Version 2009. <www.iucnredlist.org>

Kalinowski S (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5:187–189

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and Conservation Genetics. Trends Ecol Evol 21:629–637

Luikart G, Cornuet J-M (1998) Empirical evaluation of a test for detecting recent historical population bottlenecks. Conserv Biol 12:228–237

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nature Rev Genet 4:981–994

Luikart G, Zundel S, Rioux D, Miquel C, Keating KA, Hogg JT, Steele B, Foresman K, Taberlet P (2008a) Low genotyping error rates for microsatellite multiplexes and noninvasive fecal DNA samples from bighorn sheep. J Wildlife Manag 72:299–304

Luikart G, Pilgrim K, Visty J, Ezenwa VO, Schwartz MK (2008b) Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. Biol Lett 4:228–231

McKelvey KS, Schwartz MK (2005) DROPOUT: a program to identify problem loci and samples for noninvasive genetic sampling in a capture-mark-recapture framework. Mol Ecol Notes 5:716–718

Mosmann RR, Sad S (1996) The expanding universe of T-cell subsets: Th1, Th2 and more. Immunol Today 17:138–146

Ozut D (2001) Conservation genetics of Anatolian mouflon (*Ovis gmelinii anatolica*, Valenciennes, 1858). M.Sc. Middle East Technical University, Ankara, Turkey

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

Paterson S (1998) Evidence for balancing selection at the major histocopatibility complex in a free-living ungulate. J Hered 89:289–294

Payseur BA, Cutter AD, Nachman MW (2002) Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Mol Biol Evol 7:1143–1153

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Pidancier N, Jordan S, Luikart G, Taberlet P (2006) Evolutionary history of the genus Capra (Mammalia, Artiodactyla): Discordance between mitochondrial DNA and Y-chromosome phylogenies. Mol Phylogenet Evol 40:349–739

Piry S, Luikart G, Cornuet J-M (1999) Bottleneck: A computer program for detecting recent reductions in effective population size from allele frequency data. J Hered 90:502–503

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536–539

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA 94:9197–9201

Rasmussen LK, Johnsen LB, Petersen TE, Sorensen ES (2002) GLYCAM-1 mRNA is expressed in the mammary gland as splicing variants and encodes various aberrant truncated proteins. Immunol Lett 83:73–75

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Schaller B (2007) A proposal for a Pamir International Peace Park. USDA Forest Serv Proc RMRS 49:227–231

Schaller GB, Kang AL (2008) Status of Marco Polo sheep *Ovis ammon polii* in China and adjacent countries: conservation of a vulnerable subspecies. Oryx 42:100–106

Schaller GB, Li H, Lu H, Ren JR, Qiu MJ, Wang HB (1987) Status of large mammals in the Taxkorgan Reserve, Xinjiang, China. Biol Conserv 42:53–71

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33

Shackleton DM (ed) (1997) Wild sheep and goats and their relatives: status survey and conservation action plan for Caprinae. IUCN, Gland and Cambridge, 390 pp

Simões P, Pascual M, Santos J, Rose MR, Matos M (2008) Evolutionary dynamics of molecular markers during local adaptation: a case study in *Drosophila subobscura*. BMC Evol Biol 8:66–77

Taberlet P, Waits LP, Luikart G (1999) Non-invasive genetic sampling: look before you leap. Trends Ecol Evol 14:323–327

Tserenbataa T, Ramey RR II, Ryder OA, Quinn TW, Reading RP (2004) A population genetic comparison of argali sheep (*Ovis ammon*) in Mongolia using the ND5 gene of mitochondrial DNA; implications for conservation. Mol Ecol 13:1333–1339

Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22:1067–1076

Waits L, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Watterson GA (1978) The homozygosity test of neutrality. Genetics 88:405–417

Weinberg PI, Fedosenko AK, Arabuli AB, Myslenkov A, Romashin AV, Voloshina I, Zheleznov N (1997) The Commonwealth of Independent States (former USSR) pages 172–193. In: Shackleton DS (ed) Wild sheep and goats and their relatives: Status survey and conservation action plan for Caprinae. IUCN/SSC Caprinae Specialist Group, IUCN, Gland, Switzerland and Cambridge, p 390

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST doesn't equal 1/(4Nm + 1). Heredity 82:117–125

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191

Worley K, Carey J, Veitch A, Coltman DW (2006) Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*). Mol Ecol 15:623–637