Supplementary Materials

Supplementary Materials and Methods

Ethics approval and consent to participate

This study was approved by the Animal Ethics Committee of the Institute of Zoology, Chinese Academy of Sciences (IOZ20150069). Blood samples were collected from wild Chinese grouse, which were released back into the wild. The blood collection procedures were in strict accordance with the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Sample collection and sequencing

We sampled 29 individuals from eight locations (Table 1) (Song et al., 2021). The Chinese Grouse samples (n=16) were obtained from three populations (three from the Qilian Mountains (QLS), three from Zhuoni (ZN), and 10 from Lianhuashan National Nature Reserve (LHS)). The Hazel Grouse samples (n=13) were obtained from five populations (one from northeast Poland (NEP), one from the Austrian Alps, three from Bavaria in Germany (GER), three from Jämtland in Sweden (SWE), and five from northeast China (XLJ). All samples were preserved in 99% ethanol at −20 °C. Genomic DNA was extracted using a DNeasy Blood (Quintana-Murci, 2016) assays (Life Technologies, USA).

All samples were sequenced on the Illumina sequencing platform (NovaSeq 6000) at *Annoroad Gene Technology* (Beijing, China). DNA libraries (150 bp) were constructed according to the manufacturer's instructions. Analyses were based on clean reads, which were filtered following a three-step procedure: i.e., (1) removing adapter polluted reads >5 bp, (2) removing low-quality reads (quality score<19), and (3) removing Ns reads (rate>5%) (Additional File 2: Data filter summary and distribution). In total, 686.04 Gb (97%, out of 705.13 Gb) of high-quality paired-end reads were retained for further analysis (Supplementary Table S3).

Population genetic analysis

The sequences of all individuals were mapped to the Chinese Grouse reference genome and used for all subsequent analyses (Song et al., 2020). Nucleotide diversity (π) and Tajima's D of each location were calculated for all single nucleotide polymorphisms (SNPs) in unrelated individuals per site using VCFtools (v0.1.14). EIGENSOFT (v6.0.1) (Patterson et al., 2006) was run to estimate F_{ST} between sampling locations of the two species.

Linkage disequilibrium (LD) analysis

Genome-wide LD was estimated for the total panel and for each subgroup (as determined by population structure) using pairwise comparisons among the SNP markers (missing rates<0.30 and minor allele frequency (MAF)≥0.05) using r^2. For all pairs of SNPs, r^2 was calculated using PopLDdecay v.3.30 (Zhang et al., 2019).

Population structure

Principal component analysis (PCA) was carried out without calling genotypes following the procedure outlined in (Fumagalli et al., 2013). Genotype likelihoods were estimated in ANGSD v0.917 (Korneliussen et al., 2014) using the SAMtools (Li, 2011) model ("GL -1" option) with default filter settings, along with options as indicated above. These filtered genotype likelihoods were then used to infer major and minor alleles and calculate per-site allele frequencies. Allele frequency estimates for these sites were provided as priors (assuming Hardy-Weinberg equilibrium) for estimating genotype posterior probabilities for all loci mapped to putative autosomes. A covariance matrix between all samples was calculated based on the genotype probabilities using the ngsCovar utility in ngsTools (Fumagalli et al., 2014) and subjected to eigenvector decomposition to estimate the first two principal components.

Population structure was also inferred using ADMIXTURE (v1.3) (Alexander et al., 2009) with the maximum-likelihood approach. To explore genetic divergence among all individuals, the pre-defined genetic clusters (K) were set from 2 to 4, with 10 000 iterations for each run. We performed clustering analysis via the maximum-likelihood approach, implemented in ADMIXTURE, assuming 2–4 ancestral populations (K=2–10). The lowest cross-validation (CV) error was calculated for each model with nine modeled clusters. The clustering results (K=2–4) were then visualized using R.

ABBA-BABA analysis

ABBA-BABA analyses were performed with the ANGSD toolbox (v0.930) (Korneliussen et al., 2014) using *Lagopus lagopus* as an outgroup species. ABBA-BABA analyses were performed, and D statistics were calculated for analysis of all combinations of the Chinese Grouse and Hazel Grouse populations. To determine the significance of the D statistics, Z scores were calculated for 1 Mb blocks using jackknifing (Reich et al., 2009) with an R script from the ANGSD toolbox (Korneliussen et al., 2014).

Selective sweep analysis

To identify genome-wide selective sweeps associated with high-altitude adaptation, nucleotide diversity (π) and genome-wide distribution of the fixation index (F_{ST}) were calculated for the Hazel Grouse and Chinese Grouse using a sliding-window approach, with 100 kb windows and 50 kb increments. At each detected SNP position, the number of reads was counted corresponding to the most and least frequently observed allele in each group. All outlier windows were assigned to the corresponding SNPs and genes. To explore the evolution of functional categories, Kobas (Xie et al., 2011) was used to annotate the genes under selection in each species using the chicken genome (GRCg6a). These genes were submitted to Gene Ontology and KEGG databases for enrichment analysis. We used a false discovery rate (FDR)-corrected binomial distribution probability approach to test significant enrichment in gene function at $P<0.05$ (Benjamini & Hochberg, 1995).

Supplementary Tables

Supplementary Table S1. Sample information and whole-genome quality control statistics.

Species	Sample	Raw Reads	Raw Bases	Clean Reads	Clean Bases	Error Rate	Q20	Q30	GC Content
Chinese Grouse	LHS01	143 628 342	21.54G	141 357 988	21.2G	0.01%	97.52%	94.18%	41.94%
Chinese Grouse	LHS02	139 214 378	20.88G	137 702 462	20.66G	0.01%	97.86%	94.86%	41.61%
Chinese Grouse	LHS03	161 232 574	24.18G	159 463 010	23.92G	0.01%	97.69%	94.48%	41.77%
Chinese Grouse	LHS04	123 347 872	18.5G	121 834 992	18.28G	0.01%	97.71%	94.58%	41.74%
Chinese Grouse	B151466	152 899 516	22.93G	148 601 096	22.29G	0.02%	96.96%	92.91%	41.68%
Chinese Grouse	B1630	138 806 838	20.82G	133 756 888	20.06G	0.02%	97.01%	93.12%	42.69%
Chinese Grouse	B1729	142 167 828	21.32G	138 484 602	20.77G	0.02%	97.08%	93.19%	41.16%
Chinese Grouse	B1791	158 510 302	23.77G	151 902 920	22.79G	0.02%	96.98%	93.12%	43.19%
Chinese Grouse	B205	153 155 858	22.97G	147 789 652	22.17G	0.02%	96.93%	92.95%	43.60%
Chinese Grouse	B3-151032	135 113 920	20.26G	129 715 362	19.46G	0.02%	96.68%	92.46%	42.68%
Chinese Grouse	BS01	149 007 364	22.35G	147 214 746	22.08G	0.01%	97.68%	94.51%	41.94%
Chinese Grouse	BS02	140 641 794	21.09G	139 032 212	20.85G	0.01%	97.72%	94.58%	42.16%
Chinese Grouse	BS03	128 238 154	19.23G	126 787 218	19.02G	0.01%	97.65%	94.37%	42.04%
Chinese Grouse	ZN01	110 715 764	16.6G	109 339 154	16.4G	0.01%	97.69%	94.52%	42.35%
Chinese Grouse	ZN02	128 906 698	19.33G	127 424 978	19.11G	0.01%	97.81%	94.77%	41.89%
Chinese Grouse	ZN03	146 153 638	21.92G	143 980 604	21.6G	0.01%	97.61%	94.35%	42.31%
Hazel Grouse	XLJ01	127 430 906	19.11G	124 639 170	18.7G	0.01%	97.25%	93.88%	42.62%
Hazel Grouse	XLJ02	106 533 868	15.98G	104 792 878	15.72G	0.01%	97.60%	94.36%	42.35%
Hazel Grouse	XLJ03	127 053 362	19.05G	124 289 282	18.64G	0.01%	97.63%	94.62%	43.00%

Hazel Grouse	XLJ04	149 054 212	22.35G	146 683 052	22G	0.01%	97.72%	94.52%	42.23%
Hazel Grouse	XLJ05	177 222 072	26.58G	174 237 468	26.14G	0.01%	97.46%	94.22%	42.49%
Hazel Grouse	JHGO005	141 679 556	21.25G	138 673 432	20.8G	0.01%	97.00%	93.31%	43.46%
Hazel Grouse	JHGO006	138 957 440	20.84G	111 051 668	16.66G	0.01%	97.36%	93.95%	49.15%
Hazel Grouse	JHGO197	136 884 104	20.53G	134 306 646	20.15G	0.01%	97.43%	94.01%	42.53%
Hazel Grouse	JHGO046	142 837 302	21.42G	139 529 196	20.93G	0.01%	97.38%	94.08%	42.77%
Hazel Grouse	JHGO047	138 790 648	20.81G	136 819 280	20.52G	0.01%	97.63%	94.39%	42.05%
Hazel Grouse	JHGO048	165 487 810	24.82G	162 663 952	24.4G	0.01%	97.37%	93.91%	42.58%
Hazel Grouse	F3	174 722 516	26.2G	170 426 708	25.56G	0.02%	96.60%	92.21%	43.84%
Hazel Grouse	M3	147 096 620	22.06G	142 298 546	21.34G	0.02%	96.62%	92.27%	42.35%
Rock Ptarmigan	WIPI-NL-1012	162 187 862	24.32G	157 268 684	23.59G	0.02%	96.95%	92.99%	43.20%
Rock Ptarmigan	JHGO-271	165 026 116	24.75G	158 321 760	23.75G	0.01%	97.05%	93.23%	42.39%
Willow Ptarmigan	JHGO272	112 075 052	16.81G	108 520 580	16.28G	0.01%	97.20%	93.44%	42.37%

Supplementary Table S2. Nucleotide diversity (π) and Tajima's D across different sampling locations.

	Sample Location	No. of Samples	π	Tajima's D
Chinese Grouse	all	16	5.19×10^{-4}	1.49
Chinese Grouse	LHS	10	5.02×10^{-4}	1.24
Chinese Grouse	ZN	3	4.76×10^{-4}	0.33
Chinese Grouse	QLS	3	2.85×10^{-4}	0.30
Hazel Grouse	all	13	9.32×10^{-4}	1.59
Hazel Grouse	XLJ	5	8.49×10^{-4}	0.95
Hazel Grouse	SWE	3	2.94×10^{-4}	-0.10
Hazel Grouse	GER+NE Poland	5	6.70×10^{-4}	0.90

Supplementary Table S3. F_{ST} across different sampling locations.

	LHS	ZN	XLJ	SWE
Chinese grouse				
ZN	0.04	-		
QLS	0.19	0.20		
Hazel grouse				
SWE			0.26	-
GER+NE Poland			0.26	0.35

Abbreviations

LHS: Lianhuashan National Nature reserve

ZN: Zhuoni County

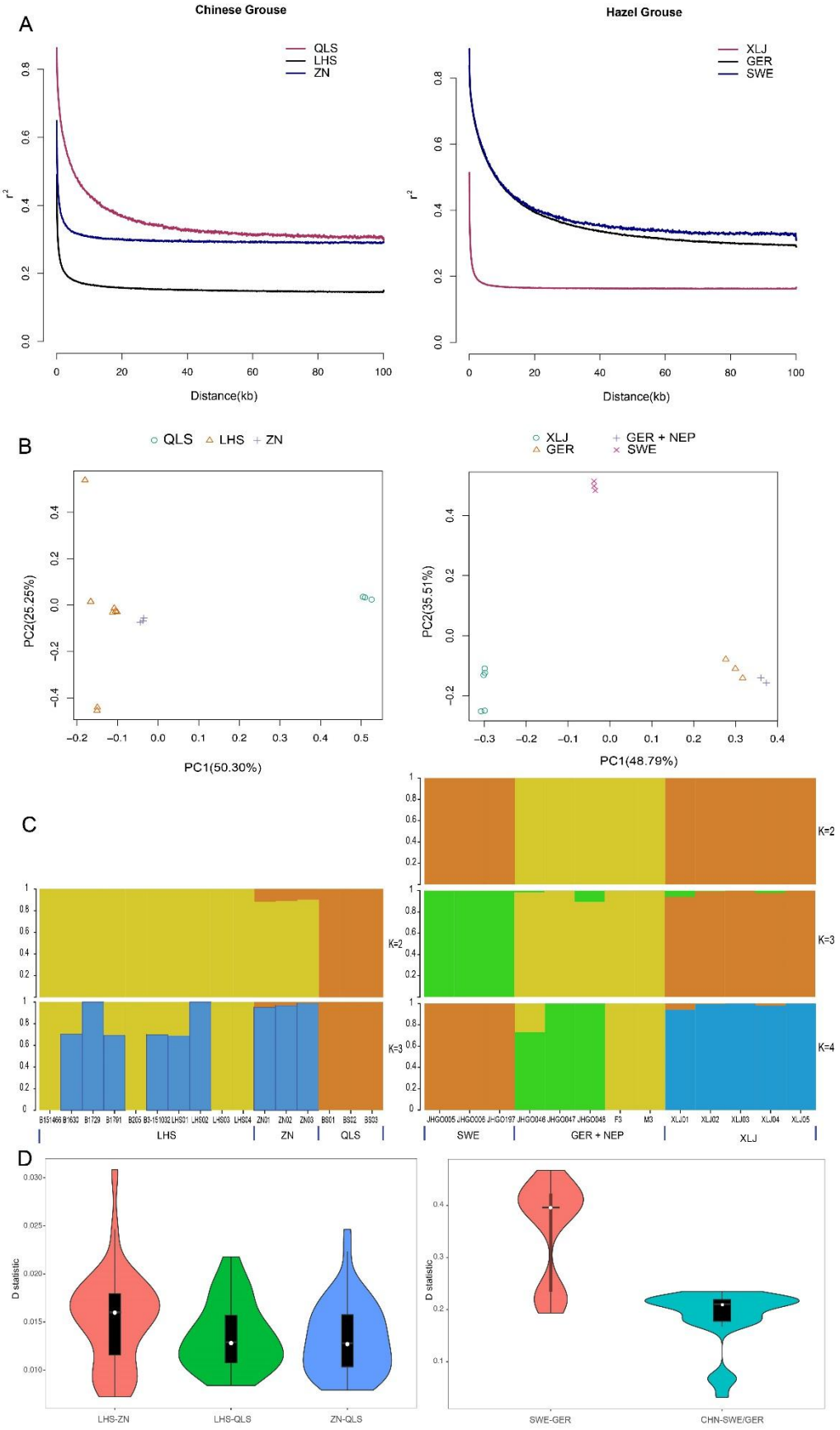
QLS: Qilian Mountains

XLJ: Northeastern China

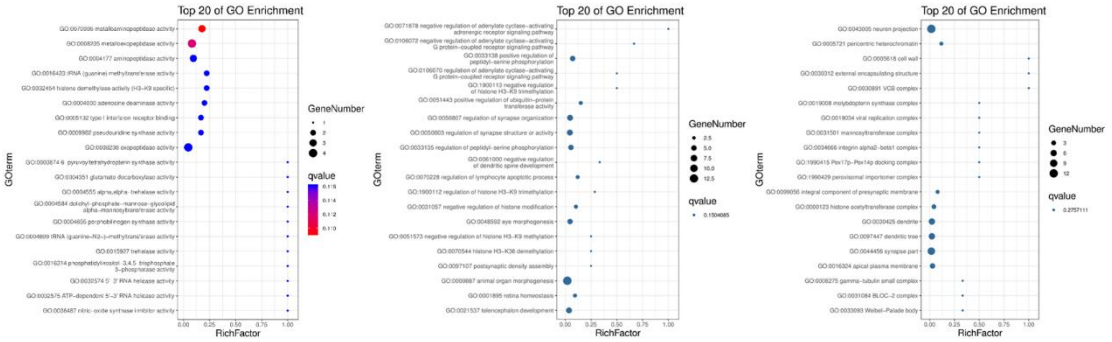
SWE: Sweden

GER: Germany

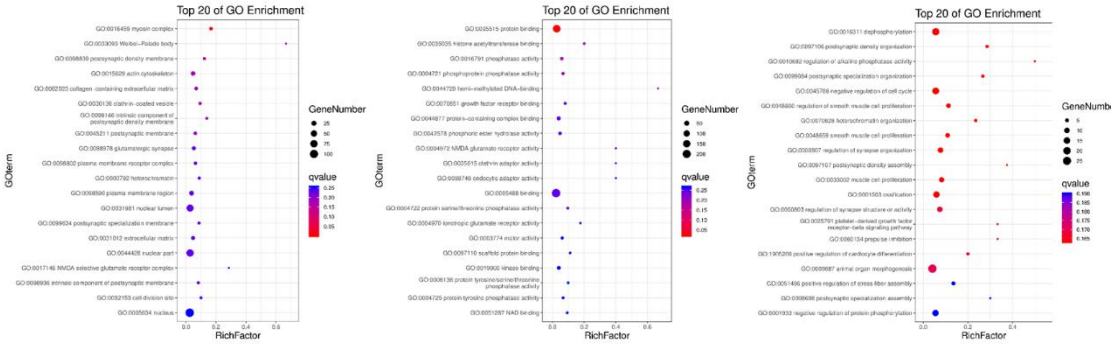
NE Poland: North East Poland


A
Chinese Grouse
QLS
LHS
ZN
r²
Distance(kb)
Hazel Grouse
XLJ
GER
SWE
B
QLS LHS ZN
PC2(25.25%)
PC1(50.30%)
XLJ
GER
GER + NEP
SWE
PC2(35.51%)
PC1(48.79%)
C
K=2
K=3
K=4
LHS
ZN
QLS
SWE
GER + NEP
XLJ
D
D statistic
LHS-ZN
LHS-QLS
ZN-QLS
SWE-GER
CHN-SWE/GER

Supplementary Figure S1. Population structure of Chinese Grouse and Hazel Grouse. **A**: Patterns of linkage disequilibrium (LD) decay across genome in different geographic populations. R^2, Pearson's correlation coefficient**. B**: Principal component analysis of Chinese Grouse and Hazel Grouse. **C**: Population structure inferred from whole-genome resequencing data using ADMIXTURE. **D**: D statistics from ABBA-BABA tests showing introgression among different populations.



Supplementary Figure S2. Three categories, i.e., cellular component (CC), molecular function (MF), and biological process (BP), of gene enrichment in Chinese Grouse from Gene Ontology (GO) database



Supplementary Figure S3. Three categories, i.e., cellular component (CC), molecular function (MF), and biological process (BP), of gene enrichment in Hazel Grouse from GO database.

REFERENCES

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**(9): 1655–1664.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(1): 289–300.

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, et al. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, **195**(3): 979–992.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. *ngsTools*: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics,* **30**(10): 1486–1487.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, **27**(21): 2987–2993.

Quintana-Murci L. 2016. Genetic and epigenetic variation of human populations: an adaptive tale. *Comptes Rendus Biologies*, **339**(7–8): 278–283.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature*, **461**(7263): 489–494.

Song K. 2020. Divergence, Selection, Demographic History and Conservation Genomics of Sibling Bird Species in Boreal Forest in Northern Eurasia and the Qinghai-Tibetan Plateau. Ph.D. dissertation, Uppsala University, Uppsala.

Song K, Gao B, Halvarsson P, Fang Y, Jiang YX, Sun YH, et al. 2020. Genomic analysis of demographic history and ecological niche modeling in the endangered Chinese Grouse *Tetrastes sewerzowi*. *BMC Genomics*, **21**(1): 581.

Song K, Gao B, Halvarsson P, Fang Y, Klaus S, Jiang YX, et al. 2021. Demographic history and divergence of sibling grouse species inferred from whole genome sequencing reveal past effects of climate change. *BMC Ecology and Evolution*, **21**(1): 194.

Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, et al. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, **39**(S2): W316–W322.

Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**(10): 1786–1788.