the **genetics**society

Check for updates

ARTICLE

Evidence of spatial genetic structure in a snow leopard population from Gansu, China

Luciano Atzeni[1], Samuel A. Cushman[2], Jun Wang[1,3], Philip Riordan[1,4,5], Kun Shi[1,6✉] and David Bauman [7,8]

Understanding the spatial structure of genetic diversity provides insights into a populations' genetic status and enables assessment of its capacity to counteract the effects of genetic drift. Such knowledge is particularly scarce for the snow leopard, a conservation flagship species of Central Asia mountains. Focusing on a snow leopard population in the Qilian mountains of Gansu Province, China, we characterised the spatial genetic patterns by incorporating spatially explicit indices of diversity and multivariate analyses, based on different inertia levels of Principal Component Analysis (PCA). We compared two datasets differing in the number of loci and individuals. We found that genetic patterns were significantly spatially structured and were characterised by a broad geographical division coupled with a fine-scale cline of differentiation. Genetic admixture was detected in two adjoining core areas characterised by higher effective population size and allelic diversity, compared to peripheral localities. The power to detect significant spatial relationships depended primarily on the number of loci, and secondarily on the number of PCA axes. Spatial patterns and indices of diversity highlighted the cryptic structure of snow leopard genetic diversity, likely driven by its ability to disperse over large distances. In combination, the species' low allelic richness and large dispersal ability result in weak genetic differentiation related to major geographical features and isolation by distance. This study illustrates how cryptic genetic patterns can be investigated and analysed at a fine spatial scale, providing insights into the spatially variable isolation effects of both geographic distance and landscape resistance.

INTRODUCTION

Improving knowledge about species' demographic and evolutionary histories is increasingly important to landscape and spatial genetics. In particular, conservation measures can be strengthened by identifying patterns of allelic differentiation in spatially complex and dynamic environments (Storfer et al. 2010; Balkenhol et al. 2016). Spatial patterns in genetic diversity result from the interacting processes of migration, dispersal and reproduction (i.e., gene flow), and are highly influenced by the nature of the intervening landscape mosaic (Cushman et al. 2006, 2013). Heterogeneous environmental patterns of factors that limit movement affect the dispersal abilities of organisms and produce spatial patterns in genetic diversity and genetic differentiation (Cushman et al. 2012, 2013; Landguth et al. 2010, 2012).

Under spatially limited dispersal, population genetic theory predicts that differentiation would increase with increased geographic distance, creating isolation-by-distance (IBD) patterns (Wright 1943). In such situations, individuals living in close proximities to each other will present genetic spatial autocorrelation, that is, they will be more genetically similar than those living further apart, as a consequence of their genetic neighbourhood size (*NS*), an area in which gene flow is high relative to drift (Wright 1946; e.g., Kuhn et al. 2017). In heterogeneous

environments, the effective distance between individuals as a function of the movement cost between them is usually most strongly related to genetic differentiation, in a process commonly termed isolation-by-resistance (IBR; e.g., Cushman et al. 2006; Shirk et al. 2017a).

To identify spatial patterns and modes of isolation, methods designed to detect barriers or clusters of sub-populations have been widely adopted in conservation genetics (Blair et al. 2012; Guillot et al. 2009; Schwartz and McKelvey 2008). However, these methods are only reliable when obvious barriers to gene flow are present, and the populations can be unequivocally identified as distinct (François and Durand 2010).

Spurious breaks and clusters can become more apparent in populations separated by IBD or IBR (Guillot et al. 2009; Schwartz and McKelvey 2008; François and Durand 2010; Cushman and Landguth 2010). This is of particular concern when populations or individuals are distributed over a continuous area, without obvious obstacles to movement (Guillot et al. 2009; Cushman et al. 2015), and inference of patterns is difficult without accounting explicitly for the spatial autocorrelation of allele frequencies (Jombart et al. 2008a; Wagner and Fortin 2012).

Methods that account for spatial autocorrelation in the distribution of genetic observations offer valuable tools to dissect

[1]Wildlife Institute, School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, China. [2]US Forest Service, Rocky Mountain Research Station, Flagstaff, AZ, USA. [3]Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK. [4]Marwell Wildlife, Winchester, UK. [5]Department of Biological Sciences, University of Southampton, Southampton, UK. [6]Eco-Bridge Continental, Beijing, China. [7]Smithsonian Environmental Research Center, Edgewater, MD, USA. [8]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK. Associate editor: Xiangjiang Zhan. ✉email: kunshi@bjfu.edu.cn

patterns of genetic autocorrelation at multiple spatial scales (Jombart et al. 2009; Wagner and Fortin 2012). These methods often rely on a spatial weighting matrix (SWM), which represents a weighted connection network expressing the links between spatial units (Dray et al. 2006; Bauman et al. 2018a, 2018b). While SWMs can be purely mathematical objects to test for IBD for example, they have also been shown to be valuable tools to incorporate ecological knowledge or landscape hypotheses of impediment to movement to optimise the detection of spatial patterns (Dray et al. 2006; Bauman et al. 2018a; Benone et al. 2020). Moran's eigenvector maps (MEM; Dray et al. 2006) is a spatial eigenvector-based method that generates spatial eigenfunctions proportional to Moran's *I* coefficient of autocorrelation by diagonalising a doubly-centred SWM (details in Dray et al. 2006; Bauman et al. 2018a). They are a powerful tool to optimally detect non-random spatial patterns ranging from simple linear gradients to complex irregular fine-scaled patterns (Dray et al. 2006). These spatial variables have proved to be accurate tools to study genetic or species composition patterns, when used in regression or constrained ordination methods (Legendre et al. 2015; Manel et al. 2010; Peres-Neto et al. 2006; Bauman et al 2018c), and have been increasingly applied in genetics studies (Manel et al. 2010; Galpern et al. 2014; Dalongeville et al. 2018; Breyne et al. 2014). These spatial methods are particularly suited to identify cryptic genetic patterns and can serve as a proxy for unmeasured landscape factors (Jombart et al. 2009; Wagner and Fortin 2016; Galpern et al. 2014).

In the context of correlative landscape genetics inferences, metrics of distance derived from Principal Component Analysis (PCA) eigenvectors have been recommended to detect significant landscape-gene relationships (Shirk et al 2017b). Their use, however, is contingent upon the identification of the number of axes and proportion of inertia that captures the spatial variance in the dataset (Shirk et al. 2017b). Concurrent to the adoption of PCA-based genetic distances in link-based landscape genetics analyses (Shirk et al. 2017b; Burgess and Garrick 2020; Savary et al. 2021), the use of ordination axes as response variables in node-based regression techniques using MEM has also been shown to successfully remove redundant genetic information, while capturing meaningful variation (Dalongeville et al. 2018; Forester et al. 2016; Breyne et al. 2014; Guerrero et al. 2018). However, in both methodological approaches, the selection of the retained PCA eigenvectors often appears arbitrary, with some exceptions (Forester et al. 2016).

We therefore suggest that assessing which ordination axes can identify significant spatial structures is essential to generating reliable landscape genetics conclusions. As species differ in their demographic traits (Shirk et al. 2017b; Hein et al. 2021) and in their genetic diversity (Landguth et al. 2012), we would expect patterns to be sensitive to the strength of the genetic signal embedded in the analysed loci (Shirk et al. 2017b; Landguth et al. 2012). Understanding how different PCA components can identify and describe spatial genetic patterns, and how a trade-off in the number of markers versus the number of samples influences such detectability (e.g. Hein et al. 2021; Landguth et al. 2012) is therefore crucial.

Species exhibiting cryptic genetic patterns, resulting from low genetic diversity, low population densities and high dispersal ability, could particularly benefit from the dimensionality reduction of a PCA (Shirk et al. 2017b). One such species is the snow leopard (*Panthera uncia*), occurring in the mountains of Central Asia and subject to population declines across its range (McCarthy et al. 2017). Despite the well-established importance of genetic surveys for species conservation (Frankham 2005), information on the genetic diversity of snow leopard populations remains scarce (Weckworth 2021). With the exceptions represented by two broad-scale genetic analyses (Janecka et al. 2017; Korablev et al. 2021), genetics studies on snow leopard are usually limited in spatial extent and sample size (e.g., Janecka et al. 2008; Karmacharya

et al. 2011; Aryal et al. 2014). The few studies that have described spatial genetic structure at a local spatial scale have inferred clusters of individuals coming from disconnected study areas (Shrestha and Kindlmann 2020; Zhang et al. 2019), possibly yielding spurious patterns due to sample biases (Schwartz and McKelvey 2008; Oyler-McCance et al. 2013).

Range-wide and landscape-specific connectivity information for this species remains limited (Riordan et al. 2016; Li et al. 2020; Shrestha and Kindlmann 2020), with little known regarding the relationships between snow leopard genetic diversity and landscape characteristics. Thus, in the context of existing and emerging threats to populations (McCarthy et al. 2017), it is important to assess the genetic status of local populations, and understand landscape features that impede connectivity, potentially causing isolation and resulting in decreased genetic diversity or inbreeding depression (Frankham 2005; Weckworth 2021).

With this study we aim to quantify the spatial genetic patterns in a snow leopard population from Gansu, China, comparing two different datasets, with different numbers of loci and alleles. We summarise genetic information using PCA, and use derived principal components, at several levels of retained variance, as dependent variables to test for spatial genetic structures. We also use spatially explicit diversity indices calculated on the extent of Wright's neighbourhood size (Shirk and Cushman 2011, 2014), and assess how these indices varied in the different sampling sub-localities. This approach complements population genetic analyses, as it enables identification of discontinuities and rapid changes in diversity indices across space, which, in turn, can help to identify demographic hypotheses in relation to the observed spatial patterns of allele distributions (e.g., Ruiz-Gonzalez et al. 2015).

We aim to answer the following questions:

(1) Are there significant spatial autocorrelation genetic patterns within a snow leopard population inhabiting an apparently continuous landscape, and how do these relate to local variability of genetic diversity indices?

(2) In the case of an absence of detectable genetic spatial patterns, can the latter be due to the species dispersal capacity, the extent of allelic diversity, or the apparent homogeneity and continuity of the landscape?

(3) In the presence of genetic spatial structures, to what extent does the structure detection depend on the number of loci/individuals, and the content of PCA variance used as response variable?

METHODS
Study area
All study areas were in Gansu Province, China (Fig. 1), in which a total of seven surveys were conducted between 2014 and 2017 (Supplementary Fig. S1). Specifically, we surveyed the localities of Yanchiwan National Nature Reserve (YCW) in Subei Mongolian Autonomous County, and parts of the Qilianshan National Nature Reserve (QLS) in Sunan Yugur County. YCW, established in 1982 and upgraded at the National level in 2006, has a total area of ~13,600 km^2, and elevations ranging from 2600 to 5483 m above sea level (a.s.l.), with wide valleys found from 3000 to 4200 m a.s.l. YCW is composed of alpine cold desert, alpine meadow grassland, wetland, and desert ecosystems. QLS was established in 1988, encompassing an area of 19,872 km^2, and has elevations ranging between 2300 and 5564 m a.s.l. The landcover is composed mainly of large areas of shrubs and alpine grasslands. Forests are found at elevations from 2300 to 3000 m a.s.l., with spruce, aspen and birch trees being the dominant species. The two reserves are located at the intersection of two eco-regions, characterised by particular vegetation and abiotic environments. Specifically, YCW is in the Kunlun alpine desert arid region, in the plateau sub-cold zone, and QLS lies in the Qilian coniferous forest and steppe semi-arid region, in the plateau temperate zone (Wu et al. 2003). Efforts by the local and national government authorities to join these two reserves and create the Qilianshan National Park (50,200 km^2) were initiated in 2018 (Atzeni et al. 2020).
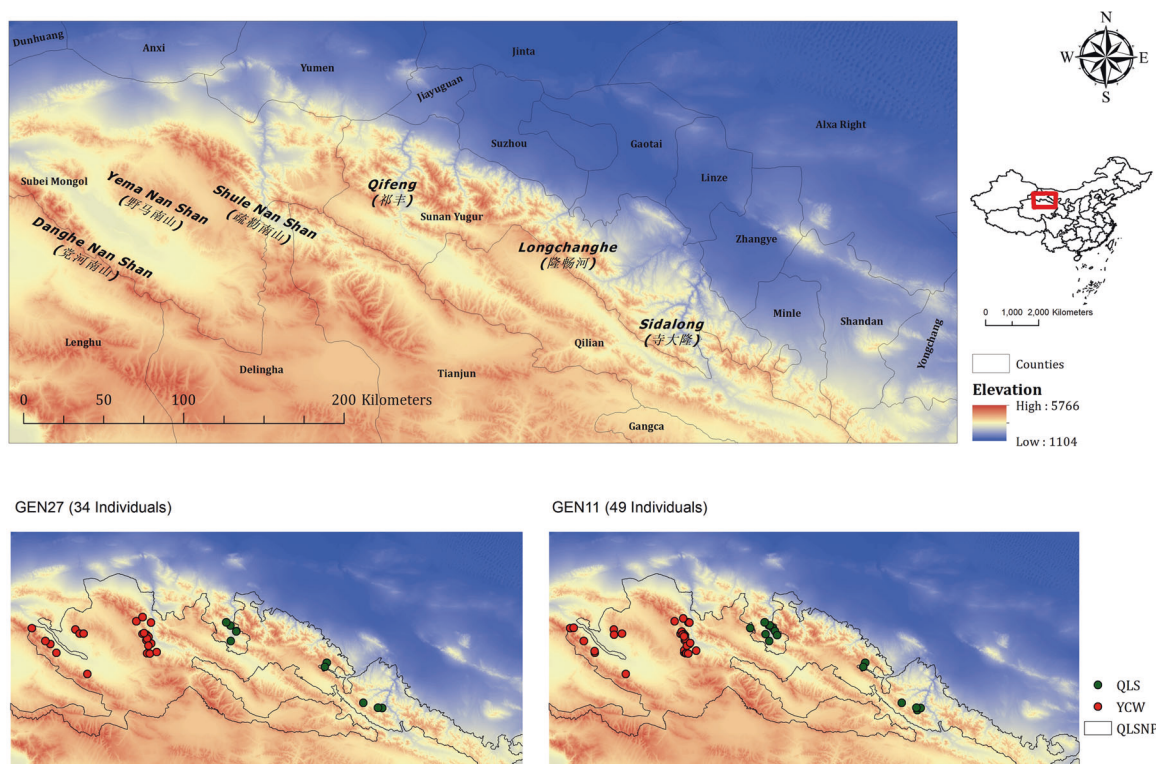
Fig. 1 Map of study areas and genetic datasets used in this study. The upper figure shows sub-localities names and County-level administrative divisions. Lower figures represent the locations of the uniquely identified individuals, divided by macro-collection area (Yanchiwan, YCW; Qilianshan, QLS). Borders represent a portion of the newly established Qilianshan National Park (QLSNP).

Sample collection

In YCW, we collected snow leopard faecal samples in the mountain ranges of Shule Nan Shan, Yema Nan Shan, Danghe Nan Shan (Fig. 1). In QLS, collection was conducted in the management areas of Qifeng, Long-changhe and Sidalong (Fig. 1). We surveyed a total extent of 4850 km², with a maximum inter-site distance of ~400 km, considering the two outermost sampling sites. We conducted line transects to collect scats believed to belong to snow leopard, using 50 mL plastic centrifuge tubes containing ~30 mL of silica gel drying agent (Janecka et al. 2008). In the field, samples were stored in dark, dry, and cool conditions, and immediately stored at −20 °C upon arrival at laboratory facilities.

Overview of genetic methods

The approaches used to generate snow leopard consensus genotypes are detailed in Supplementary Appendix 1 and summarised here. After DNA extraction, we identified carnivore species using methods described in Bai et al. (2018). We conducted individual identification using eleven snow leopard-specific microsatellite loci (Janecka et al. 2008, 2017), and employed the Quality Index criterion (QI; Miquel et al. 2006) to increase the reliability of the consensus genotypes. We clustered the profiles using allelematch (Galpern et al. 2012) to find unique multi-locus genotypes (hereafter GEN11). We then subset samples whose consensus was reached immediately after the first three replicates, identified unique individuals as before, and selected unique multi-locus genotypes to amplify further sixteen microsatellite loci (Janecka et al. 2017), totalling 27 loci (hereafter GEN27). This approach thus yielded two sets of samples, one composed of more individuals genotyped at a number of loci typically employed in non-invasive population genetics studies (GEN11), and the other composed of a higher-quality-DNA restricted pool of individuals (GEN27), allowing a more extensive genotyping and improving the trade-off between genetic profile reliability and investment of economic resources.

Summary statistics of genetic diversity

For both datasets, the number of alleles (A_N) and effective number of alleles (A_E) were summarised using GenAlEx 6.51b2 (Peakall and Smouse 2006, 2012). The software INEst 2 (Chybicki and Burczyk 2009) was used to calculate observed (H_O) and unbiased expected (H_E) heterozygosity,

adjusted estimates of observed (H_O) and expected (H_E) heterozygosity, and inbreeding coefficient ($F_{IS} = 1 − (H_O/H_E)$), accounting for the probabilistic presence of null alleles, dependent on a probabilistic genotyping failure rate. A full 'nfb' model, consisting in the probabilistic estimation of inbreeding coefficient (f), null alleles (n) and genotyping failures (b), was run for both datasets using 1,000,000 total cycles and 100,000 burn-in cycles. Global Hardy–Weinberg equilibrium was calculated using an exact test in GENEPOP (Rousset 2008) with 10,000 dememorizations, 100 batches and 10,000 iterations per batch, whilst per-locus exact tests were estimated using the R package pegas (Paradis 2010), with 10,000 Monte Carlo permutations.

Genetic profile ordination

An ordination in reduced space was implemented using PCA (Pearson 1901) on both datasets to summarise the overall variability among individuals in uncorrelated synthetic axes. Allele frequencies were centred and scaled (Jombart et al. 2009) using the function scaleGen in adegenet (Jombart 2008b), and missing allelic information was replaced with mean values (Jombart 2017). Initially, eigenvalues were converted to percentages of total variation, and then used to create four PCA objects for each of the GEN11 and GEN27 datasets, retaining the number of axes explaining ~25, 50, 75, and 100% of the total genetic inertia (hereafter PCA_25, PCA_50, PCA_75 and PCA_100). The significance of PCA axes was tested using the broken stick model (Jackson 1993) (functions bstick in the vegan package (Oksanen et al. 2019) and PCAsignificance in BiodiversityR (Kindt and Coe 2005)).

Spatially explicit indices of genetic diversity based on Wright's genetic neighbourhoods

To explore the range of spatial autocorrelation in the snow leopard genetic profiles, we tested the relationships between genetic datasets, Euclidean and resistance distances through the spatially explicit approach implemented in spatial Genetic Diversity (sGD, Shirk and Cushman 2011, 2014), an approach that overcomes bias due to Wahlund's effect (Wahlund 1928) occurring when indices are inferred at an extent greater than local population structure, and which addresses the non-transitive continuous structure of the snow leopard population (Shirk and Cushman 2011, 2014; Cushman et al. 2015).

The resistance surface in this analysis adopted the best model for this study area from Atzeni et al. (2020), which used a model selection approach based on the *MaxEnt* algorithm (Phillips et al. 2006), and found that grassland extent, landscape aggregation and fine-scale topographic position index were the main predictors of snow leopard detection in this landscape.

The habitat suitability surface from Atzeni et al. (2020) was converted to a resistance layer through a negative exponential transformation (Mateo-Sánchez et al. 2015; Wan et al. 2019), to assign higher resistance to movement only to low suitability pixels, in the form of

$$R = 100^{(-1xHS)},$$

where *HS* represents habitat suitability scores.

The resistance surface was rescaled from 1 to 20 in *spatialEco* (Evans 2020), and the original cell size (90 m) was increased to 200 m to reduce the computational time (Supplementary Fig. S2), but was of fine-enough resolution to avoid losing important landscape characteristics, thereby having a negligible overall effect on pairwise cost-distances (Cushman and Landguth 2010). Pairwise Euclidean and resistance distances were calculated through the *distmat* function in *sGD* (Shirk and Cushman 2011).

sGD calculates effective population size (N_E) at different radii, adopting the Burrow's method based on linkage disequilibrium implemented in *NeEstimator* (Do et al. 2014). N_E represents an important parameter in genetics, as it reflects the rate at which populations lose genetic diversity as a function of selection and genetic drift (Charlesworth 2009; Shirk and Cushman 2014), and it is closely related to the risk of extinction, especially in small populations (Neel et al. 2013).

In their simulations, Shirk and Cushman (2014) noted that when the extent of breeding approaches *2σ* (the circular radius defining the outer extent of Wright's neighbourhood, *NS*), the ratio between N_E:*NS* approaches unity. We selected this criterion to define *NS* in our two datasets, based on alleles with at least 10% frequency (Shirk and Cushman 2014). Breaks of 25 km were used for the Euclidean distance scenario (hereafter referred to as *geo* scenario), and 100,000 cost unit breaks for the resistance distance scenario (referred to as *res* scenario), equivalent to roughly 20 km of dispersal in optimal habitat.

We expected to observe discrepancies between the two datasets related to sample size, especially involving the resistance scenarios. Thus, as the datasets differ only in the number of loci, given that the dispersal capacity of the species is constant across the datasets, we relied on the GEN11 dataset (more individuals) for the definition of the neighbourhood sizes, constraining GEN27 to those thresholds to allow full comparability, as the same neighbourhood size must be used to compare between analyses. We predicted missing values in the calculation of indices using the package *missForest* (Stekhoven 2013), to allow a complete representation of all indices on the study area for the two datasets. The random forests machine learning algorithm was run for maximum 10 iterations of 1000 forests each.

Spatial analysis with Moran's eigenvector maps
The spatial structures of snow leopard genotypes were analysed using Moran's Eigenvectors Maps (MEM) (Dray et al. 2006). MEM are flexible and powerful eigenvector-based methods that generate a spectral decomposition of a set of spatial coordinates allowing to model multi-scale spatial structures. MEM are generated from a spatial weighting matrix (SWM), consisting of the Hadamard product of a connectivity and a weighting matrix that define which objects are connected and how connections are weighted, for example as a decreasing function of distance (Dray et al. 2006; Bauman et al. 2018a, b).

Two key steps in MEM are the selection of a SWM among a set of candidate matrices (Bauman et al. 2018a), and the definition of a subset of spatial eigenvectors to be further used as spatial predictors within the selected SWM (Bauman et al. 2018b). We optimised these selections on the basis of six SWMs built from two contrasted graph-based connection schemes (minimum spanning tree (MST) and Delaunay triangulation (DEL)), and one distance-based SWM, connecting all neighbours within a distance inferior to the smallest distance maintaining all sites connected (i.e., the longest edge of the MST graph (ND_max.edge.mst)). To weight connections along the edges of these SWMs, we tested a linear function and a concave-up function with an exponent of 0.25, yielding a total of six SWM candidates. The two weighting functions were calculated both from the Euclidean and the resistance distances, totalling two sets of six SWMs (SWMgeo and SWMres, respectively). To find the most supported topologies, we used built-in syntax of the function *listw.candidates* in

adespatial (Dray et al. 2020) to generate SWMgeo candidates, and created ad-hoc code to create SWMres candidates. These were then fed into the function *listw.select* of the same package to find the most supported SWM and subset of spatial predictors within it. The presence of spatial structures in the genetic profiles was tested on the whole set of MEM associated with positive autocorrelation structures for each SWM separately, using 9999 permutations and adjusted *p* values for multiple tests (Bauman et al. 2018a). If a SWM captured significant spatial patterns in the genetic data, we performed a spatial eigenvector selection using the forward selection procedure with double stopping criterion (Blanchet et al. 2008). The SWM and subset of spatial eigenvectors (MEMgeo and MEMres, for Euclidean and landscape resistance scenarios, respectively) yielding the highest adjusted R^2 were selected, an approach shown to produce the highest accuracy and power (Bauman et al. 2018a, b).

We assessed significance of each redundancy analysis (RDA, Wollenberg 1977) axis through the marginal method (Legendre et al. 2011) implemented in the function *anova.cca* in *vegan*, using 999 permutations. We ran the same RDAs as above using the *pcaiv* function in *ade4* (Dray and Dufour 2007), to display the fitted scores on the constrained RDA axes. All analyses were performed in the R statistical environment (R Core Team 2021; R code provided as Supplementary Appendix 3).

RESULTS
Species identification, genotyping, and clustering
A total of 475 faecal samples were collected, of which 230 were identified as snow leopards. In the interest of space, we provide full details in Supplementary Appendix 1.

Summary statistics of genetic diversity
Both datasets confirmed low allelic diversity of the population (A_N = 4.364 and 4.259; A_E = 2.627 and 2.536 for GEN11 and GEN27, respectively). The population was consistent with HWE assumptions (*p* values = 0.124 and 0.937 for GEN11 and GEN27, respectively), showing only one locus deviating from HWE expectations in both datasets (PUN132). Snow leopards in Gansu were characterised by a low inbreeding coefficient (F_{IS} equal to 0.033 and 0.011 in GEN11 and GEN27, respectively) and by intermediate values of heterozygosity in both datasets. Corrected estimates of these indices, accounting for inbreeding coefficients and null alleles, were consistent with empirical observations, due to the extremely low average rates of null alleles in both datasets (Supplementary Tables S1 and S2).

Non-spatial methods - ordination
In the two datasets (GEN11 and GEN27), the amounts of variance explained by the PCAs (25, 50, 75 and 100%) were approximately achieved by 3, 7, 13, 37 axes and 3, 8, 15, 33 axes, respectively. No significant axes were found by the broken stick models in GEN27, while the first three were identified in GEN11 (Supplementary Fig. S3). There was a clinal degree of overlap along the first three principal components in both datasets, considering genotype location (YCW and QLS), that became more marked when the number of loci was reduced (GEN11) (Fig. 2). The clinal pattern of genetic diversity was well represented by the first PC axis in both datasets, suggesting a contact zone in the area of Shule Nan Shan (Fig. 3). The subsequent two principal components highlighted finer scale differences, especially in the localities of Shule Nan Shan and Qifeng, with the two extremes of the sampled area progressively more differentiated as the inertia of the axes was reduced. Colorplots of the first three PCA axes however did not clearly identify clines or structures (Supplementary Fig. S4).

Spatially explicit neighbourhood-based diversity indices
The threshold at which the ratio N_E:*NS* approached unity based on Euclidean distances was 100 km in both datasets. In the resistance distance scenario, a threshold of 400,000 cost units was identified in GEN11, and 1.2 million cost units in GEN27 loci (Supplementary Fig. S5). Since the use of a single threshold was needed to explicitly compare results between the two analyses, we
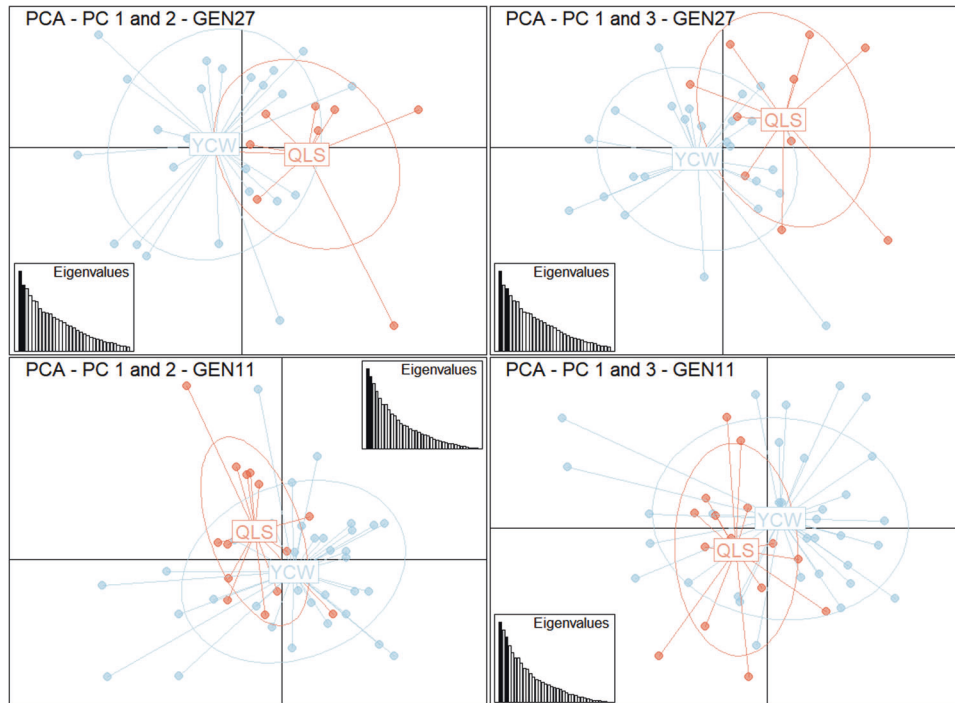
Fig. 2 Scatterplot of the first three principal component (PC) axes in the two genetic datasets. GEN27 = 34 individuals typed at 27 loci; GEN11 = 49 individuals typed at 11 loci.
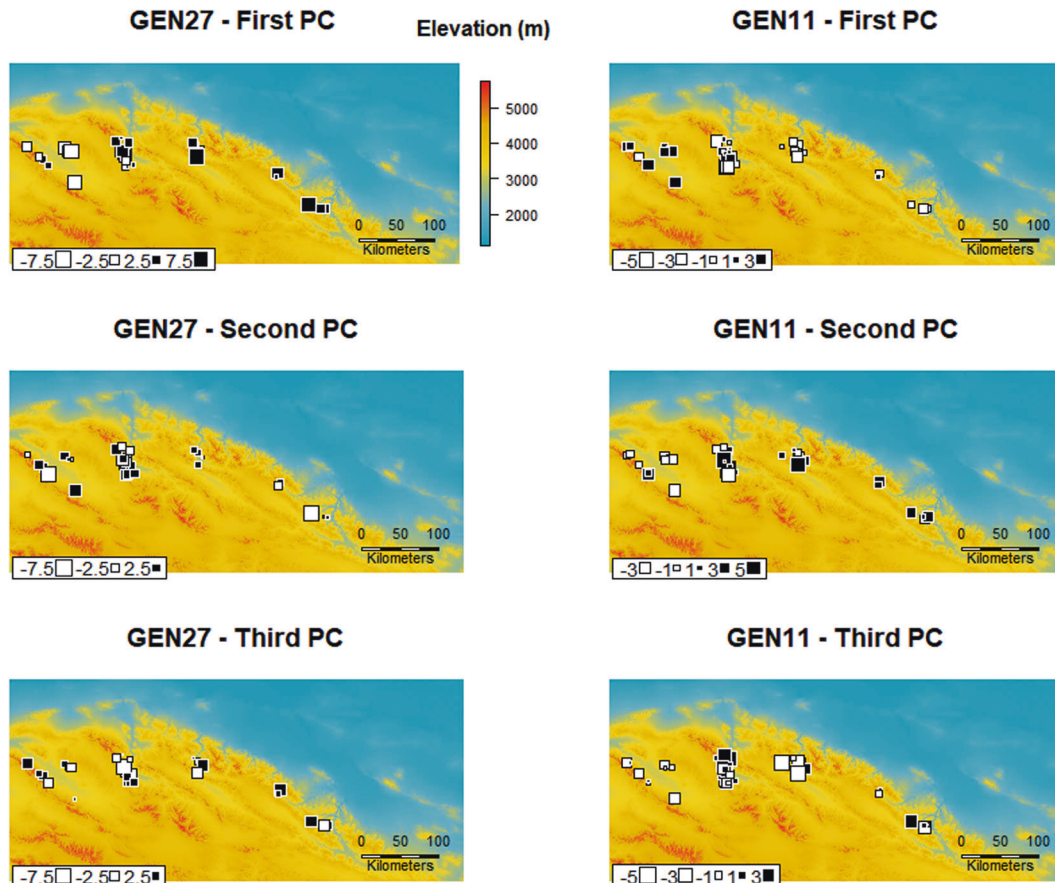

Fig. 3 Principal Component Analysis (PCA) patterns. Principal component (PC) scores relative to the first three PCA axes in the two datasets. GEN27 = 34 individuals typed at 27 loci; GEN11 = 49 individuals typed at 11 loci.

Table 1. Spatially explicit indices of genetic diversity calculated at the neighbourhood radius for which the ratio between effective population size (N_E) and neighbourhood size (N_S) approached 1 (100 km in *geo* scenario, 400,000 cost-units in *res* scenario; see main text).

Set/scenario		Spatially explicit diversity indices					
		A_N	H_O	H_E	HWE	F_{IS}	N_E
GEN27geo	Average	3.706	0.523	0.543	0.641	0.035	26.377
	SD	0.477	0.005	0.005	0.088	0.002	11.131
GEN27res	Average	3.798	0.523	0.54	0.623	0.036	30.973
	SD	0.442	0.009	0.005	0.082	0.002	14.055
GEN11geo	Average	3.909	0.553	0.587	0.496	0.063	27.973
	SD	0.531	0.02	0.02	0.161	0.023	15.315
GEN11res	Average	3.97	0.559	0.589	0.509	0.054	30.445
	SD	0.541	0.023	0.023	0.144	0.014	13.028

Values are averages and standard deviations obtained after prediction of missing values through random forest machine learning methods.
A_N number of alleles per locus, H_O and H_E observed and unbiased expected heterozygosity, HWE Hardy–Weinberg equilibrium *p* value, F_{IS} inbreeding coefficient.
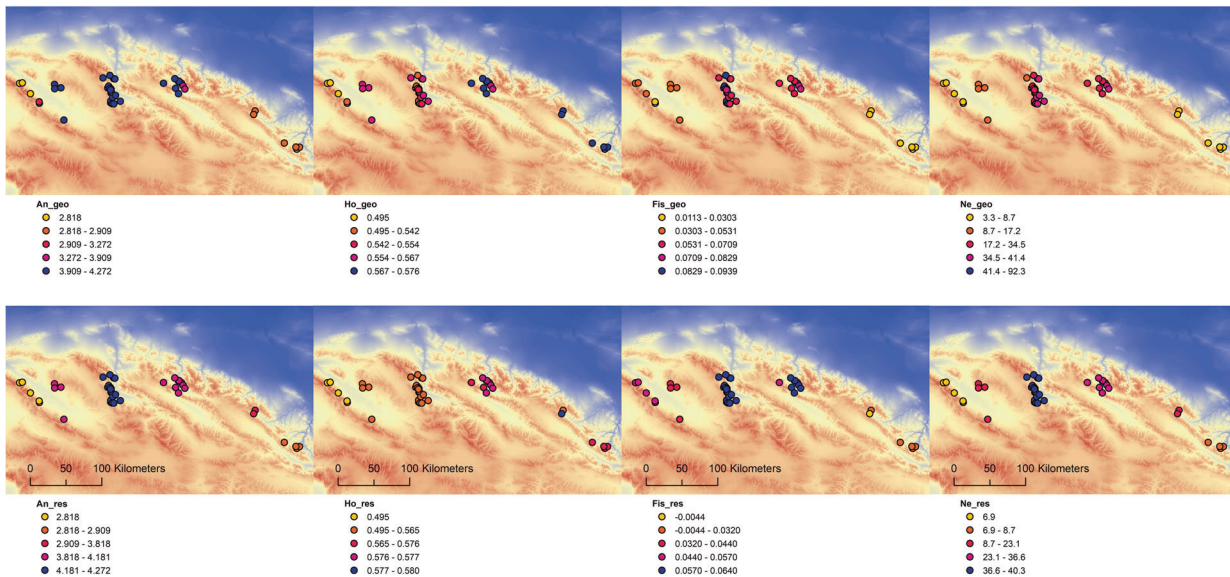

Fig. 4 **Indices of genetic diversity calculated on GEN11 dataset at the neighbourhood radius identified by the *Ne:Ns* ratio in the Euclidean distance (geo) and resistance distance (res) scenarios.** An number of alleles, Ho Observed heterozygosity, Fis inbreeding coefficient, Ne effective population size.

constrained this threshold in GEN27 to be the same as GEN11, similar to the threshold identified in the Euclidean distance scenario, amounting a distance of 80 km-equivalent circa.

All the estimated average neighbourhood-based indices remained similar within the two datasets across scenarios (Table 1). In GEN11 (Fig. 4), both scenarios showed higher average number of alleles per locus and higher N_E in the centre of the sample distribution, with progressively lower values at the periphery. Patterns of F_{IS} differed slightly between the scenarios. For Euclidean distance (GEN11geo), both edges of the study area presented the lowest values. Considering distances based on movement cost (GEN11res), only the easternmost locality displayed the lowest values, while the central localities presented F_{IS} values slightly higher than the westernmost edge (Fig. 4). Higher values of heterozygosity (H_O) were observed in the QLS portion of the study area in both scenarios. Spatially explicit indices for the GEN27 dataset (Fig. 5) were mostly concordant with those described above, especially with regard to number of alleles and N_E (lower at either edge). Heterozygosity patterns in GEN27 were also higher in the QLS region, and F_{IS} estimates were

generally lower on the eastern portion of the study area, compared to the central part (Fig. 5).

Spatial analysis - spatial weighting matrices and canonical ordination
Significant genetic spatial patterns were systematically present for both distance and resistance scenarios at all PCA inertia fractions in GEN27, and only at PCA_25 in GEN11 (Table 2; Supplementary Table S3). Usually, GEN27 sets were described by a distance-based network with a sole exception relative to PCA_75 in SWMgeo, while a graph-based connection scheme was always selected by PCA_25 in GEN11 (Table 2). The two scenarios possessed approximately the same explanatory power, with slightly higher values for SWMres in GEN27 and SWMgeo in GEN11 (Table 2). Spatial genetic structures identified in both modes of isolation were generally weak. The proportion of genetic diversity explained by eigenvectors was remarkably low when considering the full genetic variance of GEN27, and remained relatively low even when the content of PCA inertia was reduced in both datasets (Table 2). All the sets retrieved no more than four
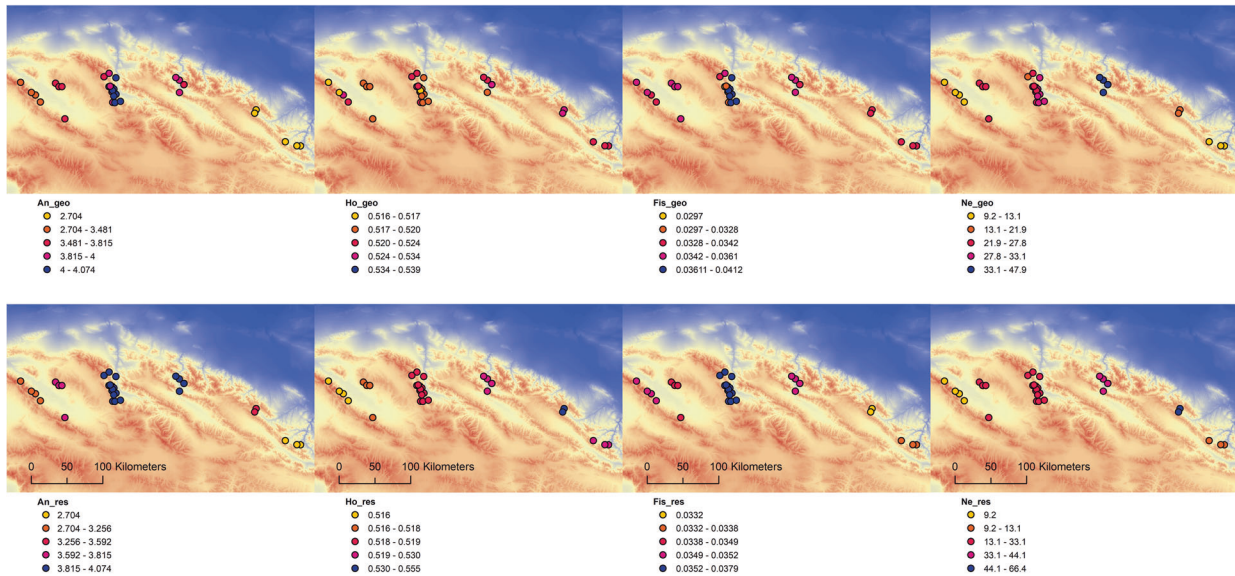
sGD indices - GEN27 dataset

Fig. 5 Indices of genetic diversity calculated on GEN27 dataset at the neighbourhood radius identified by the *Ne:Ns* ratio in the Euclidean distance (geo) and resistance distance (res) scenarios. An number of alleles, Ho Observed heterozygosity, Fis inbreeding coefficient, Ne effective population size.

significant eigenvectors (Supplementary Table S4). Significant RDA axes were always three for PCA variance equal to, or above 50%, and no more than two for the PCA_25 fractions in both GEN27 and GEN11 (Supplementary Table S5).

Major spatial patterns of snow leopard genetic diversity in Gansu

We only describe results relative to the full content of information (PCA_100) in the GEN27 dataset, as the patterns were concordant across all PCA inertia fractions (Supplementary Appendix 2, Supplementary Figs. A2.1–A2.3).

The main spatial pattern based on both RDAres and RDAgeo was a major division between the geographic locality of YCW and that of QLS. We observed admixed allelic patterns in the area of Shule Nan Shan (Fig. 6). The first axis in both RDAs found some degree of differentiation in the westernmost portion of Danghe Nan Shan mountain.

The second RDAres axis also highlighted the broad geographical division, suggesting clines in the areas of Shule and Qifeng, and from Shule to Yema/Danghe. In contrast, the second RDAgeo clustered individuals from Yema and Danghe Nan Shan, revealed a north-south gradient in Shule Nan Shan, and clearly distinguished Qifeng from the other localities in QLS.

The third RDAres and third RDAgeo axes both differentiated snow leopards at either edge of the sampling area. Overall, they both described clinal patterns within QLS (RDAres) or between QLS and YCW (RDAgeo). Shule Nan Shan was a linkage area in which genetic patterns peculiar to Danghe, Yema and Qifeng admixed (Fig. 6).

Clinal patterns in YCW were evident in the colorplot of the first three significant RDA axes (GEN27 PCA_100; Supplementary Appendix 2, Supplementary Figs. A2.5 and A2.6). While RDAres would suggest a weak, but more marked structuring in YCW, RDAgeo tended to describe this area as a continuum, with individuals at the western edge of Danghe Nan Shan always more differentiated. Both scenarios suggested a weak transition zone extending from Danghe Nan Shan to Qifeng and differentiating the cline between Longchanghe and Sidalong through augmented colour contrast (Supplementary Appendix 2, Supplementary Figs. A2.5 and A2.6).

DISCUSSION
Spatially heterogeneous genetic diversity indices

Genetic variation tends to be clinal and to vary locally when individuals are distributed over continuous areas (Chambers 1995). These variations also depend on differential patterns of landscape connectivity and habitat availability that create genetic structuring (Shirk and Cushman 2011, 2014; Jackson and Fahrig 2016). The relationships between higher genetic variation and habitat amount have been well documented in population genetics (Frankham 1996; Jackson and Fahrig 2016) and landscape genetics (Hearn et al. 2019; Macdonald et al. 2018; Bothwell et al. 2017). Availability of habitat resources and landscape continuity increase the amount of gene flow, the effective population size, and overall genetic diversity (Frankham 1996; Shirk and Cushman 2014; Bruggeman et al. 2010). Therefore, local variations in N_E in heterogeneous landscapes are driven by local population sizes, which in turn are dependent on habitat amount and connectivity (Shirk and Cushman 2014; Jackson and Fahrig 2016; Frankham 1996).

In line with these assumptions, we observed higher N_E and allelic richness in the central localities of Shule and Qifeng (Figs. 4 and 5), two areas found to harbour extensive suitable snow leopard habitats (Atzeni et al. 2020), and to be characterised by higher densities of individuals compared to the other localities analysed in this study (Wang Jun, unpublished data; Alexander et al. 2016).

When populations are genetically structured, there is a theoretical expectation for a decrease in NS and N_E from the centre to the edge of a distribution (Shirk and Cushman 2014). However, the lower values in Longchanghe and Sidalong may also be influenced by the lower number of genotyped individuals. Since the distribution of snow leopards is continuous (Atzeni et al. 2020), the current status of knowledge does not allow us to determine whether the estimated N_E and allelic number patterns reflect true demographic processes or are the effect of the sampling scheme employed in this study. However, consistent with Shirk and Cushman (2014), the lower N_E and allele numbers in Yema and Danghe correspond with lower snow leopard abundance in these sub-areas, compared to the two core central areas (Wang Jun, personal observation), a pattern driven by landscape and habitat characteristics (Atzeni et al. 2020).

Table 2. Most supported Spatial Weighting Matrices (SWM) and significant spatial functions retained (N. var) in each topology for each level of PCA inertia in the two sets.

Scenario	Set	PCA	SWM	R^2_{Adj}	p value	N.var	R^2_{Adj}.select
Resistance distance	GEN27	100	ND.max.edge.mst_up	0.0755	0.0006	4	0.0687
		75	ND.max.edge.mst_up	0.1246	0.0006	4	0.1131
		50	ND.max.edge.mst_up	0.1922	0.0006	4	0.1633
		25	ND.max.edge.mst_up	0.2871	0.0024	3	0.2462
	GEN11	100	NONE				
		75	NONE				
		50	NONE				
		25	Delaunay_up	0.2268	0.0494	4	0.1588
Euclidean distance	GEN27	100	ND.max.edge.mst_up	0.0721	0.0006	4	0.0610
		75	Delaunay_lin	0.1118	0.0149	4	0.1072
		50	ND.max.edge.mst_up	0.1822	0.0012	4	0.1417
		25	ND.max.edge.mst_up	0.2852	0.0012	3	0.2187
	GEN11	100	NONE				
		75	NONE				
		50	NONE				
		25	Delaunay_up	0.2289	0.0482	4	0.1681

R^2_{Adj} refers to the global adjusted R^2, while R^2_{Adj}.select indicates the adjusted R^2 obtained using the selected spatial functions.

We found higher levels of inbreeding in the two core areas, compared to the periphery (Figs. 4 and 5). In simulation scenarios and in empirical datasets, Shirk and Cushman (2011, 2014) postulated that lower NS and N_E at the edges of a species distribution can drive unrelated individuals to travel larger distances to mate (see also Shirk et al. 2020), resulting in an increase in heterozygosity and decrease in inbreeding coefficient. Furthermore, lower populations in peripheral areas might also drive populations inwards, causing the apparent increase in F_{IS} in the core areas (Shirk and Cushman 2014).

Recently, Shirk et al. (2020) explored the effects of gene flow from unsampled demes on the genetic composition and spatial variation of indices of the contiguous population of interest. They suggested that admixture penetrating from unsampled individuals would produce more divergence compared to less admixed individuals near the core of the sampling area. In our context, this phenomenon likely resulted in higher F_{IS} values at the centre of the sample distribution, and in lower values especially at the eastern edge, confirming evidence of genetic influence from snow leopard in the remaining portion of the Qilian mountains, and likely accounting for most of the observed genetic patterns seen in this study.

The theoretical expectation for these F_{IS} patterns may also be supported by kin structure and philopatric behaviour, a trait that has been recently observed in snow leopards (Johansson et al. 2021), and postulated to create structuring at regional level (Korablev et al. 2021). As higher kin structure is associated with connected habitats and increased gene flow (Dharmarajan et al. 2014), it is possible that areas outside the central localities represent sink habitats, characterised by lower availability of resources and composed mostly of individuals dispersing from sampled and unsampled core localities, presenting low family structure and more genetic divergence.

Spatial structures of snow leopard genetic diversity in Gansu
The clinal pattern of differentiation (Figs. 2 and 3) did not support the existence of groupings in these areas, contradicting recent suggestions (Zhang et al. 2019). Examination of PCA axes (Supplementary Fig. S4) indicated the high dispersal capacity of these animals (McCarthy et al. 2005; Johansson et al. 2018),

implying a limited role of the landscape in creating strong localised patterns of allele frequencies (Table 2).

However, RDA with MEM revealed that snow leopard genetic diversity was significantly spatially structured, albeit weakly (Table 2). Overall, the main division was between Yema and Danghe in YCW, and the entire QLS portion, with clinal patterns emerging at finer spatial scale, describing an extensive contact zone in Shule Nan Shan (Fig. 6; Supplementary Appendix 2). The differentiation of samples at either edge of the study area suggests the need for additional survey efforts along the whole Qilian mountains, and west of YCW in the Altun mountains.

The evidence generated by MEM analyses complements the patterns observed in spatial variation of diversity indices from sGD analysis (Figs. 4 and 5). Admixture in the core areas stresses the importance of these two structurally connected localities for overall landscape connectivity, especially at a transition zone between eco-regions (Wu et al. 2003). Shule and Qifeng, due to their favourable habitat characteristics, contained more individuals and received migrants from peripheral areas, creating a hotspot of genetic diversity in the northwestern portion of the Qilian mountains.

The spatial variability of patterns is indicative of non-isotropic clinal structures. Where gene flow is reduced by the effect of the landscape matrix, local densities decrease as a result of diminished functional connectivity (Kaszta et al. 2019, 2020). For example, Ruiz-Gonzalez et al. (2015) explored the spatial genetic structure of pine and stone martens in northern Spain. Their findings revealed complex clinal structures, with steeper differentiation corresponding to the boundaries of the identified clusters, likely driven by high landscape heterogeneity and fragmentation affecting gene flow. In our context, although our findings do not support the existence of discrete population units, we postulate that lower landscape connectivity in YCW might produce more differentiation over short distances, leading to the main genetic division observed in the data (Fig. 6; Supplementary Appendix 2). This results in an area of more rapid changes within a continuous genetic gradient (Fig. 6; Supplementary Appendix 2).

Strength and significance of patterns
Although significantly structured, spatial patterns were found to be relatively weak. This low spatial signal in the snow leopard
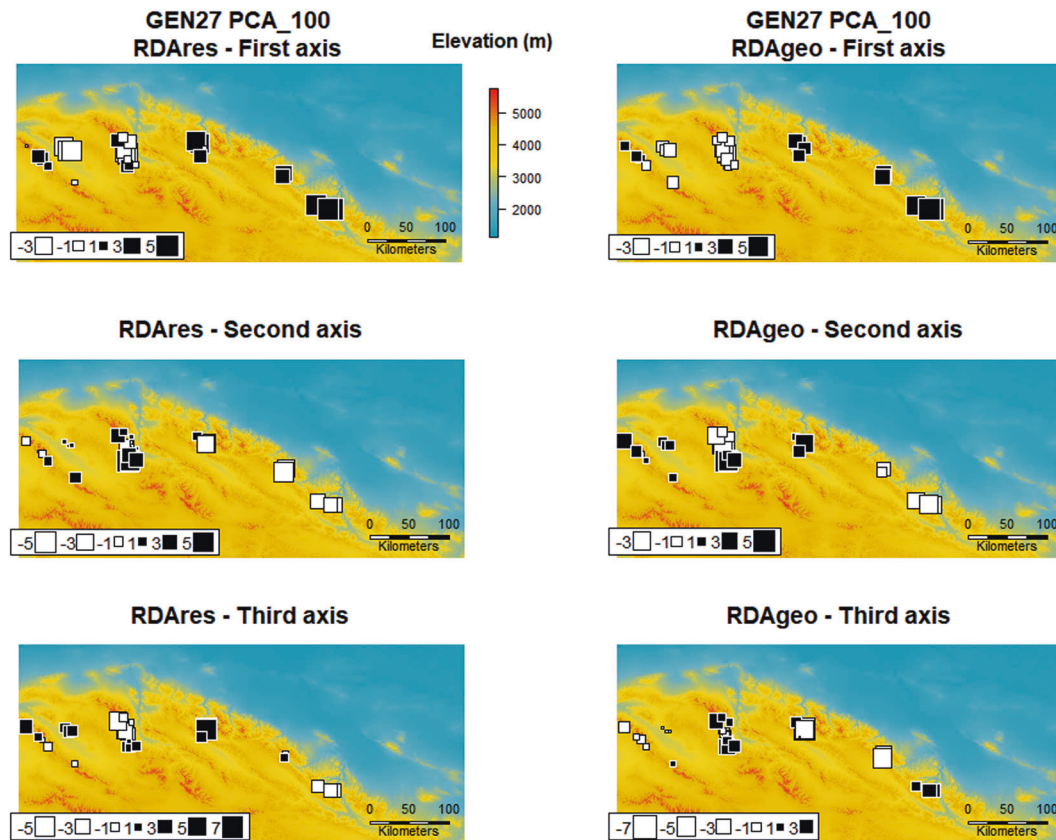
Fig. 6   Redundancy analysis (RDA) patterns. Significant axes for the landscape resistance scenario (RDAres) and Euclidean distance scenario (RDAgeo), in the set at 27 loci (GEN27), PCA variance equal to 100%.

genetic profiles may be attributable first to a combination of the inherent low genetic diversity and the high dispersal capacity of this species, and secondarily to limited sample sizes. Recently, Hein et al. (2021) demonstrated that demographic histories are among the most influential factors determining the strength of adjusted R^2 in MEM-based genetic analyses. The authors also illustrated that low sample sizes (in this paper represented by the number of nodes in the topologies) generally diminish the strength of the spatial signals. Our observations imply that, for the extent analysed in this study, neither geographical distance alone nor the landscape matrix are sufficiently strong constraints to drive the emergence of strong, localised spatial genetic patterns (Cushman et al. 2013), given snow leopards' ecology (Fig. 6; Supplementary Appendix 2).

Dispersal capacity of the species and the permeability of the landscapes are key parameters driving the signal-to-noise ratio in landscape genetics (Cushman et al. 2013; Shirk et al. 2017b, 2020). In this study, we explored the effect of the landscape matrix through an exponential transformation of a habitat model (Wan et al. 2019; Supplementary Fig. S2). Although this is a common practice in study of dispersal and/or landscape genetics (Zeller et al. 2018), additional work to optimise landscape resistance based on genetic differentiation (e.g., Shirk et al. 2010; Castillo et al. 2014; Shirk et al. 2017a) is necessary to improve the reliability of inferences regarding landscape effects on genetic diversity and genetic structure, given that habitat relationships often correlate poorly with patterns of genetic diversity and genetic differentiation (e.g., Wasserman et al. 2010).

Much emphasis has been previously given to the utility of resistance distances to weight the edges of SWMs (Bauman et al. 2018a; Galpern et al. 2014). We found that using either distance type produced comparable values of adjusted R^2 for each of the

PCA inertia fractions in the two datasets (Table 2), which could either suggest that other influential factors may determine the strength of spatial genetic structures (Hein et al. 2021; Shirk et al. 2017b), or that the response variables studied here are weakly structured at the range of spatial scales detectable by our sampling design. Further work will be necessary to gain better insight into the benefits of incorporating alternative landscape hypotheses in SWMs (Wagner and Fortin 2016).

Differences between datasets

Contrasting results were obtained from the different datasets in their ability to identify spatial structures and to describe finer scale genetic variation (consistent with Landguth et al. 2012). Hein et al. (2021), argued that the strength of the spatial structure is not affected by decreasing number of loci, and contrasted their observations to previous landscape genetics findings which demonstrated that correlations between genetic and ecological distances are sensitive to the number of loci and the level of polymorphism (Landguth et al. 2012; Oyler-McCance et al. 2013). These divergent conclusions might be resolved given that they are based on different analytical approaches and different questions: one seeking to ascertain whether there is significant genetic structure (and resulting spatial autocorrelation patterns), while the other attempting to find a landscape scenario creating those patterns. Our results partly agree with Hein et al. (2021)'s observations, and partly with those of Landguth et al. (2012) (e.g., stronger relationships in the dataset with more loci), but emphasise the importance of analysing the proportion of the allelic information which captures the spatial variation (e.g., Shirk et al. 2017b). In fact, even datasets with limited genetic resolution can describe significant spatial genetic variation (Hein et al. 2021), particularly when PCA analysis was

used to retain meaningful portions of genetic variance while reducing the effect of random variation unrelated to spatial genetic structure (e.g., Shirk et al. 2017b; Table 2).

Ordination axes as response variables in RDA and partial RDA are increasingly applied in spatial genetics studies (e.g. Dalongeville et al. 2018; Guerrero et al. 2018; Breyne et al. 2014). This is because ordination methods can distil the meaningful genetic variation in each dataset, avoiding the inclusion of unnecessary axes capturing mostly noise (Patterson et al. 2006; Shirk et al. 2017b). In link-based analyses, Shirk et al. (2017b) recommended that under the most challenging conditions for detection of landscape genetics patterns (i.e., low sample size and high dispersal ability), including additional PCA axes would improve the accuracy of inferences regarding the processes driving genetic diversity. However, this appears to be related to the variability of allelic information contained in a given dataset, in which case adding further variance does not improve accuracy (Shirk et al. 2017b). These observations seem also to apply to other analytical frameworks, as in this study. Our results highlight the importance of other explorations of the utility and the behaviour of MEM in analysing landscape variation in genetic patterns. The circumstances in which detection of patterns will be enabled or inhibited, at varying number of ordination axes (e.g. Landguth et al. 2012; Forester et al. 2016), needs to be carefully inspected across a wide array of possible factors, represented by demographic histories, number of markers, genetic variability, sample sizes (i.e., nodes in the network), and possibly alternative landscape resistance hypotheses (e.g., Landguth et al. 2012; Cushman et al. 2013).

Implications for snow leopard research and conservation

The connection between the number of genetic markers, the genetic signal and their power to describe significant patterns has practical implications in studies of spatial genetic diversity. Theoretically, a small subset of loci can suffice to identify either spatial structures (Hein et al. 2021) or the generating landscape processes in link-based analyses (e.g., Short-Bull et al. 2011; Landguth et al. 2012). Given the slow progress of snow leopard genetic research (Weckworth 2021), to date no studies, besides this one, have produced a comprehensive description of local allelic diversity to guide further genetic surveys. As it is not possible to ascertain in advance the quality of samples, the number of reliable profiles, and the degree of polymorphism, studies limited in extent should proceed iteratively, ascertaining first whether the loci chosen for individual identification are enough to capture meaningful and significant spatial variation. This of course depends on the exploration of the variance expressed by ordination axes which is representative of the whole information contained in the datasets (Forester et al. 2016; Shirk et al. 2017b).

If patterns are undetectable, then evaluating the interplay between the number of markers and PCA variance, together with the adoption of the spatial methods applied in this study, will help researchers clarify whether there is in fact no spatial genetic pattern, or if the lack of detected structure may be related to insufficient data, be it sample size (Hein et al. 2021), number of markers (Landguth et al. 2012; Oyler-McCance et al. 2013), genetic variation, or their interaction (Landguth et al. 2012). Landscape genetics inferences regarding the landscape process that have generated spatial genetic patterns are highly susceptible to the interaction between sample size, number of markers and allelic richness (Landguth et al. 2012; Oyler-McCance et al. 2013). Ongoing work is revealing that analyses of few loci produce less ecologically accurate inferences for snow leopard genetic structure (Atzeni et al. submitted), which has a larger effect than that of sample size of genotyped individuals (Atzeni et al. submitted; Landguth et al. 2012). Future directions in landscape and spatial genetics might fruitfully evaluate relationships

between the strength of a spatial signal and the accuracy of landscape genetics inferences, ideally using a simulation approach that controls the pattern-process relationships (e.g., Landguth and Cushman 2010; Landguth et al. 2012). These observations are generalisable to other highly vagile species for which cryptic genetic patterns are expected.

CONCLUSIONS

This study described the presence of weak spatial genetic structure in a snow leopard population from Gansu, China, revealing a principal geographical division between adjacent mountain ranges coupled with a cline of differentiation coincident with two admixture localities which were distinctive in their higher effective population size and allelic diversity.

Overall, spatially explicit indices of diversity, together with evidence generated through our MEM-based approach, emphasised the key importance of two core areas in providing potential snow leopard population strongholds and source for dispersal in the northwestern portion of the Qilian landscape.

Our analytical framework combined the detection of genetic structures with the assessment of the spatial variation of genetic diversity parameters. This gave us increased power to describe cryptic patterns of genetic diversity. Our approach represents a particularly effective strategy to gain insights into the localised differentiation in continuously distributed populations, providing a means to explain the nature of inferred spatial structures through demographic patterns. This enables hypothesis testing regarding the manner in which the landscape facilitates or impedes gene flow, which is essential for tailored conservation strategies.

This study also fills an important knowledge gap in snow leopard research, providing genetic baseline data for continuously distributed individuals in an under-studied landscape, at a relatively high number of microsatellite loci. The results will guide the design of future surveys to expand spatial genetics inferences to larger extents, and guide future large-scale correlative landscape genetics studies to quantify the effect of landscape structure on snow leopard gene flow across its range. As research and conservation efforts for snow leopard become more restricted to 'high-quality' patches (Johansson et al. 2016), it is increasingly vital to understand genetic structure and the landscape, management and other factors that might affect the species' survival. Finally, this study raises important methodological questions regarding the applicability of PCA axes and spatial eigenvector-based methods such as MEM in landscape genetics, which we hope will inspire further work to improve our understanding of gene-environment relationships.

Data archiving

Genotypes and geographic coordinates of snow leopard individuals in the two datasets cannot currently made available due to restrictions and directives from the competent authorities of the Popular Republic of China.

The corresponding author is willing to consider any reasonable request for data sharing and to gather the necessary permissions to do so.

R code relative to the calculation of cost distance-weighted Spatial Weighting Matrices (SWMs), is provided as appendix material to this manuscript.

REFERENCES

Alexander JS, Zhang C, Shi K, Riordan P (2016) A granular view of a snow leopard population using camera traps in Central China. Biol Conserv 197:27–31

Aryal A, Brunton D, Ji W, Karmacharya D, McCarthy T, Bencini R et al. (2014) Multi-pronged strategy including genetic analysis for assessing conservation options for the snow leopard in the central Himalaya. J Mammal 95:871–881

Atzeni L, Cushman SA, Bai D, Wang J, Chen P, Shi K et al. (2020) Meta-replication, sampling bias, and multi-scale model selection: a case study on snow leopard (*Panthera uncia*) in western China. Ecol Evol 10:7686–7712

Bai D-F, Chen P-J, Atzeni L, Cering L, Li Q, Shi K (2018) Assessment of habitat suitability of the snow leopard (*Panthera uncia*) in Qomolangma National Nature Reserve based on MaxEnt modeling. Zool Res 39:373–386

Balkenhol N, Cushman SA, Storfer AT, Waits LP, eds (2016) Landscape genetics: concepts, methods, applications, 1st ed. John Wiley and Sons Ltd. Oxford, UK

Bauman D, Vleminckx J, Hardy OJ, Drouet T (2018c) Testing and interpreting the shared space-environment fraction in variation partitioning analyses of ecological data. Oikos 128:274–285

Bauman D, Drouet T, Dray S, Vleminckx J (2018b) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. Ecography 41:1638–1649

Bauman D, Drouet T, Fortin M, Dray S (2018a) Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. Ecology 99:2159–2166

Benone NL, Soares BE, Lobato CMC, Seabra LB, Bauman D, Montag LF de A (2020) How modified landscapes filter rare species and modulate the regional pool of ecological traits? Hydrobiologia

Blair C, Weigel DE, Lazik M, Keeley AT, Walker FM, Landguth E et al. (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. Mol Ecol Resour 12:822–833

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. Ecology 89:2623–2632

Bothwell HM, Cushman SA, Woolbright SA, Hersch-Green EI, Evans LM, Whitham TG et al. (2017) Conserving threatened riparian ecosystems in the American West: Precipitation gradients and river networks drive genetic connectivity and diversity in a foundation riparian tree (*Populus angustifolia*). Mol Ecol 26:5114–5132

Breyne P, Mergeay J, Casaer J (2014) Roe deer population structure in a highly fragmented landscape. Eur J Wildl Res 60:909–917

Bruggeman DJ, Wiegand T, Fernández N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). Mol Ecol 19:3679–3691

Burgess SM, Garrick RC (2020) Regional replication of landscape genetics analyses of the Mississippi slimy salamander, *Plethodon mississippi*. Landsc Ecol 35:337–351

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. Mol Ecol 23:843–856

Chambers SM (1995) Spatial structure, genetic variation, and the neighborhood adjustment to effective population size. Conserv Biol 9:1312–1315

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Heredity 100:106–113

Cushman SA, Landguth EL (2010) Scale dependent inference in landscape genetics. Landsc Ecol 25:967–979

Cushman SA, Shirk A, Landguth EL (2012) Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landsc Ecol 27:369–380

Cushman SA, Shirk AJ, Landguth EL (2013) Landscape genetics and limiting factors. Conserv Genet 14:263–274

Cushman SA, McRae BH, McGarigal K (2015) Basics of landscape ecology: an introduction to landscapes and population processes for landscape geneticists. In: Balkhenol N, Cushman S, Storfer A, Waits L (Eds) Landscape genetics: concepts, methods, applications. Wiley, Ofxord, UK, p 11–34

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Naturalist 168:486–499

Dalongeville A, Andrello M, Mouillot D, Lobreaux S, Fortin M, Lasram F et al. (2018) Geographic isolation and larval dispersal shape seascape genetic patterns differently according to spatial scale. Evol Appl 11:1437–1447

Dharmarajan G, Beasley JC, Fike JA, Rhodes OE (2014) Effects of landscape, demographic and behavioral factors on kin structure: testing ecological predictions in a mesopredator with high dispersal capability. Anim Conserv 17:225–234

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour 14:209–214

Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Soft 22:1–20

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196:483–493

Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, et al. (2020) adespatial: multivariate multiscale spatial analysis. R package version 0.3-8. https://CRAN.R-project.org/package=adespatial

Evans JS (2020) spatialEco. R package version 1.3-1, https://github.com/jeffreyevans/spatialEco

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Mol Ecol 25:104–120

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784

Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conserv Biol 10:1500–1508

Frankham R (2005) Genetics and extinction. Biol Conserv 126:131–140

Galpern P, Peres-Neto PR, Polfus J, Manseau M (2014) MEMGENE: Spatial pattern detection in genetic distance data. Methods Ecol Evolution 5:1116–1120

Galpern P, Manseau M, Hettinga P, Smith K, Wilson P (2012) Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. Mol Ecol Resour 12:771–778

Guerrero J, Byrne AW, Lavery J, Presho E, Kelly G, Courcier EA et al. (2018) The population and landscape genetics of the European badger (*Meles meles*) in Ireland. Ecol Evol 8:10233–10246

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18:4734–4756

Hearn AJ, Cushman SA, Goossens B, Ross J, Macdonald EA, Hunter LTB et al. (2019) Predicting connectivity, population size and genetic diversity of Sunda clouded leopards across Sabah, Borneo. Landsc Ecol 34:275–290

Hein C, Moniem HEA, Wagner HH (2021) Can we compare effect size of spatial genetic structure between studies and species using moran eigenvector maps? Frontiers. Ecol Evol 9:612718

Jackson DA (1993) Stopping rules in principal components analysis: a comparison of heuristic and statistical approaches. Ecology 74:2204–2214

Jackson ND, Fahrig L (2016) Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landsc Ecol 31:951–968

Janecka J, Jackson R, Yuquang Z et al. (2008) Population monitoring of snow leopards using noninvasive collection of scat samples: a pilot study. Anim Conserv 11:401–411

Janecka JE, Janecka JE, Yu-Guang Z, Di-Qiang L, Munkhtsog B, Bayaraa M et al. (2017) Range-wide snow leopard phylogeography supports three subspecies. J Hered 108:597–607

Johansson Ö, Rauset G, Samelius G, McCarthy T, Andrén H, Tumursukh L et al. (2016) Land sharing is essential for snow leopard conservation. Biol Conserv 203:1–7

Johansson Ö, Koehler G, Rauset G, Samelius G, Andrén H, Mishra C et al. (2018) Sex-specific seasonal variation in puma and snow leopard home range utilization. Ecosphere 9(8):e02371. https://doi.org/10.1002/ecs2.2371.

Johansson Ö, Ausilio G, Low M, Lkhagvajav P, Weckworth B, Sharma K (2021) The timing of breeding and independence for snow leopard females and their cubs. Mamm Biol 101:173–180

Jombart T (2008b) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. Heredity 102:330–341

Jombart T, Devillard S, Dufour A-B, Pontier D (2008a) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:hdy200834

Jombart T (2017) An introduction to adegenet 2.1.0. https://github.com/thibautjombart/adegenet/wiki/Tutorials

Karmacharya DB, Thapa K, Shrestha R, Dhakal M, Janecka JE (2011) Noninvasive genetic population survey of snow leopards (*Panthera uncia*) in Kangchenjunga conservation area, Shey Phoksundo National Park and surrounding buffer zones of Nepal. BMC Research Notes 4:516

Kaszta Ż, Cushman SA, Hearn AJ, Burnham D, Macdonald EA, Goossens B et al. (2019) Integrating Sunda clouded leopard (*Neofelis diardi*) conservation into development and restoration planning in Sabah (Borneo). Biol Conserv 235:63–76

Kaszta Ż, Cushman SA, Htun S, Naing H, Burnham D, Macdonald DW (2020) Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, *Neofelis nebulosa*. Landsc Ecol 35:727–746

Kindt R, Coe R (2005) Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi (Kenya). http://www.worldagroforestry.org/output/tree-diversity-analysis

Korablev MP, Poyarkov AD, Karnaukhov AS, Zvychaynaya EYU, Kuksin AN, Malykh SV et al. (1776) Large-scale and fine-grain population structure and genetic

diversity of snow leopards (*Panthera uncia* Schreber, 1776) from the northern and western parts of the range with an emphasis on the Russian population. Conserv Genet 22:397–410

Kuhn A, Bauman D, Darras H, Aron S (2017) Sex-biased dispersal creates spatial genetic structure in a parthenogenetic ant with a dependent-lineage reproductive system. Heredity 119:207–213

Landguth E, Cushman S, Schwz M, Mckelvey K, Mury M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4179–4191

Landguth EL, Cushman SA (2010) cdpop: A spatially explicit cost distance population genetics program. Mol Ecol Resour 10:156–161

Landguth EL, Fedy BC, Oyler-Mccance SJ, Garey AL, Emel SL, Mumma M et al. (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Mol Ecol Resour 12:276–284

Legendre P, Oksanen J, ter Braak CJF (2011) Testing the significance of canonical axes in redundancy analysis. Methods Ecol Evol 2:269–277

Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? Methods Ecol Evol 6:1239–1247

Li J, Weckworth BV, McCarthy TM, Liang X, Liu Y, Xing R et al. (2020) Defining priorities for global snow leopard conservation landscapes. Biol Conserv 241:108387

Macdonald EA, Cushman SA, Landguth EL, Hearn AJ, Malhi Y, Macdonald DW (2018) Simulating impacts of rapid forest loss on population size, connectivity and genetic diversity of Sunda clouded leopards (*Neofelis diardi*) in Borneo. Plos One 13:e0196974

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. Mol Ecol 19:3824–35

Mateo-Sánchez MC, Balkenhol N, Cushman S, Pérez T, Domínguez A, Saura S (2015) A comparative framework to infer landscape effects on population genetic structure: are habitat suitability models effective in explaining gene flow? Landscape Ecol 30:1405–1420

McCarthy TM, Fuller TK, Munkhtsog B (2005) Movements and activities of snow leopards in Southwestern Mongolia. Biol Conserv 124:527–537

McCarthy T, Mallon D, Jackson R, Zahler P, McCarthy K (2017) *Panthera uncia*. The IUCN red list of threatened species 2017: e.T22732A50664030. https://doi.org/10.2305/IUCN.UK.2017-2.RLTS.T22732A50664030.en. Accessed 29 June 2019

Miquel C, Bellemain E, Poillot C, Bessiere J, Durand A, Taberlet P (2006) Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. Mol Ecol Notes 6:985–988

Neel MC, McKelvey K, Ryman N, Lloyd MW, Bull RS, Allendorf FW et al. (2013) Estimation of effective population size in continuously distributed populations: there goes the neighborhood. Heredity 111:189–199

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. (2019) vegan: community ecology package. R package version 2.5-6. https://CRAN.R-project.org/package=vegancitat

Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. Conserv Genet 14:275–285

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genet 2:e190

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

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Riordan P, Cushman SA, Mallon D, Shi K, Hughes J (2016) Predicting global population connectivity and targeting conservation action for snow leopard across its range. Ecography 39:419–426

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

Ruiz-Gonzalez A, Cushman SA, Madeira MJ, Randi E, Gómez-Moliner BJ (2015) Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. Mol Ecol 24:5110–5129

Savary P, Foltête J, Moal H, Vuidel G, Garnier S (2021) Analysing landscape effects on dispersal networks and gene flow with genetic graphs. Mol Ecol Resour 21:1167–1185

Schwartz MK, McKelvey KS (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv Genet 10:441

Shirk AJ, Cushman SA (2011) sGD: software for estimating spatially explicit indices of genetic diversity. Mol Ecol Resour 11:922–934

Shirk AJ, Cushman SA (2014) Spatially-explicit estimation of Wright's neighborhood size in continuous populations. Front Ecol Evolution 2:62

Shirk AJ, Landguth EL, Cushman SA (2020) The effect of gene flow from unsampled demes in landscape genetic analysis. Mol Ecol Resour 21:394–403

Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new model selection framework. Mol Ecol 19:3603–3619

Shirk AJ, Landguth EL, Cushman SA (2017a) A comparison of regression methods for model selection in individual-based landscape genetic analysis. Mol Ecol Resour 18:55–67

Shirk AJ, Landguth EL, Cushman SA (2017b) A comparison of individual-based genetic distance metrics for landscape genetics. Mol Ecol Resour 17:1308–1317

Short-Bull RA, Cushman S, Mace R, Chilton T, Kendall K, Landguth E et al. (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. Mol Ecol 20:1092–1107

Shrestha B, Kindlmann P (2020) Implications of landscape genetics and connectivity of snow leopard in the Nepalese Himalayas for its conservation. Sci Rep 10:19853

Stekhoven DJ (2013) missForest: nonparametric missing value imputation using random forest. R package version 1.4

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? Mol Ecol 19:3496–3514

Wagner HH, Fortin M-J (2012) A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14:253–261

Wagner HH, Fortin MJ (2016) Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, eds. Landscape genetics: concepts, methods, applications, 1st ed. John Wiley and Sons Ltd. Oxford, UK. pp. 77–98

Wahlund S (1928) Composition of populations and correlation appearances viewed in relation to the studies of inheritance. Hereditas 11:65–106

Wan HY, Cushman SA, Ganey JL (2019) Improving habitat and connectivity model predictions with multi-scale resource selection functions from two geographic areas. Landsc Ecol 34:503–519

Wasserman TN, Cushman SA, Schwartz MK, Wallin DO (2010) Spatial scaling and multi-model inference in landscape genetics: *Martes americana* in northern Idaho. Landsc Ecol 25:1601–1612

Weckworth B (2021) Snow leopard (*Panthera uncia*) genetics: the knowledge gaps, needs, and implications for conservation. J Indian I Sci 101:279–290

Wollenberg AL, van den (1977) Redundancy analysis. An alternative for canonical correlation analysis. Psychometrika 42:207–219

Wright S (1943) Isolation by distance. Genetics 28:114–138

Wright S (1946) Isolation by distance under diverse systems of mating. Genetics 31:39–59

Wu S, Qinye Y, Du Z (2003) Delineation of eco-geographic regional system of China. J Geogr Sci 13:309

Zeller KA, Jennings MK, Vickers TW, Ernest HB, Cushman SA, Boyce WM (2018) Are all data types and connectivity models created equal? Validating common connectivity approaches with dispersal data. Divers Distrib 24:868–879

Zhang Y, Hacker C, Zhang Y, Xue Y, Wu L, Dai Y et al. (2019) An analysis of genetic structure of snow leopard populations in Sanjiang-Yuan and Qilianshan National Parks. Acta Theriologica Sin 39:442–449

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

LA, SC, PR, SK and DB designed the study. WJ collected samples in YCW Nature Reserve. LA conducted laboratory and statistical analyses, and led the writing of the manuscript. DB supervised the statistical analyses relative to multivariate methods and the manuscript writing. SC supervised genetic analyses relative to spatially explicit indices and the manuscript writing. All authors actively discussed the ecological implications of the results and substantially contributed to the final version through comments and revisions.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-021-00483-0.

Correspondence and requests for materials should be addressed to Kun Shi.

Reprints and permission information is available at http://www.nature.com/reprints

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