

Habitat fragmentation affects genetic diversity and differentiation of the Yarkand hare

Yonghua Wu · Lin Xia · Qian Zhang ·
Qisen Yang

Received: 16 February 2009 / Accepted: 3 December 2009 / Published online: 24 December 2009
© Springer Science+Business Media B.V. 2009

Abstract The Yarkand hare, *Lepus yarkandensis*, is an endemic, endangered species restricted to the Tarim Basin of the Xinjiang Uygur Autonomous Region, China. The Yarkand hare is distributed in scattered oases which are physically isolated by the desert. Its natural fragmentation habitat makes it an ideal object for studying effect of habitat fragmentation on its genetic structure. To evaluate the effects of habitat fragmentation on genetic diversity of the species, we assessed genetic diversity for 20 sampling populations based on control region and Cytb markers. Relatively low levels of gene diversity are found in most of isolated populations in the southern margin of the Taklamakan Desert. Furthermore, a positive correlation is found between gene diversity and the size of historical effective population. Significant genetic differentiation is detected among most populations by pairwise F_{ST} analyses, which is characterized by an isolation by distance pattern. Additionally, the AMOVA results show highly significant population structure among seven geographical groups. High migration rates are found among continuous populations, while very low levels of migration rates are found among the relatively isolated populations, suggesting that the desert may make an effective barrier against gene flow. Finally, the control region shows four clades by the phylogenetic analyses, three of which are present in nearly all sampling populations. The observed pattern of the lineage mixing, also shown by the Cytb data, may be caused by extensive gene flow among populations, and could be explained by possible demographical expansion of the

Yarkand hare during the late Pleistocene interglacial period.

Keywords Control region · Cytb · Effective population size · Gene diversity · Demographical expansion

Introduction

Habitat fragmentation often results in generation of isolated populations. As a consequence, in small isolated populations, inbreeding may increase, genetic drift tends to predominate over mutation, selection and migration, making them prone to lose adaptive genetic diversity. Ultimately, the species may risk extinction due to loss of genetic diversity (Wright 1943; Frankham et al. 2002).

Small isolated populations are particularly susceptible to genetic drift and inbreeding, both of which are thought to reduce genetic diversity (Frankham et al. 2002; Gaggiotti 2003). The loss of genetic diversity in turn enhances inbreeding, and finally results in inbreeding depression (Reed and Frankham 2003). Low genetic diversity and/or inbreeding depression have been reported from many small and isolated populations, e.g., black-footed rock-wallaby (Eldridge et al. 1999), hairy-nosed wombat (Taylor et al. 1994; Beheregaray et al. 2000), red-cockaded woodpecker (Stangel et al. 1992; Daniels and Walters 2000), prairie chicken (Bouzat et al. 1998), adders (Madsen et al. 1996), and archaic reptile (Miller et al. 2008). Moreover, gene flow among isolated populations may reduce greatly; as a result, we may expect increased genetic differentiation among isolated populations.

To understand how habitat fragmentation affects genetic diversity and differentiation of isolated populations, a lot of

Y. Wu · L. Xia · Q. Zhang · Q. Yang (✉)
Key Laboratory of Zoological Systematics and Evolution,
Institute of Zoology, Chinese Academy of Sciences,
Datun Road, Chaoyang, Beijing 100101, China
e-mail: yangqs@ioz.ac.cn

studies have been done. However, due to relatively rare occurrences of ideal natural fragmented habitats, the issue is not well understood. The Yarkand hare, *Lepus yarkandensis*, is an ideal object for such study in that it is distributed in scattered oases around the Taklamakan Desert. The Yarkand hare is an endemic, endangered species restricted to the Tarim Basin, in southern Xinjiang Uygur Autonomous Region in northwest China. Surrounded by several ranges of mountains and with the Taklamakan Desert located in the center, the climate in the Tarim Basin is extremely arid. The main water resources of the oases are the streams descending from the surrounding mountains, in the form of seasonal flooding rivers by snow melting in mountains, and the rivers are crucial for formation and survival of these oases in the Tarim Basin. As

these rivers flow through different regions of the Tarim Basin, scattered oases are formed around the Taklamakan Desert in the Basin (Fig. 1a). Many oases in the Tarim Basin are physically isolated by the desert, and the oases at southern margin of Tarim Basin that are suffered from severe desert encroachment are being fragmented or buried by sand (Zhu et al. 1991; Fan 1993; Zhong and Xiong 1999; Yang 2001; Bruelheide et al. 2003; Zu et al. 2003).

The distribution of the Yarkand hare is limited to those scattered oases around the Taklamakan Desert in the Tarim Basin, as shown in Fig. 1a, and many populations of the Yarkand hare are physically isolated by desert from populations in other oases. To our current knowledge, no individual of the Yarkand hare have been found outside the Tarim Basin. Based on geological evidence, Sun et al. (2009) indicates that the Taklamakan Desert started to form at about 7 Ma, and then it evolved into present great landscape at around late Pliocene and Holocene (Zhand and Men 2002). The relatively long geological time of existence of the desert indicates that the population isolation of the Yarkand hare in different oases have lasted for a long time, and hence great differentiations among populations may be expected. Moreover, the different oases patches inhabited by the Yarkand hare have varied observed area sizes, so effective population sizes and genetic diversity may differ in those populations. The above features make the Yarkand hare an ideal object for studying the effect of habitat fragmentation on genetic structure.

At present time, the population size of the Yarkand hare is not only suffered from desert encroachment, but also has been affected by human factors like traffic, agricultural fields and hunting. So there is an urgent need to protect this species. It was listed in the Second Category of State Key Protected Wildlife List in 1988, and as Near Threatened in 1996 IUCN Red List of Threatened Animals (Wang 1998). The hunting is currently prohibited as a practical action.

To date, the biology details of the Yarkand hare, especially the population genetic structure, are not well documented. A previous phylogenetic analysis based on mtDNA data showed that the speciation of the Yarkand hare was at about 0.64 ± 0.26 Ma due to the event of peripheral isolate speciation (Wu et al. 2005). A recent mtDNA study (Li et al. 2006) indicated that there is pronounced genetic differentiation among populations, suggesting that the desert between oases makes an effective barrier against gene flow of the Yarkand hare. However, in the study by Li et al. (2006), only limited sampling sites and small sample sizes were used, so the results based on the incomplete data set may be statistically biased. In the present study, to estimate whether the habitat fragmentation could have a serious effect on the genetic diversity of this species, we use two mtDNA markers, and relatively large sample sizes and

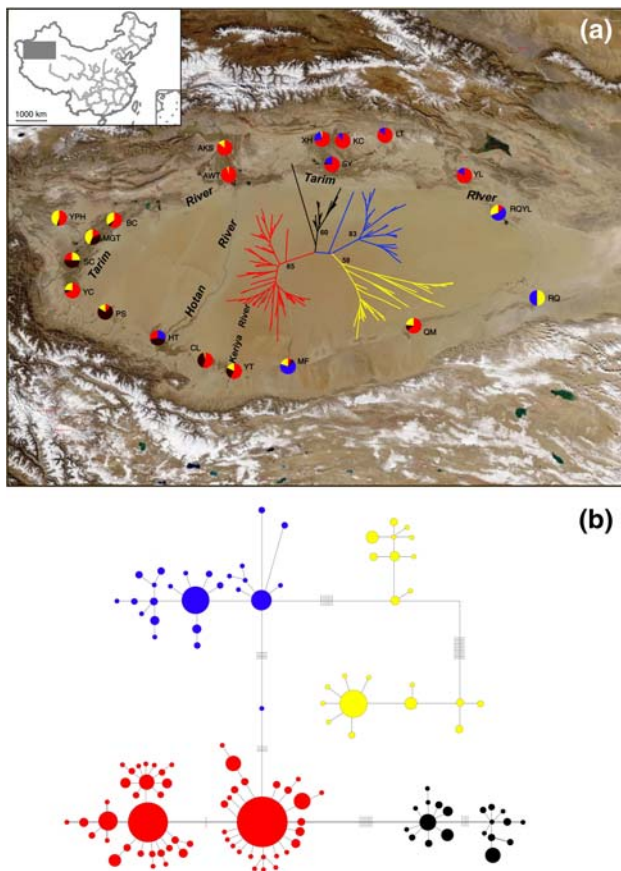


Fig. 1 The distribution of 20 sampling sites of *L. yarkandensis* and the phylogenetic relationships of mtDNA haplotypes. **a** represents the locations of sampling populations and neighbour-joining tree based on control region. The different color areas of the circle are proportional to the individual frequencies clustered in four different neighbour-joining phylogenetic clades based on control region. The gray square in the inset map indicates the location of research area in China map. **b** represents median-joining network for 123 Cytb haplotypes. The area of the circle is proportional to the haplotype frequency. The short lines in the Cytb network represent mutational steps (only primary mutational steps are shown)

sampling sites throughout the range of the Yarkand hare were acquired. Furthermore, the genetic differentiations among populations were also analyzed.

Materials and methods

Sampling

We collected about 518 different hare samples with certain geographical locations during our field sampling from 2002 to 2006. Among them, 468 hare samples were collected from local forest conservation field stations (including roadkills and confiscated specimens from hunters), 50 different hare samples were collected from roadkills. A total of 518 different hare samples are used in this study. These samples come from 20 sampling sites throughout the range of the Yarkand hare (Fig. 1a). For all samples obtained, the muscle or skin tissue samples were taken and preserved in 98% ethanol.

Molecular experiments

Genomic DNA was isolated from ethanol-preserved muscle or skin tissue using genomic DNA extraction kits (Sangon, Beijing). Control region was amplified as described by Li et al. (2006) and the primer L15926 was used for sequencing. Control region sequences were successfully obtained in 518 hares (Table 1), and a 535 bp alignment was constructed, including 60 bp tRNA and 475 bp of mtDNA control region. Overall, 191 haplotypes were detected among 518 individuals. The haplotypes obtained in this study were submitted to GenBank with accessions (EU734186-90, EU734192-95, EU734198-223, EU734225-31, EU734234-35, EU734237-41, EU734243-50, EU734258-60, EU734262-67, EU734269-82, EU734290-95, EU734297-308, EU734310-26, EU734338-47, EU734350, EU734352-64, EU734368-89).

For Cytb, the PCR primers were modified from Irwin et al. (1991), including L14724: 5'-GAT ATG AAA AAC CAT CGT TG-3', H15915: 5'-CGG ATT TCC ATT TTT GGT TTA CAA GAC-3'. The Cytb DNA was amplified in similar conditions like the control region except that they were annealed at 52°C for 1 min, and extended at 72°C for 1.5 min. Overall, a total of 501 hares were successfully sequenced (Table 1). Complete Cytb sequences (1140 bp) were aligned and analyzed. In total, 123 haplotypes were detected and submitted to GenBank with accessions (EU729130-66, EU729169-70, EU729172, EU729174-76, EU729183-91, EU729197-221, EU729225-27, EU729229-30, EU729232-44, EU729247-58, EU729262, EU729265-66, EU729268-71, EU729273, EU729275-76, EU729280, EU729283-87).

Data analyses

All sequences were firstly checked for nucleotide mismatch, and then aligned using the DNAMAN software (version 5.2.2.0, Lynnon Biosoft, USA). To examine genetic diversity for each population, the nucleotide diversity and gene diversity based on the control region and Cytb data were calculated using Arlequin software version 3.1. (Excoffier et al. 2005). Tests of mutation-drift equilibrium (Tajima 1989) for each population and entire data set were also conducted using the same software.

To examine genetic differentiations among 20 populations, the pairwise F_{ST} values among these populations were calculated using Arlequin software. Additionally, the correlations between genetic $F_{ST}/(1 - F_{ST})$ and geographical distances were tested (Rousset 1997) using a Mantel test implemented in the software IBD on the web (<http://ibdws.sdsu.edu/>). For this analysis, the geographical distances between neighboring populations were measured by their shortest straight-line distances in the map, and the geographical distances between non-neighboring populations were calculated as the sum of geographical distances between neighboring populations. Moreover, given the “ring”-shape habitat chain of the Yarkand hare, all populations were divided into two sub-data sets and analyzed separately: one sub-data set includes 13 populations located in the Tarim River drainage (YC, SC, MGT, YPH, BC, AWT, AKS, LT, KC, XH, SY, YL, RQYL); and the other sub-data set includes eight populations (YC, PS, HT, CL, YT, MF, QM, RQ). Taken into consideration the possibility that the population YC could have gene flow with its neighboring population PS, the population YC was also included in the second sub-data set, which also increased the sample size of the second sub-data set.

The genetic structure was examined by performing a hierarchical analysis of molecular variance (AMOVA, Excoffier et al. 1992) implemented in the Arlequin software. For this analysis, 20 sampling populations were pooled into seven geographical groups according to their geographical isolation as follows: northern populations group (LT, KC, XH, SY, AKS, AWT), western populations group (BC, MGT, YPH, SC, YC), eastern populations group (YL, RQYL) and four southern populations groups: (PS), (HT, CL, YT), (MF), (QM, RQ). The seven geographical groups locate in seven relatively separated oases, respectively, and they are greatly isolated from each other by desert (Fig. 1a). For example, the populations PS and MF were treated as two separate groups due to their considerable isolation from their neighboring populations.

For control region data, historical migration rates and effective population size (N_e) were estimated using maximum-likelihood approach based on coalescence theory implemented in software Migrate (Beerli and Felsenstein

Table 1 The sample information for 20 sampling sites with abbreviations based on control region (Dloop) and Cytb data

Sampling sites	Dloop		Cytb		H _d		π		D		Dloop		N _e
	N	n	N	n	Dloop	Cytb	Dloop	Cytb	Dloop	Cytb	Θ	N _e	
	Yuli (YL)	30	26	30	21	0.99 ± 0.01	0.97 ± 0.02	0.025 ± 0.012	0.004 ± 0.002	-0.34	-1.65*	0.05460	
Luntai (LT)	21	12	25	15	0.91 ± 0.05	0.93 ± 0.03	0.022 ± 0.012	0.005 ± 0.003	-0.34	-1.05	0.01260	101613	
Kuqa (KC)	33	17	31	15	0.95 ± 0.02	0.94 ± 0.02	0.023 ± 0.012	0.003 ± 0.002	0.22	-0.98	0.00755	60887	
Xinhe (XH)	28	16	30	12	0.95 ± 0.02	0.91 ± 0.03	0.027 ± 0.014	0.005 ± 0.003	-0.04	-0.82	0.01231	99274	
Xayar (SY)	52	28	52	30	0.97 ± 0.01	0.96 ± 0.02	0.027 ± 0.014	0.005 ± 0.003	-0.22	-1.57*	0.02599	209597	
Awat (AWT)	24	8	24	6	0.84 ± 0.04	0.63 ± 0.10	0.015 ± 0.008	0.003 ± 0.002	-0.59	-2.12**	0.00099	7984	
Aksu (AKS)	7	6	7	6	0.95 ± 0.10	0.95 ± 0.10	0.022 ± 0.013	0.004 ± 0.002	-0.62	-0.99	0.03502	282419	
Bachu (BC)	34	20	32	13	0.95 ± 0.02	0.90 ± 0.03	0.027 ± 0.014	0.010 ± 0.005	0.34	0.67	0.02277	183629	
Markit (MGT)	32	17	25	8	0.90 ± 0.05	0.83 ± 0.05	0.028 ± 0.014	0.012 ± 0.006	0.18	1.72	0.01105	89113	
Yopurga (YPH)	20	13	20	6	0.95 ± 0.03	0.81 ± 0.05	0.028 ± 0.015	0.010 ± 0.005	0.99	1.23	0.02186	176290	
Shache (SC)	12	8	12	8	0.92 ± 0.06	0.92 ± 0.06	0.027 ± 0.015	0.011 ± 0.006	0.39	0.52	0.00325	26210	
Yecheng (YC)	17	7	16	4	0.72 ± 0.11	0.58 ± 0.12	0.022 ± 0.012	0.009 ± 0.005	0.22	0.64	0.00255	20565	
Pishan (PS)	19	8	19	7	0.85 ± 0.05	0.84 ± 0.05	0.018 ± 0.009	0.008 ± 0.004	-0.36	-0.04	0.01035	83468	
Hotan (HT)	25	12	25	11	0.92 ± 0.03	0.92 ± 0.02	0.025 ± 0.013	0.006 ± 0.003	1.56	0.84	0.01489	120081	
Qira (CL)	31	5	30	5	0.62 ± 0.06	0.62 ± 0.06	0.019 ± 0.010	0.007 ± 0.004	-0.02	1.36	0.00146	11774	
Yutian (YT)	33	10	34	8	0.78 ± 0.07	0.67 ± 0.08	0.025 ± 0.013	0.009 ± 0.005	0.84	0.56	0.00521	42016	
Minfeng (MF)	30	12	23	6	0.71 ± 0.09	0.56 ± 0.11	0.017 ± 0.009	0.006 ± 0.003	-0.96	-0.73	0.00553	44597	
Qiemo (QM)	17	7	18	8	0.77 ± 0.10	0.88 ± 0.05	0.030 ± 0.016	0.009 ± 0.005	-0.1	-0.49	0.00532	42903	
Ruoqiang (RQ)	20	12	20	9	0.93 ± 0.03	0.84 ± 0.06	0.024 ± 0.013	0.005 ± 0.003	1.41	-0.16	0.00723	58306	
Ruoqiangyuli (RQYL)	33	21	28	14	0.95 ± 0.02	0.89 ± 0.04	0.032 ± 0.016	0.005 ± 0.003	0.12	-0.42	0.03314	267258	
ALL	518	191	501	123	0.99 ± 0.00	0.96 ± 0.00	0.031 ± 0.015	0.008 ± 0.004	-0.64	-1.58*	-	-	

Number of individuals (N), number of haplotypes (n), gene diversity (H_d), and nucleotide diversity (π) are shown. The values of the Tajima's D are shown with statistical tests of significance. The Θ values and their corresponding effective population sizes (N_e) were calculated by maximum-likelihood methods based on the control region data

* P < 0.05, ** P < 0.001

1999, 2001). The maximum-likelihood approach is generally used to estimate long-term asymmetric migration rates and inbreeding effective population sizes (Pearse and Crandall 2004). The Θ value obtained by the maximum-likelihood approach was used to calculate effective population size by the equation $\Theta = N_e \mu$, where N_e is the effective female population size for mtDNA markers and μ is the mutation rate per generation per site. For the Yarkand hare, $\mu = 1.24 \times 10^{-7}$, which is specific for control region of *Lepus* (Pierpaoli et al. 1999). In this analysis, we used 20 short chains sampling 2,500,000 genealogies, among which, 5,000 were used for the parameter estimation, and five long chains sampling 25,000,000 genealogies, 50,000 of which were used for estimation. First genealogy was started using a random tree, and F_{ST} was used as a starting value for the maximum likelihood estimates. To validate the result, two independent runs (run with replicates = 5) with different random numbers were performed and they produced similar results (paired-samples T test, $t = 0.788$, $df = 47$, $P = 0.435$); the maximum likelihood estimate of final run was reported.

To examine the population demography of the Yarkand hare, the mismatch distribution and Fu's F_s values based on the control region and Cytb data were analyzed separately using Arlequin software. For the mismatch distribution analyses, 1,000 replicates were used, and a test statistic SSD (the sum of squared differences) between the observed and expected distribution under population expansion was used to test the validity of the estimated demographic expansion. Usually, populations at demographic equilibrium could show multimodal distribution, whereas populations which have gone through a recent demographic expansion are expected to be unimodal (Rogers and Harpending 1992). The demographic expansion time was estimated using equations $\tau = 2ut$ and $u = \mu k$, where u is the mutation rate for the whole sequence, t is the number of generations since the expansion, and k is the length of sequence used. For control region analyses, we used $\mu = 1.24 \times 10^{-7}$ (substitutions per nucleotide per generation). In the Cytb analyses, we used the molecular clock 4% sequence divergence per Ma (Pierpaoli et al. 1999). The length of the Yarkand hare generation is around 1 year according to the study of Chinese hare breeding biology (Luo 1988; Lu 1995). For the Fu's F_s analyses, the significance of F_s was based on 1,000 simulated samples.

To examine the phylogenetic structure of the Yarkand hare, the phylogenetic relationships based on the control region and Cytb data were constructed separately. For control region data, phylogenetic relationship was constructed using Neighbour-Joining (NJ) method. The NJ tree was constructed based on the kimura-two parameter model implemented in Mega software version 4.0 (Tamura et al. 2007) with 1,000 bootstrap replicates. For Cytb haplotypes,

the median-joining (MJ) method was used to construct network with epsilon equal to zero. The MJ network was constructed using the NETWORK software version 4.5.0.0 (Bandelt et al. 1999) with epsilon equal to zero.

Results

Genetic diversity

Based on control region and Cytb data, many haplotypes are shared by neighboring populations in the northern, western and eastern areas, while few haplotypes are shared among southern populations. On the contrary, the private haplotypes dominate in the southern populations. For example, the control region result showed that, for the population RQ, eight of 12 haplotypes were private haplotypes. Similarly, six of seven in population QM, nine of 12 in population MF, eight of ten in population YT, two of five in population CL, 10 of 12 in population HT and five of eight in population PS were private haplotypes.

To examine the genetic diversity level for each population, we calculated gene diversity and nucleotide diversity based on the control region and Cytb data, respectively (Table 1). The gene diversity values ranged from 0.62 in population CL to 0.99 in population YL based on the control region data, and from 0.56 in population MF to 0.97 in population YL based on the Cytb data. The nucleotide diversity values ranged from 0.015 in population AWT to 0.032 in population RQYL based on the control region data, and from 0.003 in population AWT to 0.012 in population MGT based on the Cytb data. The global gene diversity of the entire data set was 0.99 based on the control region data, and 0.96 based on the Cytb data. The nucleotide diversity was 0.031 based on the control region data, and 0.008 based on the Cytb data.

Moreover, our results show that the southern populations with small habitat patches tend to exhibit relatively low gene diversities and small effective population sizes. First, gene diversity values in five of seven southern populations (PS, CL, YT, MF, QM) are much lower than nearly all other populations. To further analyze whether there is a significant difference in gene diversity between the southern populations (including seven populations: PS, HT, CL, YT, MF, QM, RQ) and other populations, we conducted statistical t -test between them. And the results based on the control region showed significant lower level of gene diversity in the southern populations than other populations ($t = 2.995$, $df = 18$, $P = 0.004$), while the Cytb result showed no significant difference between them ($t = 1.610$, $df = 18$, $P = 0.062$). Furthermore, the nucleotide diversity results showed no significant difference between southern populations and other populations (data not shown).

In addition to gene diversity results, the southern populations tend to have low N_e values calculated from software Migrate (Table 1). So we further analyzed the correlation between N_e and gene diversity based on control region, and the highly significant positive correlation between N_e and gene diversity was detected (Spearman's $r = 0.805$, $P = 0.000$; Fig. 2).

Genetic differentiation

To determine the genetic differentiation among 20 sampling populations, we first calculated their pairwise F_{ST} values based on the control region and Cytb data. The majority of populations have significant genetic differentiation (Table 2). While the continuous populations tend to show no significant genetic differentiation, most of the isolated populations have significant or highly significant genetic differentiation.

To examine whether there is a pattern of isolation by distance over large geographical scale, we further analyzed correlation between genetic and geographical distances with the Mantel test. Almost all results based on control region and Cytb data show significant positive correlation between genetic and geographical distances. For one sub-data set including 13 populations, both the control region (Fig. 3a) and Cytb results (Fig. 3b) showed highly significant positive correlations between genetic and geographical distances (for control region, $P = 0.001$, $r = 0.502$, $Z = 17120.140$; for Cytb, $P = 0.000$, $r = 0.520$, $Z = 16373.115$). For the other sub-data set including eight populations, the Cytb result (Fig. 3d) showed significant positive correlation between genetic and geographical distances ($P = 0.034$, $r = 0.472$, $Z = 3747.426$), while the control region result (Fig. 3c) showed no significance between them ($P = 0.060$, $r = 0.426$, $Z = 6519.593$).

The genetic structure of seven geographical groups was examined by the hierarchical AMOVA. Both the control region and Cytb results indicate highly significant differentiations at three different levels of hierarchical AMOVA

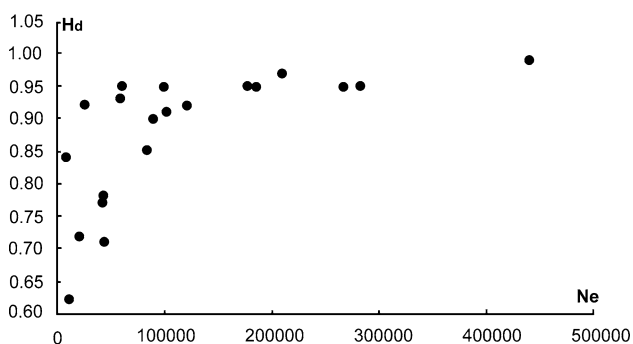


Fig. 2 Scatterplot between gene diversity (H_d) and effective population size (N_e) based on the control region data

(Table 3). First, differentiations among seven geographical groups explained 11.89% variation based on control region ($F_{CT} = 0.12$, $P < 0.001$) and 16.21% variation based on Cytb ($F_{CT} = 0.16$, $P < 0.001$). Second, differentiations among populations within groups explained 12.69% variation based on control region ($F_{SC} = 0.14$, $P < 0.001$) and 6.71% variation based on Cytb ($F_{SC} = 0.08$, $P < 0.001$). Differentiations within populations explained the rest of variation, which accounted for the majority part of variation (for control region, $F_{ST} = 0.25$, $P < 0.001$; for Cytb, $F_{ST} = 0.23$, $P < 0.001$).

The phylogenetic structures among haplotypes of control region and Cytb were constructed separately. The results based on the control region data (Fig. 1a) showed four separate clades. According to the control region result, clade 1 (red) scattered throughout the 20 sampling populations with only one exception (the population RQ), clade 2 (black) was present in partial southern populations, and the other two clades (blue and yellow) scattered among the majority of the sampling populations. Note that the clade 3 (blue) was not present in the western populations, and the clade 4 (yellow) was not present in several northern and southern populations. Similar phylogenetic structure was constructed from the Cytb data except that those haplotypes clustered within the yellow clade of control region were separated into two divergent subclades (indicated by yellow color) in the Cytb network (Fig. 1b). The geographical distribution of the Cytb phylogenetic clades (indicated by four different colors) is consistent with that of control region. Intriguingly, from the Cytb network, the two divergent subclades within the yellow clade showed quite different geological distribution except that they overlapped in two populations (YT and QM): one subclade is mainly distributed in populations located at eastern area of old Keriya River (LT, XH, RQYL, RQ, QM, CL, YT), but the other one occurred mainly in populations located at western area of the River (AWT, BC, MGT, YPH, SC, YC, PS, YT, MF, QM).

Gene flow among populations

To examine whether the desert could make an effective barrier against gene flow, we analyzed migration rates among populations using the maximum-likelihood approach. Based on the control region, highly asymmetric migration rates among populations are detected (Fig. 4), and the continuous populations tend to present high migration rates, while the isolated populations show very low migration rates.

First, within the northern population group, high migration rates are found among continuous populations, while very low levels of migration rates are found among the relatively isolated populations. For examples, high

Table 2 Pairwise F_{ST} values among 20 populations based on the control region (below the diagonal) and Cytb data (above the diagonal)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1 YL		0.032	0.017	0.091	<i>0.030</i>	0.065	<i>0.105</i>	0.191	0.340	0.336	0.290	0.229	0.442	0.267	0.196	0.118	0.395	0.100	0.449	0.340
2 LT	0.028		0.052	0.099	<i>0.046</i>	0.098	<i>0.117</i>	0.150	0.287	0.276	0.231	0.177	0.395	0.242	0.175	0.094	0.334	<i>0.084</i>	0.371	0.277
3 KC	0.014	<i>0.05</i>		0.046	0.004	<i>0.043</i>	0.027	0.186	0.348	0.304	0.304	0.223	0.455	0.279	0.211	0.131	0.412	0.136	0.472	0.369
4 XH	0.077	0.150	0.032		0.001	0.104	0.056	0.144	0.282	0.277	0.214	0.163	0.377	0.193	0.194	0.116	0.270	0.118	0.346	0.238
5 SY	0.022	0.066	-0.004	0.003		0.058	0.039	0.164	0.323	0.308	0.246	0.183	0.399	0.204	0.188	0.113	0.305	0.109	0.367	0.258
6 AWI	0.115	0.206	<i>0.056</i>	0.114	0.081		0.033	0.160	0.328	0.306	0.301	0.190	0.461	0.317	0.221	0.135	0.442	0.151	0.493	0.401
7 AKS	0.049	<i>0.116</i>	-0.012	0.008	-0.014	0.049		0.141	0.274	0.260	0.227	<i>0.156</i>	0.402	0.287	<i>0.193</i>	<i>0.118</i>	0.417	<i>0.125</i>	0.465	0.391
8 BC	0.137	0.173	0.104	0.091	0.080	0.147	0.028		<i>0.054</i>	-0.001	0.039	-0.040	0.238	0.211	0.206	0.096	0.174	<i>0.080</i>	0.232	0.199
9 MGT	0.249	0.265	0.237	0.201	0.184	0.332	<i>0.157</i>	0.076		0.039	-0.024	0.049	<i>0.117</i>	0.203	0.248	0.155	0.159	0.166	0.237	0.242
10 YPH	0.183	0.208	0.158	0.126	0.118	0.239	0.079	-0.005	<i>0.060</i>		0.082	-0.020	0.297	0.323	0.306	0.186	0.221	<i>0.148</i>	0.267	0.264
11 SC	0.215	0.247	0.211	0.184	0.155	0.327	<i>0.132</i>	0.088	-0.019	<i>0.092</i>		0.048	0.033	0.066	<i>0.146</i>	0.057	<i>0.116</i>	<i>0.106</i>	0.235	0.214
12 YC	0.183	0.237	0.160	0.154	0.130	0.240	0.104	0.104	0.193	0.086	<i>0.191</i>		0.265	0.250	0.227	<i>0.107</i>	0.202	0.080	0.257	0.224
13 PS	0.327	0.360	0.318	0.289	0.250	0.460	0.310	0.230	0.087	0.245	0.027	0.355		0.135	0.249	0.199	0.288	0.275	0.391	0.381
14 HT	0.186	0.237	0.204	0.160	0.141	0.340	0.193	0.181	0.122	0.190	<i>0.071</i>	0.236	0.118		<i>0.106</i>	<i>0.074</i>	0.231	0.187	0.360	0.289
15 CL	0.142	0.187	0.173	0.214	0.147	0.294	<i>0.180</i>	0.200	0.202	0.235	<i>0.141</i>	0.228	0.241	0.135		0.029	0.316	0.141	0.370	0.317
16 YT	0.113	0.117	0.129	0.191	0.123	0.229	<i>0.136</i>	0.150	0.185	0.185	<i>0.139</i>	0.205	0.241	0.152	0.103		0.208	0.056	0.273	0.218
17 MF	0.435	0.467	0.434	0.355	0.342	0.579	0.475	0.365	0.289	0.368	0.349	0.461	0.396	0.294	0.485	0.423		0.206	0.186	0.109
18 QM	<i>0.062</i>	0.128	0.112	0.156	0.105	0.228	<i>0.113</i>	0.146	0.219	0.174	0.195	0.206	0.333	0.207	0.161	0.127	0.437		0.201	0.160
19 RQ	0.481	0.508	0.484	0.411	0.415	0.594	0.475	0.396	0.341	0.375	0.380	0.479	0.463	0.380	0.511	0.458	0.365	0.396		0.032
20 ROYL	0.326	0.338	0.331	0.258	0.267	0.441	0.300	0.260	0.204	0.231	0.221	0.319	0.295	0.211	0.356	0.320	0.161	0.267	0.050	

Bold/italic characters indicate $P < 0.01$, $P < 0.05$, respectively

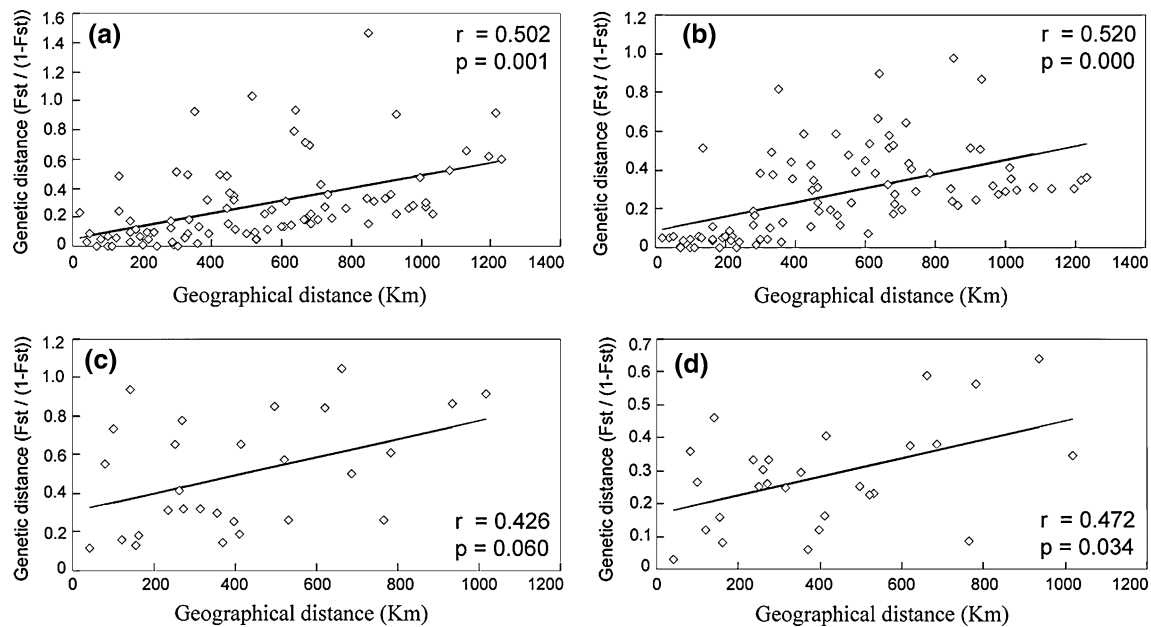


Fig. 3 Genetic distance $F_{ST}/(1 - F_{ST})$ against geographical distance (Km). The correlation coefficient (r) and significance level (p) are also shown. For the first sub-data set including 13 populations, the

results based on the control region (a) and Cytb (b) are shown. For the second sub-data set including eight populations, the results based on the control region (c) and Cytb (d) are shown

Table 3 Analyses of molecular variance based on the control region and Cytb data

Source of variation	d.f.	Percentage of variation	Fixation indices
Control region			
Among groups	6	11.89	$F_{CT} = 0.12^{***}$
Among populations			
Within groups	13	12.69	$F_{SC} = 0.14^{***}$
Within populations	498	75.42	$F_{ST} = 0.25^{***}$
Cytb			
Among groups	6	16.21	$F_{CT} = 0.16^{***}$
Among populations			
Within groups	13	6.71	$F_{SC} = 0.08^{***}$
Within populations	481	77.08	$F_{ST} = 0.23^{***}$

*** $P < 0.001$

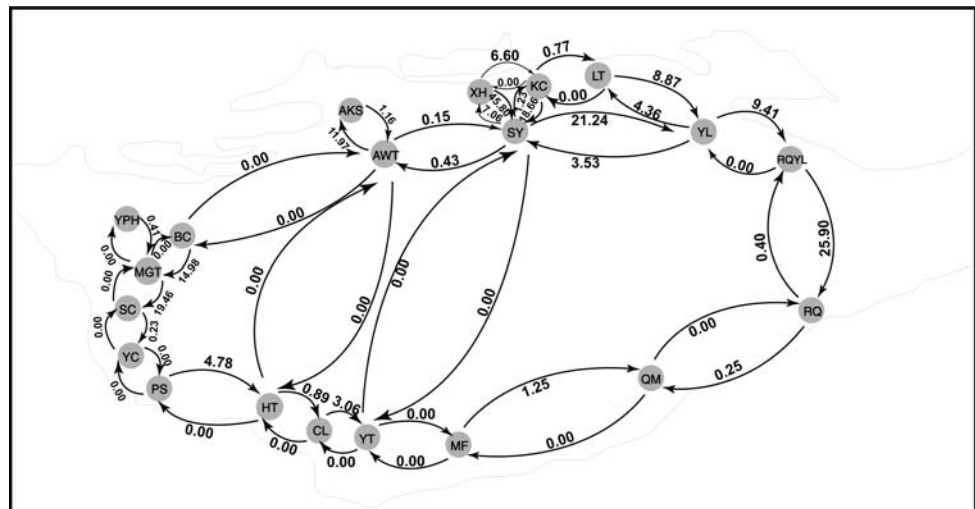
migration rates (11.97) were found from populations AWT to AKS, which are two continuous populations within a relatively separated area. Similarly, high migration rates were also observed among three continuous populations (KC, XH and SY) within another relatively separated area: 45.60 from populations XH to SY, 7.06 from populations SY to XH, 6.60 from populations XH to KC, and 18.66 from populations KC to SY. Unlike those continuous populations, very low levels of migration rates were observed for isolated populations within the northern population group: 0.15 from populations AWT to SY, 0.43 from populations SY to AWT, and 0.77 (unidirectional migration rate) from populations KC to LT.

Second, within the western population group, highly unidirectional migration rates are found between three continuous populations BC, MGT and SC; however, two relatively isolated populations YPH and YC present low migration rates to their neighboring populations MGT and SC, respectively. For the population YPH, which is relatively isolated from its neighboring population MGT due to desert existence between them, only low unidirectional migration rate (0.41) was found from YPH to MGT. Similarly, for the population YC, which is relatively isolated by the Tarim River from its neighboring population SC, only lowly unidirectional migration rate (0.23) was found between them.

Finally, for two eastern populations YL and RQYL, high migration rates are found among them and their neighboring populations. For the southern populations, low unidirectional migration rates are found between nearly all of them (e.g., 4.78 from populations PS to HT, 0.89 from populations HT to CL, 3.06 from populations CL to YT, zero from populations YT to MF, 1.25 from populations MF to QM, and 0.25 from populations RQ to QM).

Furthermore, we also analyzed migration rates between populations along two rivers: Hotan River and Keriya River (Fig. 1a). The Hotan River flows across the desert from the south to the north, and the Keriya River, a dry river at present time, used to flow through the Taklamakan Desert. The migration rates results showed that there were zero migration rates between populations HT and AWT along the Hotan River, populations YT and SY along the old Keriya River, respectively. Similarly, we also found

Fig. 4 Migration rates among populations of *L. yarkandensis* based on control region data



zero migration rates among other populations, like populations BC and AWT, populations YC and PS, populations YT and MF.

Population demography

To examine population demography of the Yarkand hare, we performed mismatch distribution analyses and Fu’s F_s tests based on the control region (Fig. 5a) and Cytb data (Fig. 5b). The mismatch distributions fit unimodal curves typically associated with recent population expansion (for control region, $SSD = 0.0027$, $P = 0.674$; for Cytb, $SSD = 0.0025$, $P = 0.936$), and the demographic expansion is further supported by significant Fu’s F_s values (for control region, Fu’s $F_s = -23.47$, $P = 0.009$; for Cytb, Fu’s $F_s = -23.90$, $P = 0.001$). The demographic expansion time for the Yarkand hare is 0.172 Ma (95% CI 0.104–0.217 Ma) and 0.143 Ma (95% CI 0.018–0.707 Ma) based on the control region and Cytb data, respectively.

Discussion

Positive correlation between gene diversity and effective population size

Based on our mtDNA result, the Yarkand hare populations with small habitat patches, e.g., southern populations, tend to present relatively low gene diversity and small effective population size (N_e). We also detected positive correlation between gene diversity and effective population size in the Yarkand hare (Fig. 2), like what have been found in many plant and other animal species (Soule 1976; Frankham 1996; Knaepkens et al. 2004).

The positive correlation between gene diversity and N_e obtained in this study is characterized by a convex curve

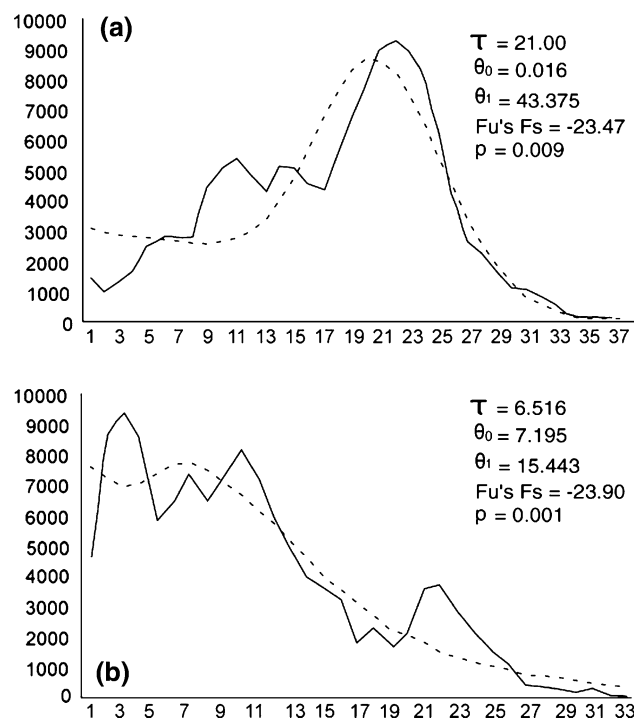


Fig. 5 The demography analyses of mismatch distribution and Fu’s F_s for *L. yarkandensis* based on the control region (a) and Cytb (b) data. The observed and expected mismatch distributions are indicated by solid and dashed lines, respectively. The θ_0 and θ_1 represent pre-expansion and postexpansion population sizes, respectively. The τ is the date of the growth or decline measured in units of mutational time. The statistical significance (p) of Fu’s F_s value is also shown

(Fig. 2), which is consistent with theoretical expectation. According to previous theory work (Crow and Kimura 1970), at equilibrium between neutral mutation and drift, the relationship between heterozygosity (H) and N_e is indicated by Eq. 1, suggesting that the heterozygosity increases with the increase of N_e , and a convex-curve correlation between the two is predicted by the equation.

The heterozygosity is equivalent to the gene diversity in a randomly mating diploid population, both of which are calculated by Eq. 2 (n is sample size, X_i is the frequency of the i -th haplotype) (Nei 1987). Hence, in our case, the relationship between gene diversity and N_e can also be depicted by the Eq. 1, with a convex-curve correlation like what we plotted from our control region data which meets the equilibrium assumption with no significant Tajima's D values detected (Table 1).

$$H = 4N_e\mu / (4N_e\mu + 1) \quad (1)$$

(Crow and Kimura 1970)

$$H = [n/(n-1)](1 - \sum X_i^2) \quad (2)$$

(Nei 1987)

Therefore, the effective population size (N_e) may be an important factor that affects gene diversity, and a certain N_e may be necessary for a population to maintain adaptive genetic diversity. As suggested by the concept of minimum viable population sizes (MVP), it is necessary to maintain the smallest population size so that it can exist without facing extinction from demographic, environmental and genetic stochasticity (Shaffer 1981). Previous study by Lynch and Lande (1998) reported the minimum effective population sizes should be in the order of 1,000–5,000 to maintain adaptive genetic variation, to avoid the accumulation of mildly deleterious mutations. In addition, a median MVP of 5816 adults was reported by Reed et al. (2003), and a recent median MVP of 4169 individuals was reported by Traill et al. (2007) based on a meta-analysis of many literaturally reported species.

Strong genetic differentiations among most populations

The F_{ST} analyses revealed significant genetic differentiations among most populations (Table 2), especially for those isolated populations in the southern margin of the basin, in consistence with the results obtained by Li et al. (2006). Similar strong genetic differentiations among fragmented populations have been reported in many other species, like red-cockaded woodpecker, *Picoides borealis* (Stangel et al. 1992), protozoan parasite, *Plasmodium falciparum* (Anthony et al. 2005), wind-pollinated tree, *Juniperus communis* (Provan et al. 2008). Highly significant differentiations among seven geographical groups were also detected by our AMOVA analyses (Table 3).

The strong genetic differentiations among most populations of the Yarkand hare indicate that the desert may make an effective barrier for population dispersal. This is further supported by our migration rates analyses. It was detected that the continuous populations tend to show high migration rates, while the isolated populations have very low migration rates (Fig. 4). Moreover, our analyses of

isolation by distance showed that genetic distances significantly increase as the geographical distances increase. So the strong genetic differentiations among most populations may also attribute to spatially restricted gene flow over large geographic scales (Fig. 3).

The genetic drift could have a slight effect on differentiations of the Yarkand hare, especially for the southern isolated populations, which may be supported by the lack of shared haplotypes among the southern isolated populations. Of course, the lack of the shared haplotypes among isolated populations may also directly attribute to restricted gene flow among them (Schmitt 1978; Robinson et al. 1996; Seddon and Baverstock 1999).

Despite of strong genetic differentiations among populations and geographical groups, we haven't found obvious phylogenetic structure corresponding to geographical areas based on phylogenetic analyses (Fig. 1a and b). A possible explanation is that the extensive gene flow among populations might occur during ancient geological period. This may be supported by two pieces of evidence observed in this study. The first piece of evidence to support the extensive gene flow among populations is that some haplotypes are shared by distant populations. For example, for the control region data, the haplotype H69 was shared by populations MF, RQYL and LT; the haplotype H116 was shared by populations RQ, RQYL and SY; the haplotype H78 was shared by populations MF and MGT; the haplotype H7 was shared by populations CL and AWT. The most common haplotype H4 was found in seven northern populations. Similarly, for the Cytb data, four widespread haplotypes are found to scatter in many populations throughout the range of the Yarkand hare.

Another piece of evidence to support the extensive gene flow among populations is demographical expansion of the Yarkand hare. Based on our mismatch distribution analyses (Fig. 5), the time of the Yarkand hare demographical expansion could be potentially linked to the late Pleistocene interglacial period (0.07–0.15 Ma) (Shi et al. 2006). The demographical expansion might cause range expansion of the Yarkand hare, followed by extensive gene flow among populations.

In addition, some geological evidence suggests that the previous habitats of the Yarkand hare could be more continuous than their current habitats. For example, according to fluvial deposits and older riverbed of the Keriya River, a dry river at present time, it used to flow through the Taklamakan Desert and reached the Tarim River of the northern fringe of the Tarim Basin in several different geological times, with quite abundant total runoff during the late period of the last glaciation, 2000a B.P. and the little ice ages owing to increased melting water of the glaciers and snows (Yang 2001; Yang et al. 2002). Hence, this river could have been a basis for the formation of oases

and green oasis corridors where hares may disperse along, causing extensive gene flow among populations along it. However, we should note that although the Keriya River might promote historical gene flow between northern and southern populations, it could also act as a barrier for gene flow between eastern and western populations, supported by the occurrence of different geographical distribution of the two divergent subclades (yellow) observed in our Cytb network.

Conservation implications

The conservation of genetic diversity has emerged as one of the central issues in conservation biology (Frankham et al. 2002). It is thought to be necessary to maintain abundant genetic diversity in a population so that it can exist without facing extinction from environmental changes (Shaffer 1981; Frankham et al. 2002). In the present study, we estimated the effects of habitat fragmentation on the genetic diversity of the Yarkand hare. The mtDNA result showed that those populations with small habitat patches tend to exhibit relatively low gene diversities and small N_e , and a significant positive correlation between gene diversity and N_e is found. Furthermore, the migration rate result based on control region showed that continuous populations tend to have high migration rates while those isolated populations, especially southern populations, have very low values. Usually, gene flow between populations is thought to increase gene diversity in a population. Based on these results, we can conclude that it is necessary to maintain a relatively large and continuous habitat to keep a population high gene diversity. For the Yarkand hare, because those southern populations with small habitat patches have relatively low gene diversities and limited gene flow, they are in more danger of extinction facing future environment changes and more attention should be paid to them in further conservation. The habitat protection of the Yarkand hare should be taken into consideration for further road and agricultural field construction, especially in the southern area. Local hunting should always be prohibited and roadkills should be reduced.

Acknowledgments We are grateful to Haifeng Wang for helping with this paper's writing. Two anonymous reviewers provided very useful comments on the manuscript. This research was supported by the National Natural Science Foundation of China (Project No. 30370212) and CAS Innovation Program (KSCX3-IOZ-0810).

References

- Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, Shamsul S, Singh B (2005) Fragmented population structure of plasmodium falciparum in a region of declining endemicity. *J Infect Dis* 191(9):1558–1564
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of migration rates and effective population numbers in n subpopulations using a coalescent approach. *Proc Natl Acad Sci USA* 98:4563–4568
- Beheregaray LB, Sunnucks P, Alpers DL, Taylor AC (2000) A set of microsatellite loci for the hairy-nosed wombats (*Lasiorhinus krefftii* and *L. latifrons*). *Conserv Genet* 1:89–92
- Bouzat JL, Lewin HA, Paige KN (1998) The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. *Am Nat* 152:1–6
- Bruelheide H, Jandt U, Gries D, Thomas FM, Foetzki A, Buerkert A, Wang G, Lan Z, Zhang XM, Runge M (2003) Vegetation changes in a river oasis on the southern rim of the Taklamakan Desert in China between 1956 and 2000. *Phytocoenologia* 33:801–818
- Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row Publishers, New York
- Daniels SJ, Walters JR (2000) Inbreeding depression and its effects on natal dispersal in red-cockaded woodpeckers. *Condor* 102:482–491
- Eldridge MDB, King JM, Loupis AK, Spencer PBS, Taylor AC, Pope LC, Hall GP (1999) Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby. *Conserv Biol* 13:531–541
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform* 1:47–50
- Fan ZL (1993) A study on the formation and evolution of oases in Tarim basin. *Acta Geophys Sinica* 48:421–427
- Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conserv Biol* 10:1500–1508
- Frankham R, Ballou JD, Briscoe DA (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge
- Gaggiotti OE (2003) Genetic threats to population persistence. *Ann Zool Fenn* 40:155–168
- Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome b gene of mammals. *J Mol Evol* 32:128–144
- Knaepkens G, Bervoets L, Verheyen E, Eens M (2004) Relationship between population size and genetic diversity in endangered populations of the European bullhead (*Cottus gobio*): implications for conservation. *Biol Conserv* 115:403–410
- Li ZC, Xia L, Li YM, Yang QS, Liang MY (2006) Mitochondrial DNA variation and population structure of the yarkand hare *Lepus yarkandensis*. *Acta Theriol* 51(3):243–253
- Lu X (1995) Studies on breeding biology of the Cape hare (*Lepus capensis*). *Acta Theriol Sin* 15:122–127
- Luo ZX (1988) The Chinese hare. China forestry publishing house, Beijing
- Lynch M, Lande R (1998) The critically effective size for a genetically secure population. *Anim Conserv* 1:70–72
- Madsen T, Stille B, Shine R (1996) Inbreeding depression in an isolated population of adders *Vipera berus*. *Biol Conserv* 75: 113–118
- Miller HC, Miller KA, Daugherty CH (2008) Reduced MHC variation in a threatened tuatara species. *Anim Conserv* 11(3):206–214
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

- Pearse DE, Crandall KA (2004) Beyond F_{ST} : analysis of population genetic data for conservation. *Conserv Genet* 5:585–602
- Pierpaoli M, Riga F, Trocchi V, Randi E (1999) Species distinction and evolutionary relationships of the Italian hare (*Lepus corsicanus*) as described by mitochondrial DNA sequencing. *Mol Ecol* 8:1805–1817
- Provan J, Beatty GE, Hunter AM, McDonald RA, McLaughlin E, Preston SJ, Wilson S (2008) Restricted gene flow in fragmented populations of a wind-pollinated tree. *Conserv Genet* 9:1521–1532
- Reed DH, Frankham R (2003) Correlation between population fitness and genetic diversity. *Conserv Biol* 17:230–237
- Reed DH, O'Grady JJ, Brook BW, Ballou JD, Frankham R (2003) Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biol Conserv* 113:23–34
- Robinson T, Canty P, Mooney T, Rudduck P (1996) South Australia's offshore islands. Department of Environment and Natural Resources, Canberra
- Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F -statistics under isolation by distance. *Genetics* 145:1219–1228
- Schmitt LH (1978) Genetic variation in isolated populations of the Australian bush rat *Rattus fuscipes*. *Evolution* 32:1–14
- Seddon JM, Baverstock PR (1999) Variation on islands: major histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Mol Ecol* 8:2071–2079
- Shaffer ML (1981) Minimum population sizes for species conservation. *Bioscience* 31(2):131–134
- Shi YF, Cui JZ, Su Z (2006) The quaternary glaciations and environmental variations in China. Hebei Science and Technology Publishing House, Hebei
- Soule ME (1976) Allozyme variation-its determinants in space and time. In: Ayala FJ (ed) *Molecular evolution*. Sinauer Associates, Sunderland Massachusetts
- Stangel PW, Lennartz MR, Smith MH (1992) Genetic variation and population structure of red-cockaded woodpeckers. *Conserv Bio* 6:283–292
- Sun JM, Zhang ZQ, Zhang LY (2009) New evidence on the age of the Taklimakan Desert. *Geology* 37:159–162
- Tajima F (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Taylor AC, Sherwin WB, Wayne RK (1994) Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat *Larsiohinus kreftii*. *Mol Ecol* 3:277–290
- Traill LW, Bradshaw JA, Brook BW (2007) Minimum viable population size: a meta-analysis of 30 years of published estimates. *Biol Conserv* 139:159–166
- Wang S (1998) China red data book of endangered animals. Science Press, Beijing
- Wright S (1943) Isolation by distance. *Genetics* 28:114–138
- Wu CH, Wu JP, Bunch TD, Li QW, Wang YX, Zhang YP (2005) Molecular phylogenetics and biogeography of *Lepus* in Eastern Asia based on mitochondrial DNA sequences. *Mol Phylogen Evol* 37:45–61
- Yang XP (2001) The relationship between oases evolution and natural as well as human factors—evidences from the lower reaches of the Keriya river, southern Xinjiang, China. *Earth Sci Front* 8:83–89
- Yang XP, Zhu ZD, Jaekel D, Owen LA, Han JM (2002) Late Quaternary palaeoenvironment change and landscape evolution along the Keriya River, Xinjiang, China: the relationship between high mountain glaciation and landscape evolution in foreland desert regions. *Quaternary Int* 97(98):155–166
- Zhand HY, Men GF (2002) Stratigraphic subdivision and climatic change of the quaternary of the center Taklimakan Desert. *Xinjiang Geology* 20:256–261
- Zhong W, Xiong H (1999) Paleo-climatic and environmental development since about 4 Ka BP and the relation with abandonments of ancient cities in Southern Xinjiang. *J Desert Res* 19:343–347
- Zhu ZD, Chen ZP, Wu Z, Li JZ, Li BY, Wu GC (1991) Study on the geomorphology of wind-drift sands in the Taklamakan Desert. Science Press, Beijing
- Zu RP, Gao QZ, Qu JJ, Qiang MR (2003) Environmental changes of oases at southern margin of Tarim Basin, China. *Environ Geol* 44:639–644