Journal of Animal & Plant Sciences, 32(6): 2022, Page: 1542-1551
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2022.6.0563

GENETIC DIVERSITY AND POPULATION STRUCTURE OF THE ASIAN HOUSE RAT *RATTUS TANEZUMI* IN SHANXI, CHINA

X. Yang [1, 2], T. Wang [2], H. Guo [1], J. Yang [2], M. Zhang [1], B. Zou [2], Z. Ren [1], J. Zhang [1*], and J. Zhang [3, 4*]

[1]Institute of Applied Biology, School of Life Science, Shanxi University, Taiyuan, 030006, China
[2]Shanxi Key Laboratory of Integrated Pest Management in Agriculture, College of Plant Protection, Shanxi Agricultural University, Taiyuan, 030031, China.
[3]The State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China
[4]CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing, 100049, China

Corresponding Authors' Email: zjz@sxu.edu.cn; zhangjx@ioz.ac.cn

ABSTRACT

Asian house rat *Rattus tanezumi*, a common commensal rat in southern China, was first discovered in Shanxi Province, northern China, in the early 1990s. Its rapid expansion threatens to reduce the diversity of native species and poses a serious threat to local biodiversity. In this study, samples were collected from fourteen different locations using one-night trapping method, and eight populations of *Rattus tanezumi* were captured. Seventy-six samples were successfully genotyped at 12 microsatellite loci and 70 mitochondrial DNA (mtDNA) cytochrome c oxidase subunit I gene sequences were amplified and sequenced. The analysis of genetic diversity and population structure about *R. tanezumi* significantly reveal evolutionary forces and non-random mating within populations in Shanxi, which could be concluded as a bottleneck or selective sweep. The genetic and geographical distances were uncorrelated, suggesting that geographic distance had no (or a weak) effect on genetic distance. Furthermore, we found four new haplotypes that are most similar to a nonrandom subset of those in their native range in southern China. The haplotypes could be more common in Shanxi due to a genetic bottleneck or natural selection. Further evidence is needed to reveal the mechanisms of genetic exchange and molecular evolution underlying these patterns. Our results provide fundamental insights into the potential introduction routes and the relevant features of successful invasions.

Keywords: Diversity, Population structure, Asian house rat, Gnomic analysis, Microsatellite loci, COI gene sequence.

Published first online June 10, 2022. Published final November 20, 2022

INTRODUCTION

The Asian house rat *Rattus tanezumi* (Rodentia: Muridae), one of the major commensal rat species in the southern provinces of the Yangtze River basin in China, is regarded as a pest related to agriculture and public health concerns (Feng *et al.*, 2007; Guo *et al.*, 2017; Shi *et al.*, 2018). The pest not only causes crop damage and household damage, but also spread plague (Yersinia pestis), haemorrhagic fever with renal syndrome (Hantaan virus), and leptospirosis (Leptospira) and other zoonoses (Blasdell *et al.*, 2011; Huang *et al.*, 2013; Plyusnina *et al.*, 2009). In recent years, this species has gradually migrated north of the Yangtze River and has become the dominant species in Shanxi, Shaanxi, Hebei, Henan, Qinghai, and other provinces that belong to the Yellow River Basin (Yang *et al.*, 2019; Zheng *et al.*, 2008; Zhang *et al.*, 2000; Hou and Jiang, 2008; Ma *et al.*, 2011; Wu *et al.*, 2017). The main reasons for expanding this species may be related to human activities and transportation lines such as railways and highways and the developed logistics network (Wilmshurst *et al.*, 2008; Tollenaere *et al.*, 2010; Guo *et al.*, 2019). *R. tanezumi* was first discovered in Southern Shanxi Province, North China, in the early 1990s, and by 2016 the rats had spread to Wutai County in the north of Shanxi (Yang *et al.*, 2019). Further migration of *R. tanezumi* to the north may be inevitable with the rapid development of transportation services and logistics. However, the expansion of a species usually reduces the diversity of native species and may pose a severe threat to local biodiversity (Russell *et al.*, 2014; Stuart *et al.*, 2016; Santicchia *et al.*, 2018). Therefore, it is of practical significance to carry out genetic diversity and population structure monitoring of *R. tanezumi* using molecular genetic markers.

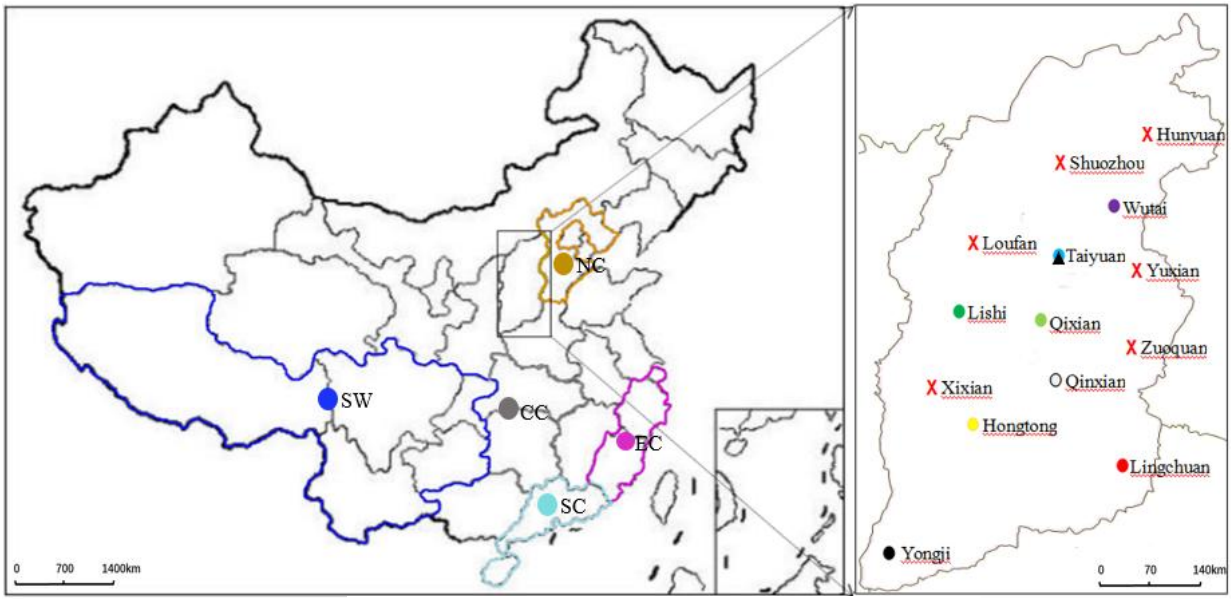
Invasive rodents impact biodiversity, human health, and food security worldwide (Godwin *et al.*, 2019). Thus, the genetic structure of rodent populations after their spread to new regions has aroused increasing interest in recent decades. In Senegal, the house mouse *Mus musculus domesticus*, even between close sites, showed a complex interplay of different demographic processes during its spatial expansion, including

sequential founder effects and stratified dispersal due to human transport along major roads (Lippens *et al.*, 2017). The genetic diversity of the invasive crested porcupine (*Hystrix cristata*) populations in Italy was lower relative to other native populations from sub-Saharan and southern Africa (Trucchi *et al.*, 2016). The study on the molecular characterization of *R. tanezumi* in southern Africa revealed that factors other than climate might influence this species' distribution (Bastos *et al.*, 2011). However, it should be realized that knowledge gaps exist in some regions, and more studies are needed on invasive rodents.

In this study, we aimed to measure the genetic diversity and population structure of *R. tanezumi* after its invasion into Shanxi using microsatellite loci and mitochondrial DNA (mtDNA) cytochrome c oxidase subunit I (COI) gene sequences. Our results provide a scientific basis for fundamentally understanding the features of successful invasive species.

MATERIALS AND METHODS

Sample collection and population distribution: To understand the distribution and genetic characteristics of *R. tanezumi* in Shanxi Province, we selected fourteen different locations with a linear distance more than 50 km from each other, where no specific permits were required. samples were collected from chicken or pig farms in fourteen different locations. The locations were Yongji (YOJ), Lingchuan (LIC), Hongtong (HOT), Xixian (XIX) and Qinxian (QNX) in the south; Zuoquan (ZOQ), Qixian (QIX), Lishi (LIS), Taiyuan (TAY) and Loufan (LOF) in the central area; and Yuxian (YUX), Wutai (WUT), Shuozhou (SOZ) and Hunyuan (HNY) in the north (Fig. 1, Table 1). A one-night trapping method was used to catch individuals in chicken or pig farms with peanuts, walnuts and apples as bait. The traps were open for at least two days and checked daily and re-baited. The number of traps were depended on the size of the location. Nevertheless, the total number of night traps at each site was more than 150. The tail tips of samples were cut and stored at −80°C in absolute ethanol until DNA extraction. All procedures above are following Chinese laws and ethical regulation "Regulations for the Administration of Laboratory Animals established by the Ministry of Science and Technology of the People's Republic of China (2017 Revision)".



A b

Fig. 1. (a) Regions where the samples of COI sequences in GenBank were collected. SW: Yunnan, Chongqing, Sichuan, Guizhou and Tibet of Southwest China; SC: Guangdong and Hainan of South China; CC: Hunan and Hubei of Central China; EC: Zhejiang, Fujian and Shanghai of East China; NC: Hebei of North China. (b) The sampling locations of Shanxi Province. The capital of Shanxi (Taiyuan) is marked with a black triangle. Locations where no R. tanezumi specimens were caught are marked with a red X.

DNA extraction: According to the animal tissue experimental, the genomic DNA of *R. tanezumi* samples was extracted from the tail tips using a Blood/Cell/Tissue Genomic DNA Extraction Kit (DP304, produced by Tiangen Biotech Co. Ltd., Beijing, China) protocol provided by the manufacturer. The DNA was eluted in TE buffer (the mixture of 100 mmol / L Tris-HCl (pH

8.0) and 10 mmol / L EDTA (pH 8.0)) and stored at −20°C.

Microsatellite amplification: *R. tanezumi* samples were only captured in eight of fourteen locations in Shanxi Province. The eight locations were YOJ, LIC, HOT, QNX, QIX, LIS, TAY and WUT (Fig. 1b). Twelve microsatellite loci with high polymorphism were obtained from literature (Guo *et al.*, 2012). DNA from all captured *R. tanezumi* individuals (N = 76) distributed the eight locations was successfully amplified in a thermocycler (BIOER TECHINOLOGY, G1000) in 25 µL reaction mixtures containing 1 µL of total DNA (30-50 ng), 12.5 µL of 2×Taq PCR MasterMix (KT201, produced by Tiangen Biotech Co. Ltd., Beijing, China), and 1 µL of each primer. Cycling conditions were as follows: 180 s at 94°C followed by 30 cycles of 30 s at 94°C, 30 s annealing at 56°C or 58°C (depending on the different primers from Guo *et al.* (2012)) and 30 s at 72°C followed by extension for 300 s at 72°C. The length of the PCR products was determined manually using GeneMapper v. 4.0 software (Applied Biosystems).

Table 1. Populations and sizes of *R. tanezumi* samples used in microsatellite analysis in the present study.

Province	Code	Locations	Number of traps	No. of samples	Geographic coordinates	Elevation (m)	Distance from Taiyuan (km)	Site
Shanxi	YOJ	Yongji	200	5	34.86° N; 110.43° E	350	392	Chicken farms
	LIC	Lingchuan	180	11	35.74° N; 113.27° E	1080	248	Chicken farms
	HOT	Hongtong	200	8	36.25° N; 111.59° E	450	203	Pig farms
	XIX	Xixian	240	0	36.50° N; 110.85° E	1020	192	Chicken farms
	QNX	Qinxian	210	10	36.74° N; 112.69° E	1140	128	Chicken farms
	ZOQ	Zuoquan	160	0	37.05° N; 113.31° E	1080	115	Pig farms
	QIX	Qixian	210	7	37.35° N; 112.31° E	770	86	Chicken farms
	LIS	Lishi	330	7	37.52° N; 111.98° E	1230	135	Chicken farms
	TAY	Taiyuan	750	21	37.80° N; 113.27° E	780	0	Pig farms
	LOF	Loufan	200	0	38.05° N; 111.86° E	1320	69	Pig farms
	YUX	Yuxian	160	0	37.71° N; 113.35° E	1250	78	Pig farms
	WUT	Wutai	460	7	38.64° N; 113.20° E	1060	108	Pig farms
	SOZ	Shuozhou	270	0	39.11° N; 112.30° E	1140	162	Pig farms
	HNY	Hunyuan	270	0	39.85° N; 113.96° E	1550	227	Pig farms

COI gene amplification and sequencing: We amplified the COI gene fragment of *R. tanezumi* in Shanxi with universal primers BatL5310 (5'-CCTACTCRGCCATTTTACCTATG-3') and R6036R (5'-ACTTCTGGGTGTCCAAA GAATCA-3') to explore its sequence polymorphism and population genetics. The PCR reactions were performed in a thermocycler (BIOER TECHINOLOGY, G1000) in 25 µL reaction mixture containing 1.5 µL of total DNA (30-50 ng), 12.5 µL of 2×Taq PCR MasterMix (KT201, produced by Tiangen Biotech Co. Ltd., Beijing, China), and 1 µL of each primer. Cycling conditions were as follows: 300 s at 95°C followed by 35 cycles of 45 s at 95°C, 60 s at the annealing temperature of 54°C, 60 s at 72°C, and then 600 s at 72°C. For sequencing, the amplified products were identified by 1% agarose gel electrophoresis and then sent to Shanghai Personal Biotechnology Co. Ltd., Shanghai, China.

Microsatellite DNA analysis: GENEPOP version 4.0 (https://genepop.curtin.edu.au/) was used to test the deviations from Hardy-Weinberg equilibrium (HWE) and to test for linkage disequilibrium for each population (Rousset, 2008). The genetic diversity per locus of each population was characterized by estimating the average number of alleles (N_a), allelic richness (R_s), observed heterozygosity (H_o) and expected heterozygosity (H_e) using Microsatellite Analyzer (MSA) version 4.05 (Dieringer and Schlötterer, 2003). The inbreeding coefficient (F_{IS}) and pairwise fixation indices (F_{ST}) were estimated with ARLEQUIN version 3.5.1.2 (Excoffier and Lischer, 2010). Pairwise fixation indices (F_{ST}) and *P*-values for statistical significance were also calculated to estimate the significance of population differentiation using ARLEQUIN version 3.5.1.2 (Excoffier and Lischer, 2010).

To examine the population genetic structure of *R. tanezumi*, STRUCTURE version 2.3.4 was used to estimate the number of clusters (*K*) based on information from individuals (Evanno *et al.* 2005). Twenty independent repeated runs for each *K*-value from one to ten were carried out to estimate the most likely number of clusters. Each replicate run consisted of a burn-in period of 50,000 steps followed by 1,000,000-step Markov

Chain Monte Carlo (MCMC) runs for estimating the true number of clusters in the samples (*K*-value) (Evanno *et al.*, 2005; Yang *et al.*, 2016). The results were compressed and uploaded to the Structure Harvester website (http://taylor0.biology.ucla.edu/ structure Harvester/) to estimate the *K*-value, and then repeatedly sampled with CLUMPP 1.1.1 to optimize the coefficients (Jakobsson and Rosenberg, 2007). A structure chart was drawn with DISTRUCT Version 1.1 (Rosenberg, 2004).

The pairwise Nei's genetic distances were computed using Popgen 32 (Krawczak *et al.*,2006). The linear geographical distances between each location is obtained from Google Maps (http://www.gugeditu.net).

Sequence analysis: The program Muscle aligned COI gene sequences within MEGA X (Kumar *et al.* 2018; Edgar 2004). The number of segregating sites (*S*), the number of haplotypes (*H*), haplotype diversity (H_pD), nucleotide diversity (π), and the average number of nucleotide differences (*K*) were calculated with DnaSP version 5.10 (Librado and Rozas, 2009).

The median-joining networks of COI gene haplotypes were constructed by NETWORK 10.1.0.0 (Bandelt *et al.*, 1999). Given the short expansion history of *R. tanezumi* in Shanxi Province (Yang *et al.*, 2019), we obtained a total of 499 known COI sequences from GenBank (Accession No. MH303212-MH303710) to understand the genetic relationship of *R. tanezumi* populations between Shanxi and other regions of China. The obtained COI sequences were geographically grouped into SW (from Yunnan, Chongqing, Sichuan, Guizhou and Tibet of southwest China), SC (from Guangdong and Hainan of southern China), CC (from Hunan and Hubei of central China), EC (from Zhejiang, Fujian and Shanghai of eastern China) and NC (from Hebei in northern China, a neighboring province to Shanxi that also is in the species' expansive range; Zhang *et al.*, 2000) (Fig. 1a).

Statistical analysis: HWE *P*-values based on microsatellite loci was obtained using GENEPOP version 4.0 (https://genepop.curtin.edu.au/). The paired sample *t*-test in SPSS 20.0 was used to test the significant difference between the mean H_o and the mean H_e of *R. tanezumi* populations. The single sample *t*-test in SPSS 20.0 was used to test whether the average of F_{IS} was significant more than zero. The F_{ST} *P*-values for statistical significance were calculated to estimate the significance of population differentiation using ARLEQUIN version 3.5.1.2 (Excoffier and Lischer, 2010). We analyzed the correlation of pairwise genetic distance and the natural logarithm of linear geographical distance using SPSS 20.0. Tajima's *D* test was used in DnaSP version 5.10 to determine whether the COI gene sequences deviated from neutral evolution.

RESULTS

Genetic diversity based on Microsatellite analysis: *R. tanezumi* samples were collected in 14 locations and were captured only in 8 locations, indicating that *R. tanezumi* might not be distributed in the other 6 locations. Seventy-six individuals belonging to the eight populations (N = 5-21 rats per population, the whole trap-success rate was 1.98%) were successfully genotyped at 12 microsatellite loci in this study (Table 1). All loci were polymorphic but were not in linkage disequilibrium. The Hardy-Weinberg exact tests indicated that these populations were in HWE except for TAY (HW $P < 0.05$, Table 2). The average N_a ranged from 3.67 to 7.08 (Table 2). The maximum R_s was 4.03 (QIX), and the minimum R_s was 3.07 (YOJ) (Table 2). The H_o in each population ranged from 0.508 to 0.679. The minimum H_o was in WUT and the maximum H_o in QIX (Table 2). The H_e ranged from 0.610 (LIS) to 0.788 (TAY) (Table 2). The F_{IS} over 12 loci ranged from 0.056 to 0.356 with an average of 0.232 (Table 2).

Table 2. Genetic variation in eight populations averaged over 12 microsatellite loci.

Index	YOJ	LIC	HOT	QNX	QIX	LIS	TAY	WUT
N_a	3.67	5.00	4.75	5.42	5.67	3.75	7.08	4.50
R_s	3.07	3.52	3.62	3.47	4.03	3.22	3.91	3.73
H_o	0.596	0.606	0.617	0.550	0.679	0.563	0.622	0.508
H_e	0.662	0.739	0.739	0.721	0.756	0.610	0.788	0.757
F_{IS}	0.056	0.244	0.175	0.356	0.186	0.275	0.214	0.350
HW-*P*	0.306	0.142	0.210	0.055	0.326	0.362	**0.036**	0.234

N_a, Number of alleles; R_s, Allelic richness; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, the inbreeding coefficient; HW-*P*, *P*-value for Hardy-Weinberg equilibrium (the HW-*P* given in bold denotes $P<0.05$). The above data are mean values over twelve loci.

Population structure: Bayesian analysis of population genetic structure based on the results from STRUCTURE revealed that the most likely *K*-value was three (posterior probability of Bayesian clustering Ln (*D*) likelihood score optimal for $K = 3$, Fig. 2a; Fig. 2b). Cluster 1 consisted of almost all individuals of LIC and HOT (Fig. 2c). The YOJ and QNX populations were assigned to Cluster 2 (Fig. 2c). The LIS population was mainly composed of Cluster 1 and Cluster 3 (Fig. 2c). The QIX,

TAY and WUT were mixtures of the three clusters (Fig. 2c).

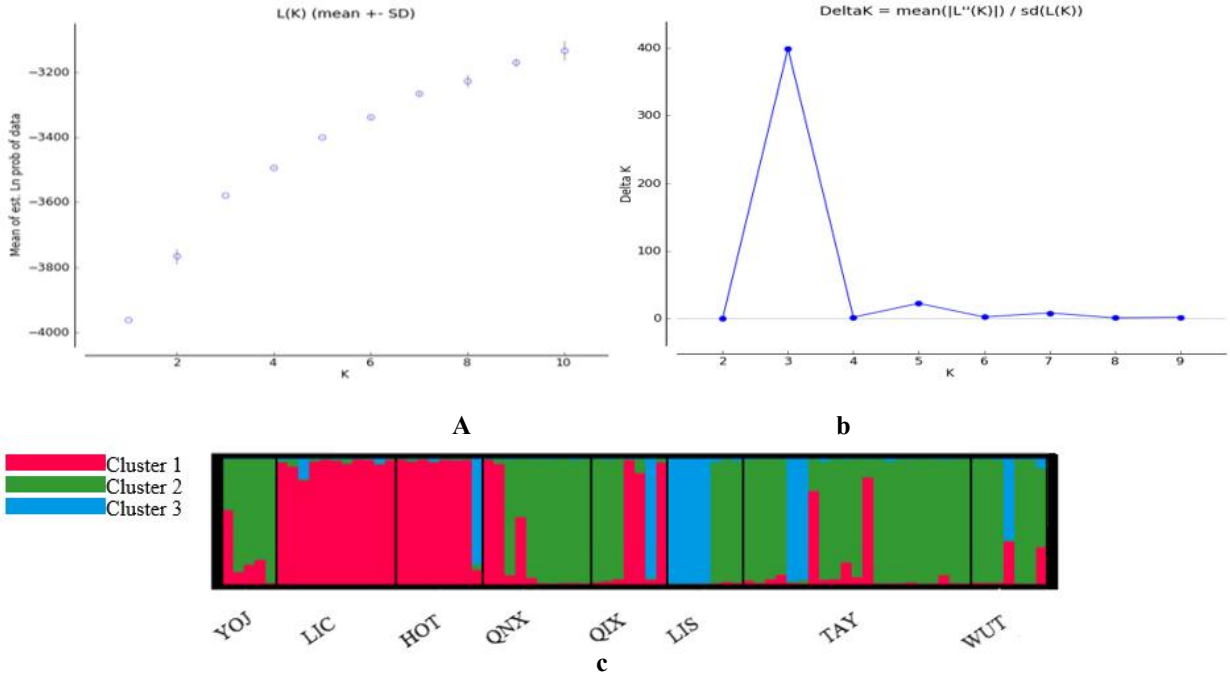


Fig. 2. Bayesian analyses of population genetic structure revealed that the most likely K-value was three. (a) Mean K values (mean L(K)(±SD) over 20 runs for each value of K between 1 and 10). (b) ΔK values (ΔK over 20 runs for each value of K between 1 and 10). (c) Clustering of the analyzed populations (each individual is represented by a line partitioned into K colored segments according to the individuals' estimated membership fractions in each of the K clusters).

Pairwise F_{ST} values of eight populations ranged from 0.047 to 0.294. The overall F_{ST} across all loci was 0.130 (Table 3). The results showed that 13.0% of the genetic variation existed among populations, while most (87.0%) of the variation was within populations. The *P*-values of pairwise F_{ST} for 28 population pairs (significance level, $P = 0.05$) showed that only the differences between population pairs of YOJ-QIX, YOJ-WUT and TAY-WUT were not significant (Table 3). The other 25 pairwise comparisons were significant or extremely significant (global F_{ST} *P*-value < 0.001) (Table 3). Furthermore, seventeen pairwise comparisons showed moderate genetic differentiation (F_{ST} = 0.050-0.150) and 10 showed great genetic differentiation (Table 3; Wright, 1931).

Table 3. Pairwise F_{ST} values of eight populations and the significance of population differentiation.

	YOJ	LIC	HOT	QNX	QIX	LIS	TAY
LIC	0.137**						
HOT	0.162**	0.136**					
QNX	0.118**	0.102**	0.153**				
QIX	0.062	0.110**	0.128*	0.113**			
LIS	0.272**	0.252**	0.294**	0.201**	0.261**		
TAY	0.087**	0.111**	0.133**	0.054**	0.090**	0.183**	
WUT	0.061	0.133**	0.171**	0.090**	0.064*	0.179**	0.047

*, $P < 0.05$; **, $P < 0.01$.

Genetic distance and correlation with linear geographical distance: For the pairwise genetic distance, the minimum value was 0.2852 for TAY-QNX, and the maximum value was 1.3505 for LIS-HOT (Table 4). In addition, the values of TAY-LIC, TAY-WUT, LIC-QNX and QNX-QIX were less than 0.5, indicating the close relationship of LIC, QNX, QIX, TAY and WUT (Table 4).

We analyzed the correlation of pairwise genetic distance and the natural logarithm of linear geographical distance (Table 4). The genetic distance and geographical distance were uncorrelated (R^2 = 0.038, P = 0.317),

suggesting that geographic distance had no (or a weak) effect on genetic distance.

Table 4. Geographical distance (km) (above the diagonal) and pairwise Nei's distances based on 12 microsatellite loci (below the diagonal) of eight populations of *R. tanezumi.*

Code	YOJ	LIC	HOT	QNX	QIX	LIS	TAY	WUT
YOJ	—	270	193	293	331	305	392	500
LIC	0.8431	—	146	122	198	266	248	331
HOT	0.8031	0.5585	—	106	136	149	203	306
QNX	0.8113	0.4375	0.8197	—	77	158	128	225
QIX	0.5624	0.5425	0.7349	0.4793	—	99	86	171
LIS	1.2816	0.9379	1.3505	0.7087	1.2212	—	135	225
TAY	0.6584	0.4616	0.7866	0.2852	0.5106	0.5815	—	108
WUT	0.5721	0.8111	1.1186	0.5168	0.5889	0.5722	0.3664	—

Genetic diversity based on COI genes: The COI genes of 70 *R. tanezumi* individuals from eight populations in Shanxi Province were successfully amplified and sequenced. There were ten segregating sites for all populations in Shanxi, and the number of haplotypes was six per population ranging from one to five (Table 5). The nucleotide diversity (π = 0.00108, H_pD = 0.513, K = 0.733) was lower than in a previous study (π = 0.00287, H_pD = 0.747, K = 2.016; Guo *et al*. 2019). In Shanxi, the YOJ population showed the highest nucleotide diversity (π = 0.00387, H_pD = 0.700, K = 2.700), while the nucleotide diversity of QNX and WUT were the lowest (π = 0, H_pD = 0, K = 0) (Table 5). Tajima's D values were negative and statistically significant for YOJ (D = −1.369, $P < 0.01$) and for Shanxi as a whole (D = −1.869, $P < 0.05$), indicating evolutionary selection of *R. tanezumi* in Shanxi (Table 5).

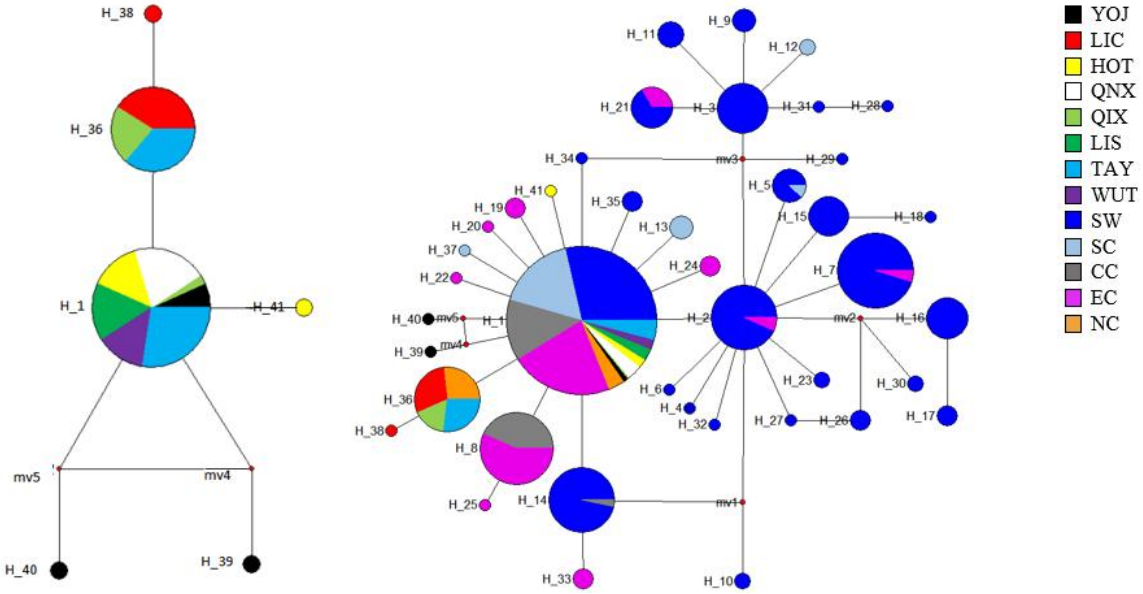
Phylogenetic analysis: For individuals of *R. tanezumi* populations in Shanxi, the median-joining network based on COI gene sequences showed that six haplotypes formed two distinct haplogroups (Fig. 3a). The haplogroup centered on H36 contained 35.7% of 70 individuals from Shanxi and was composed of the individuals from LIC, QIX and TAY populations (Fig. 3a). The most frequent haplotype H1 (62.9%) was observed in all populations in Shanxi except LIC, and H39, H40, and H41 were around H1 (Fig. 3a).

The median-joining network based on 70 individuals obtained in this study and 499 individuals downloaded from GenBank showed forty-one haplotypes and three distinct haplogroups (Fig. 3b). The haplotype H1 contained individuals of all groups and populations from Shanxi (Fig. 3b). The haplotype H36, as a branch of H1, was composed of the individuals from LIC, QIX, TAY and NC (Hebei province) (Fig. 3b). An individual from LIC formed H38 (Fig. 3b). Similarly, H39, H40, and H41 also contained one individual. The individuals of H39 and H40 were from YOJ, and H41 was from HOT (Fig. 3b).

Table 5. COI gene genetic diversity and Tajima's test for *R. tanezumi* populations.

Populations		*N*	*S*	H_p	H_pD	π	*K*	Tajima's *D*
	YOJ	5	6	3	0.700	0.00387	2.700	−1.369**
	LIC	10	5	5	0.667	0.00182	1.267	−1.136
	HOT	7	5	4	0.714	0.00301	2.095	0.132
	QNX	9	0	1	0.000	0.00000	0.000	—
Shanxi	QIX	6	1	2	0.333	0.00048	0.333	−0.933
	LIS	7	1	2	0.476	0.00070	0.476	0.559
	TAY	20	1	2	0.505	0.00073	0.505	1.430
	WUT	6	0	1	0.000	0.00000	0.000	—
	Overall	70	10	6	0.513	0.00108	0.733	−1.869*

N, simple size; *S*, the number of segregating sites; H_p, the number of haplotypes; H_pD, haplotype diversity; π, nucleotide diversity; *K*, the average number of nucleotide differences; *, $P < 0.05$; **, $P < 0.01$.



A b

Fig. 3. Median-joining Networks based on the COI gene. Each circle represents a haplotype. Cycle size is roughly proportional to the haplotype frequency. Colors of a circle refers to the individual sampling locations. (a) Median-joining Network of six haplotypes detected in 70 individuals from eight populations in Shanxi province. (b) Median-joining Network of 41 haplotypes detected in 569 individuals, including individuals from Shanxi and GenBank.

DISCUSSION

In this study, we only collected *R. tanezumi* samples in 8 of 14 locations. The altitude of the six locations where this species was not captured was more than 1000 m (Table 1). Thus, *R. tanezumi* may only be distributed in low altitude areas or in places with convenient transportation to the south of Wutai in Shanxi. The distribution of *R. tanezumi* in Shanxi could be related to its adaptation and expansive mode in light of its short expansion history from south to north China.

A significant departure from HWE was observed in TAY of the eight populations, suggesting evolutionary forces. Tajima's *D* values also indicated evolutionary selection of *R. tanezumi* for Shanxi as a whole. In addition, the mean H_o (0.593) of *R. tanezumi* populations in Shanxi was significantly lower than the mean H_e (0.722) (paired sample *t*-test using SPSS 20.0, $P = 0.001 < 0.01$), and the average of F_{IS} (0.232) was significantly more than zero (single sample *t*-test using SPSS 20.0, $P = 0.000 < 0.001$), suggesting non-random mating within populations. Although the exact pressures driving these facts remain to be determined, we believe an important factor is the low density of this species in Shanxi (the whole trap-success rate was 1.98%).

Compared with a previous study of other provinces of mainland China (Guo *et al.*, 2019), our data for both microsatellites and mtDNA showed lower diversity (average N_a per locus ranged from 3.67 to 7.08; π ranged from 0 to 0.00387). We believe that this can be attributed to genetic drift and founder effects (Mayr, 1942), observed in many other invasive species. Sika deer *Cervus nippon yakushimae* native to East Asia lost neutral diversity after being introduced to Delmarva Peninsula, USA about 100 years ago (Kalb *et al.*, 2019). In Argentina, low levels of genetic diversity of *Lagostomus maximus* was considered to be caused by a relatively recent expansion event (Gariboldi *et al.*, 2019). After *Myocastor coypus* was introduced released into the wild of South Korea in 1987, its genetic diversity showed a low level (Kim *et al.*, 2019). In this study, low genetic diversity might mean fewer invasion source and inbreeding within *R. tanezumi* populations.

Bayesian analysis of the microsatellite loci revealed that the populations were all a mixture of the genetic clusters. However, the gene cluster compositions of the populations were different and did not correspond strictly to geography (Fig. 2). Considering the deviations from neutral sequence evolution within *R. tanezumi* as a whole (Tajima's *D* test of the COI sequences, Table 5), it could be explained by a bottleneck or selective sweep in Shanxi. The extensive long-distance gene exchange may be via human-mediated dispersal. Similarly, the genetic structure of *Mus musculus domesticus* in Senegal revealed strong bottlenecks and a complex interplay occurring during its spatial expansion, including stratified dispersal due to human transport along major roads (Lippens *et al.*, 2017). *M. musculus* colonizing Gough

Island 100 years ago also showed substantial reductions in mitochondrial and nuclear sequence variation and weak reductions in microsatellite diversity compared with Western European populations, consistent with a population bottleneck (Gray *et al.*, 2014). The genetic data of *R. rattus*, a major invader in Senegal, West Africa, revealed that its invasion pathways closely paralleled the history of human trade routes, and genetic bottlenecks and admixture have played a major role in shaping the genetics of it (Konečný *et al*., 2013).

The expansion of a species, especially rodents, to a non-residential area can often be attributed to frequent human activity. For example, the brown rat *R. norvegicus*, which originated in northern China and Mongolia, expanded to Southeast Asia along with the southward human migration across China between 800-1550 AD, and then rapidly expanded worldwide via existing maritime trade routes (Puckett and Munshi-South, 2019). The expansion of the house mouse *Mus musculus* out of India and Southeast Asia took place for thousands of years and was perhaps associated with the spread of agricultural practices (Suzuki *et al.*, 2013). The three most invasive rat species, the black or ship rat *R. rattus*, the brown or Norway rat *R. norvegicus*, and the Pacific rat *R. exulans*, were incrementally introduced to islands as humans explored the world's oceans (Harper and Bunbury, 2015). The invasion of *R. tanezumi* in Shanxi was also probably due to modern transportation services that provided opportunities for long-distance dispersal. As a result, the gene flow of *R. tanezumi* in Shanxi over long distances may mainly occur via human-mediated dispersal mechanisms such as transportation and logistical services.

For phylogenetic analysis of COI gene sequences, the individuals from Shanxi displayed a star-like pattern centered around the haplotype H1 that contained individuals of all groups and populations from Shanxi in the median-joining network (Fig. 4B). The haplotype H36 contained individuals from LIC, QIX, TAY and NC (Hebei Province) (Fig. 4B). In view of the short history of *R. tanezumi* in Shanxi and Hebei, some individuals may have spread from the same region. Interestingly, four new haplotypes were found in this study (H38, H39, H40 and H41). The COI haplotypes of the species in Shanxi are most closely related to a nonrandom subset of those in their native range in southern China. These haplotypes could be more common in Shanxi due to a genetic bottleneck or natural selection. Further evidence is needed to reveal their means of genetic exchange and molecular evolution.

In this study, the sample size collected is small due to the low density of *R. tanezumi* in Shanxi, limiting our data analysis in depth. In the future, we may need a larger sample size and more data analysis to reveal this rat species' expansive pathway and molecular evolutionary history.

Conclusion: Our results, which showed low genetic diversity of *R. tanezumi* populations and the cluster with a nonrandom subset in Shanxi based on microsatellite and COI gene analysis, significantly revealed the evolutionary forces and non-random mating within populations. It may mean fewer invasion source and inbreeding of *R. tanezumi*, consistent with a population bottleneck or selective sweep. The extensive long-distance gene exchange may be via human-mediated dispersal. Further evidence is needed to reveal the genetic exchange and molecular evolution of *R. tanezumi*. This study provide fundamental insights into the potential introduction routes and the relevant features of successful invasions.

Conflict of Interest: The authors declare no potential conflict of interest.

Acknowledgements: This research was supported by the Open Project of the State Key Laboratory of Integrated Management of Pest Insects and Rodents (ChineseIPM1618), the Special Investigation on Basic Resources of Science and Technology, China (2019FY100300), the Key R&D Project (Agriculture) in Shanxi Province, China (201803D221015-1) and the Beijing Natural Science Foundation (5192015). We thank LetPub (www.letpub.com) for its linguistic assistance and scientific consultation during the preparation of this manuscript.

REFERENCES

Bandelt, H. J., P. Forster and A. Röhl (1999). Median-joining networks for inferring Intraspecific phylogenies. Mol. Bio. Evol. 16(1): 37-48.

Bastos, A.D., D. Nair, P.J. Taylor, H. Brettschneider, F. Kirsten, E. Mostert, E. Maltitz, J.M. Lamb, P. Hooft, S.R. Belmai, G. Contrafatto, S. Downs and C.T. Chimimba (2011). Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive *Rattus* congeners in South Africa. BMC Genet. 12: 26.

Blasdell, K., J.F. Cosson, Y. Chaval, V. Herbreteau, B. Douangboupha, S. Jittapalapong, A. Lundqvist, J.P.S. Morand and P. Buchy (2011). Rodent-borne hantaviruses in Cambodia, Lao PDR, and Thailand. Ecohealth. 8(4): 432-443.

Dieringer, D. and C. Schlötterer (2003). Microsatellite Analyser (MSA): A platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes. 3 (1): 167-169.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic. Acids. Res. 32 (5): 1792-1797.

Evanno, G., S. Regnaut and J. Goudet (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Eco. 14 (8): 2611-2620.

Excoffier, L. and H.E.L. Lischer (2010). Arlequin Suite Ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10 (3): 564-567.

Feng, L., Y.L. Zhao, M.W. Zhang, Y. Wang, Z.G. Xu, J. Peng and S.Y. Wang (2007). Distribution pattern and diversity of rodent communities at beach and lakeside areas in the Dongting Lake region. Acta Ecol. Sin. 37 (17): 5771-5779.

Gariboldi, M.C., P.I.F. Inserra, S. Lucero, M. Failla, S.I. Perez and A.D. Vitullo (2019). Unexpected low genetic variation in the South American hystricognath rodent *Lagostomus maximus* (Rodentia: Chinchillidae). PLoS One. 14(9): e0221559.

Godwin, J., M. Serr, S.K. Barnhill-Dilling, D.V. Blondel, P.R. Brown, K. Campbell, J. Delborne, A.L. Lloyd, K.P. Oh, T.A.A. Prowse, R. Saah and P. Thomas (2019). Rodent gene drives for conservation: opportunities and data needs. Proc. Biol. Sci. 286 (1914): 20191606.

Gray, M.M., D. Wegmann, R.J. Haasl, M.A. White, S.I. Gabriel, J.B. Searle, R.J. Cuthbert, P.G. Ryan and B.A. Payseur (2014). Demographic history of a recent invasion of house mice on the isolated Island of Gough. Mol Ecol. 23(8): 1923-1939.

Guo, H.L., H.J. Teng, J.H. Zhang, J.X. Zhang and Y.H. Zhang (2017). Asian house rats may facilitate their invasive success through suppressing brown rats in chronic interaction. Front. Zool. 14: 20.

Guo, S., G.C. Li, J.L. Liu, J. Wang, L. Lu and Q.Y. Liu (2019). Dispersal route of the Asian house rat (*Rattus tanezumi*) on mainland China: insights from microsatellite and mitochondrial DNA. BMC Genet. 20: 11.

Guo, S., L. Lu, H.L. Ma and Q.Y. Liu (2012). Isolation of microsatellite markers in *Rattus tanezumi.* Chin. J. Vector Biol. & Control. 23 (3): 198-201.

Harper, G.A. and N. Bunbury (2015). Invasive rats on tropical islands: their population biology and impacts on native species. Glob. Ecol. Conser. 3 (2015): 607-627.

Hou, Y.F. and X.C. Jiang (2008). The investigation of *Rattus tanezumi* in Shijiazhuang during 2004-2007. Chin. J. Vector Biol. & Control. 19(2): 125.

Huang, L.Q, X.G. Guo, J.R. Speakman and W.G. Dong (2013). Analysis of gamasid mites (Acari: Mesostigmata) associated with the Asian house rat, *Rattus tanezumi* (Rodentia: Muridae) in Yunnan Province, Southwest China. Parasitol Res. 112(5): 1967-1972.

Jakobsson, M. and N.A. Rosenberg (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23 (14): 1801-1806.

Kalb, D.M., D.A. Delaney, R.W. DeYoung and J.L. Bowman (2019). Genetic diversity and demographic history of introduced sika deer on the Delmarva Peninsula. Ecol. Evol. 9 (19): 11504-11517.

Kim, D.Y., H.Y. Suk, S.K. Park, S.Y. Kang, S.H. Seok, S.K. Lee and S.C. Yeon (2019). Development of microsatellite markers and the genetic diversity of *Myocastor coypus* introduced to South Korea. J Vet Med Sci. 81(3): 499-503.

Konečný, A., A. Estoup, J.M. Duplantier, J. Bryja, K. Bâ, M. Galan, C. Tatard and J.F. Cosson (2013). Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa. Mol Ecol. 22(2): 286-300.

Krawczak, M., S. Nikolaus, H. Eberstein, P.J.P. Croucher, N.E.E. Mokhtari and S. Schreiber (2006). PopGen: population based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet. 9 (1):55-61.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6): 1547-1549.

Librado, P. and J. Rozas (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25 (11): 1451-1452.

Lippens, C., A. Estoup, M.K. Hima, A. Loiseau, C. Tatard, A. Dalecky, K. Bâ, M. Kane, M. Diallo, A. Sow, Y. Niang, S. Piry, K. Berthier, R. Leblois, J.M. Duplantier and C. Brouat (2017). Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: a legacy of colonial and contemporary times. Heredity (Edinb). 119 (2): 64-75.

Mayr, E. (1942). Systematics and the Origin of Species. Columbia University Press; New York.

Ma, Y., Y.W. Wei, J. Luo, L. Lu and Q.Y. Liu (2011). Spatial niche analysis of the small mammals in Haidong region of Qinghai Province, China. Chinese J. Zoology. 46 (4): 126-130.

Plyusnina, A., I.N. Ibrahim and A. Plyusnin (2009). A newly recognized hantavirus in the Asian house rat (*Rattus tanezumi*) in Indonesia. J Gen Virol. 90(Pt 1): 205-209.

Puckett, E.E. and J. Munshi-South (2019). Brown rat demography reveals pre-commensal structure in eastern Asia before expansion into Southeast Asia. Genome Res. 29 (5): 762-770.

Rosenberg, N.A. (2004). Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes. 4 (1): 137-138.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8 (1): 103-106.

Russell, J.C., N.S. Sataruddin and A.D. Heard (2014). Over-invasion by functionally equivalent invasive species. Ecology 95 (8): 2268-2276.

Santicchia, F., B. Dantzer, F. Kesteren, R. Palme, A. Martinoli, N. Ferrari and L.A. Wauters (2018). Stress in biological invasions: Introduced invasive grey squirrels increase physiological stress in native Eurasian red squirrels. J. Anim. Ecol. 87 (5): 1342-1352.

Shi, L.Y., G.R. Yang, Z.K. Zhang, L.X. Xia, Y. Liang, H.L. Tan, J.R. He, J.G. Xu, Z.Z. Song, W. Li and P. Wang (2018). Reemergence of human plague in Yunnan, China in 2016. PLoS One. 13 (6): e0198067.

Stuart, A.M., C.V. Prescott and G.R. Singleton (2016). Can a native rodent species limit the invasive potential of a non-native rodent species in tropical agroforest habitats? Pest Manag. Sci. 72 (6): 1168-1177.

Suzuki, H., M. Nunome, G. Kinoshita, K.P. Aplin, P. Vogel, A.P. Kryukov, M.L. Jin, S.H. Han, I. Maryanto, K. Tsuchiya, H. Ikeda, T. Shiroishi, H. Yonekawa and K. Moriwaki (2013). Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. Heredity (Edinb). 111 (5): 375-390.

Tollenaere, C., C. Brouat, J.M. Duplantier, L. Rahalison, S. Rahelinirina, M. Pascal, Moné, G. Mouahid, H. Leirs and J.F. Cosson (2010). Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. J. Biogeogr. 37 (3): 398-410.

Trucchi, E., B. Facon, P. Gratton, E. Mori, N.C. Stenseth and S. Jentoft (2016). Long live the alien: is high genetic diversity a pivotal aspect of crested porcupine (*Hystrix cristata*) long-lasting and successful invasion? Mol. Ecol. 25 (15): 3527-3539.

Wilmshurst, J.M., A.J. Anderson, T.F.G. Higham and T.H. Worthy (2008). Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. PNAS. 105 (22): 7676-7680.

Wright, S. (1931). Evolution in Mendelian population. Genetics. 16 (3): 91-159.

Wu, H.X., L. Lu, F.X. Meng, Y.H. Guo and Q.Y. Liu (2017). Reports on national surveillance of rodents in China, 2006-2015. J. Vector Biol. & Control. 28 (6): 517-522.

Yang, J., Z.F. Liu, J.Q. Fan, Y.P. Wu, R.Y. Ma and R.J. Fan (2016). Genetic variation and population structure of the oriental fruit moth *Grapholita molesta* in Shanxi, a major pome fruits growing region in North China. J. Asia-Pac. Entomol. 19: 1131-1137.

Yang, X.G., J.X. Zhang, J.Z. Zhang, T.L. Wang, B. Zou, W.Y. Chang, Y. Hou and J. Liu (2019). Investigation on distribution status and age structure of *Rattus flavipectus* population in Shanxi. J. Shanxi Agric. Sci. 47 (1): 106-108.

Zhang, M.W., A.G. Chen, Y. Wang, C. Guo, H.F. Liu and B. Li (2000). The biological characteristics of the buff-breasted rats (*Rattus flavipectus*) in Yangtze Valley in China. Acta Theriologica Sinica 20 (3): 200-211.

Zheng, Z.M., Z.K. Jiang and A.G. Chen (2008). Conspectus of Glires. Shanhai Jiao Tong University Press; Shanghai (China). pp. 156-157.