Title Page

**Parapatric speciation with recurrent gene flow of two sexual
dichromatic pheasants**

Zheng Li¹, Jie Zhou¹, Minzhi Gao¹, Wei Liang², Lu Dong^{1,*}

1. Ministry of Education Key Laboratory for Biodiversity and Ecological
Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875,
China

2. Ministry of Education Key Laboratory for Ecology of Tropical Islands, College of
Life Sciences, Hainan Normal University, Haikou 571158, China

*Correspondence: donglu@bnu.edu.cn

Abstract

Background: Understanding speciation has long been a fundamental goal of evolutionary biology. It is widely accepted that speciation requires an interruption of gene flow to generate strong reproductive isolation between species, in which sexual selection may play an important role by generating and maintaining sexual dimorphism. The mechanism of how sexual selection operated in speciation with gene flow remains an open question and the subject of many research. Two species in genus *Chrysolophus*, Golden pheasant (*C. pictus*) and Lady Amherst's pheasant (*C. amherstiae*), which both exhibit significant plumage dichromatism, are currently parapatric in the southwest China with several hybrid recordings in field.

Methods: In this research, we estimated the pattern of gene flow during the speciation of two pheasants using the Approximate Bayesian Computation (ABC) method based on the multiple genes data. With a new assembled *de novo* genome of Lady Amherst's pheasant and resequencing of widely distributed individuals, we reconstructed the demographic history of the two pheasants by pairwise sequentially Markovian coalescent (PSMC).

Results: The results provide clear evidence that the gene flow between the two pheasants were consistent with the prediction of isolation with migration model for allopatric populations, indicating that there was long-term gene flow after the initially divergence (ca. 2.2 million years ago), and further support the secondary contact when included the parapatric populations since around 30 ka ongoing gene flow to now, which might be induced by the population expansion of the Golden pheasant in late Pleistocene.

Conclusions: The results of the study support the scenario of speciation between Golden pheasant (*C. pictus*) and Lady Amherst's pheasant (*C. amherstiae*) with cycles

53 of mixing-isolation-mixing due to the dynamics of natural selection and sexual
 54 selection in late Pleistocene that provide a good research system as evolutionary model
 55 to test reinforcement selection in speciation.

56

57 **Keywords:** Golden pheasant (*Chrysolophus pictus*), Lady Amherst's pheasant
 58 (*Chrysolophus amherstiae*), speciation, gene flow, Approximate Bayesian
 59 Computation (ABC), Pairwise Sequentially Markovian coalescent (PSMC).

60

61

Background

The fascination that evolutionary biologists have long had with speciation has increased in recent decades because of a growing understanding of the complex mechanisms that shape biodiversity. Allopatric speciation driven by geographical isolation, implying in the absence of gene flow, has been the standard model to interpret the classical speciation progress (Mayr 1942; Coyne & Orr 2004). However, allopatric speciation models without gene flow seem oversimplified since the recurrence of land bridges during glacial periods of Pleistocene has provided opportunities for gene exchange between populations (Rohling et al. 1998, Lambeck and Chappell 2001). Since more empirical examples of divergence in sympatry or parapatry were reported, it is widely accepted that speciation can occur in the presence of gene flow (Rice and Salt 1988; Bolnick and Fitzpatrick 2007; Nosil 2008). To understand how speciation may proceed in the face of gene flow, one must understand the pattern and role of gene flow in the speciation process.

Many researches so far has focus on the speciation process in monomorphic species of birds, such as Hwameis (Li et al. 2010), Crows (Poelstra et al. 2014), Darwin's Finches (Lamichhaney et al. 2015) and Bamboo Partridges (Wang et al. 2021), most of that revealed the important impacts of disruptive selection or assortative mating that promote the speciation with gene flow. However, the research on evolutionary mechanisms and speciation in sexual dichromatic species has been lagging behind (Servedio 2011). Compared with sexual monomorphic species, sexual dichromatic species might be exposed to a greater selection, namely Fisherian sexual selection, which plays an important role in driving population divergence as an essential driver of pre-zygotic isolation. However, there are only some theoretical researches based on model simulation (Servedio 2012; Servedio and Bürger 2014; Servedio 2016) for the

pattern and role of gene flow in speciation under Fisherian sexual selection (but see Dong et al. 2013 for a case study in birds). More empirical researches are still needed to support theoretical expectations.

To examine the pattern and role of gene flow in speciation between sexual dichromatic species, we focus specifically on two species in genus *Chrysolophus*, golden pheasant (*C. pictus*) and Lady Amherst's pheasant (*C. amherstiae*) which both exhibit significant plumage dichromatism. The males of golden pheasant and Lady Amherst's pheasant have extreme phenotypic differences in their plumage color, but the females are very similar, both with a duller mottled brown plumage. The golden pheasants are mainly distributed in the mountains of central and west China from the Qinling Mountains in the north to the Minshan Mountain in the west, while the Lady Amherst's pheasants are mainly distributed in the Yunnan-Guizhou Plateau and the western and southern edge of the Sichuan Basin (Fig. 1). These two pheasants are currently parapatric in southwest China with several hybrid recordings in west Sichuan (He and Wei 1993; Shi et al. 2018) and northeast Yunnan (Deng 1974), and even a hybrid recording in a painting attributed to Emperor Huizong of the Song Dynasty around one thousand years ago (Peng et al. 2016). These ongoing hybridization records implied incomplete pre-zygotic reproductive isolation in the sympatric zone. And the captive observation also revealed incomplete post-zygotic reproductive isolation in the F1 hybrids and backcross hybrids with higher proportion of the male hybrids which corresponded to the prediction of Haldane's rule (Arrieta et al. 2013).

The niche model analysis explained the divergence history since the last interglacial (LIG) of the two pheasant species. The significantly different ecological niches occupied by the *C. pictus* and *C. amherstiae* according to current distribution and the isolated glacial shelters in the late Pleistocene may promote the allopatric

speciation of the two species (Lyu et al. 2015). However, without genetically based coalescence analysis, this model is limited to reconstructing the divergent history before the segregation of the two species before the LIG and distinguishing the mode of the speciation process, thus what happened since the divergence of *C. pictus* and *C. amherstiae* cannot be estimated. The ecological niches between the two *Chrysolophus* species were different in the late Pleistocene, allopatric with distribution gap, and re-parapatric in/after LGM that may be induced by expansion of the golden pheasant, that provide a good research system for studying the speciation with gene flow under both sexual selection and natural selection.

In this study, we combined the multi-locus data from the two closely related species of *Chrysolophus* and a *de novo* genome of *C. amherstiae* with several resequencing individuals from two pheasant species to achieve two main goals. First, we aimed to detect the candidate hybrids with phylogenetic analysis and further explore the mode of gene flow in speciation of *C. pictus* and *C. amherstiae* through Approximate Bayesian Computation (ABC) analysis (Beaumont et al. 2002). Second, we employed pairwise sequentially Markovian coalescent (PSMC) analyses (Li and Durbin 2011) to inspect the evolutionary history between *C. pictus* and *C. amherstiae*, basing on the draft genome and re-sequencing data. These analyses allowed us to explore the pattern and role of gene flow in speciation of sexual dichromatic species and provide a better understanding of how speciation with gene flow proceed.

Methods

Sampling, amplification and sequencing

A total of 38 individuals (23 individuals of *C. pictus*, 15 individuals of *C. amherstiae*) from 15 locations were sampled across the geographic range of *C. pictus* and *C. amherstiae*, 30 of them were used for ABC analysis and 19 for PSMC analysis (Fig. 1; Table S1). One Temminck's Tragopan *Tragopan temminckii*, collected from Shimian County, Sichuan Province, was used as the outgroup in phylogenetic analysis. DNA was extracted from ethanol-preserved muscle, blood or feather using TIANamp Genomic DNA kit (Tiangen Biotech Ltd., Beijing) following the manufacturer's protocol and dissolved in 100 µl of TE buffer. In total of 37 exon loci (Table S2) including 34 unlinked autosomal loci and two Z-linked loci developed by Liu et al. (2018) and the mitochondrial cytochrome b (*cyt b*) gene were amplified by polymerase chain reaction (PCR). The Touchdown PCR (Don et al. 1991) was performed in a PTC-220 PCR thermal cycler system (Bio-Rad, USA). The 40 µl reaction for PCR contained 2.4 µl template DNA, with 20ul mixed concentrations of 2 × ExTaq buffer (Takara, Japan), 0.8ul of each forward and reverse primer (10uM), and 16ul ddH₂O. The initial denaturation at 94°C for 3 min, followed by 20 cycles of 94°C for 30 s, 62–52 °C (decreasing the annealing temperature by 0.5 °C per cycle) for 30 s and 72 °C for 1min, and another 15 cycles of 94°C for 30 s, 52 °C for 30s and 72 °C for 1min, ended with a final extension at 72°C for 10 min. PCR products were sequenced bidirectionally using the 3730XL series platform Sequencer (Applied Biosystems, USA).

Sequences analysis

The obtained sequences were edited using the software CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA), aligned with Clustal W (Thompson et al. 1994) in MEGA v5.10 (Tamura et al. 2013) and removed the sequences with

multiple deletions. Haplotypes of the diploid nuclear sequences were separated using DNASP v5.10.01 (Librado and Rozas 2009) based on Markov chain Monte Carlo (MCMC) simulation approach with 500 bootstrap replicates and the results of the first 100 times were discarded. Due to the small sample size, haplotypes of the diploid nuclear sequences of *T. temminckii* were manually separated. The phased nuclear sequences and cytb sequences were aligned between *C. pictus*, *C. amherstiae* and *T. temminckii* with Clustal W (Thompson et al. 1994) in MEGA v5.10 (Tamura et al. 2013) and saved as fasta format for subsequent analyses.

We tested the neutrality of each gene using Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1987) and Tajima's D (Tajima 1989). We calculated intra- and interspecific summary statistics for DNA polymorphisms of each gene (Table S3) by DnaSP v5.10.01 (Librado and Rozas 2009) for the subsequent ABC analysis (Beaumont et al. 2002). No significant λ^2 (HKA) and Tajima's D were observed in all genes except for the Tajima's D of the mitochondrial cytb gene (Table S4). Nuclear substitution rate was calculated by taking the average ratio of within group and outgroup pairs and then multiplied it by the rate of 1.035×10^{-8} per site per million years (molecular clock for the passerine mitochondrial cytb gene, Weir and Schluter 2008). The mitochondrial cytb gene was not included in the ABC and PSMC analyses because its evolutionary rate highly differs from that of nuclear genes.

Detection of hybrids

We used the NewHybrids (Anderson and Thompson 2002) to calculate the posterior probability of sampled individuals belonging to parental or hybrid categories based on the genotypic information of the two pheasants. In this study, individuals collected 50

km away without niche similarity from the sympatric zone (Lyu et al. 2015) were used as the parent group to estimate the allele frequencies of individuals collected in the suture zone or apart within 50 km. The dataset was analyzed three times using the uniform prior for both θ and the mixing proportion π . The burn-in periods and numbers of sweeps were set as recommended value in the manual. The mean value was used to infer hybrids between the two pheasants.

To infer the phylogenetic relationship between pheasants sampled, we used the ancestral reconstruction algorithm implemented in BEAST 1.8 (Drummond et al. 2012) based on autosome nuclear genes, Z chromosome genes and cyt b gene respectively. The substitution model of cyt b gene were determined with jModelTest2.1.3 (Darriba et al. 2012) based on the Bayesian information criterion (BIC). Strict clock was set for each gene, the rate of cyt b gene were set to 1.0 and the substitution rate of nuclear genes were estimated relative to the rate of cyt b gene. Length of MCMC chain for the BEAST analysis were set at 1×10^7 steps and the first 10% steps of each run were discarded as burn-in. Convergence were checked with Tracer 1.5 (Rambaut and Drummond 2009) to make sure the ESS values of main parameters above 200.

Approximate Bayesian Computation (ABC)

To infer the speciation history of the two *Chrysolophus* species, we used ABC analysis, which has been widely applied to phylogeographic studies (Beaumont et al. 2002; Beaumont 2010; Bertorelle et al. 2010; Duchon et al. 2013; Cornejo-Romero et al. 2017) to statistically test four alternative models of demographic history: (1) Isolation model, no gene flow occurs throughout the whole divergence process. (2) Isolation with migration model (IM model), gene flow continuously occurs after an initial population

split. (3) Early gene flow model, gene flow only occurs at the initial stages of divergence. (4) Secondary contact model, gene flow only exists in the late period of divergence. The parameters to be estimated for the four models are as follows: divergence time (T_{div}), Effective population size (N_e) of *C. pictus* (N_{cp}) and *C. amherstiae* (N_{ca}), N_e of the most common ancestor (N_A). For the model with gene flow, the number of migratory in each generation was estimated: M_1 represents the number of migrants of *C. pictus*, M_2 represents the number of migrants from *C. amherstiae*, m_{cp} represents the migration rate of *C. pictus*, and m_{ca} represents the migration rate of *C. amherstiae*.

For each speciation model, we used the software msABC (Pavlidis et al. 2010) to perform multi-locus coalescent simulations for 1×10^6 times and each divergence time was normalized with $4N_e$. Demographic parameters were randomly sampled from a uniform distribution within the prior ranges. The ranges of the priors for each demographic parameter in all models and the command line for each model used in msABC are listed in Additional file 1: Tables S5 and S6, respectively.

We performed the model selection, and parameter estimation under the most probable model using “abc” package (Csilléry et al. 2012) in R statistical programming environment (R Core Team 2020). With the 1×10^6 simulations, posterior probabilities for each demographic model were computed using a multinomial logit regression (tolerance level=0.001) (Beaumont et al. 2008). We selected the model with the highest posterior probability as the most probable model and performed 3×10^6 coalescent simulations under the most probable model to calculate the parameters. The logit transformations of parameters were applied to estimate the posterior distributions of parameters under the most probable model by neural network regression (Blum and François 2010). Each simulated dataset was weighted by its Euclidian distance through

an Epanechnikov kernel (Jean-Marie et al. 2008), and shown in mean and median (Table 1).

Considering that the sympatric populations and hybrids may affect the result of model selection, analyses were conducted on two datasets respectively: (1) Allopatric populations of *Chrysolophus* (excluding the individuals collected from the hybrid zone and 50 km apart from the hybrid zone: site 1, 7 and 9 in Fig. 1); (2) All sampled individuals of two *Chrysolophus* species.

Whole-genome assembly and mapping of reads

De novo sequencing of an individual of one male *C. amherstiae* (collected in Dali, Yunnan, China) was carried out on an Illumina HiSeq 2000 platform, by Novogene Company (Beijing, China). In total of 116.4Gb data was sequenced and assembled by ALLPATHS-LG v41687 (Butler et al. 2008) using default parameters (Table S7).

We resequenced 19 individuals from two *Chrysolophus* species (13 golden pheasants and six Lady Amherst's pheasants, Table S4) with an average of 35× depth using Illumina NovaSeq 6000 dual-index sequencing libraries by Novogene Company (Beijing, China). The trimmed clean reads of each individual were mapped back onto the assembled reference genome of *C. amherstiae* by BWA-MEM v0.7.12 (Li 2013) using default parameters. We converted SAM files into binary BAM files, sorted the BAM files and removed the PCR repeats using SAMtools v0.1.19 (Li et al. 2009). The final files were used as the input files of PSMC analyses.

PSMC analyses

By utilizing information from whole genome of a single diploid individual, the PSMC analysis (Li and Durbin 2011) can represent an appropriate method for estimating population size changes deep into the early Pleistocene. We applied the PSMC approach on the genomes of *C. pictus* (13 individuals) and *C. amherstiae* (6 individuals) to reconstruct demographic history of each species. The consensus sequence (in BAM format) for each individual of each species was divided into non-overlapping 100-bp bins, where each bin is scored as heterozygous ('1') if there is a heterozygote in the bin, or as homozygous ('0') otherwise implemented with the fq2psmcfa command from R package "psmc" and the quality cut off (-q) was set to 20. The psmcfa sequences were taken as the input of the PSMC estimation to deduce the past N_e over time. The number of iterations (-N) was set to 25, Tmax (-t) was set to 15, initial mutation/recombination ratio (-r) was set to 5 and the atomic time interval pattern (-p) was '4+25*2+4+6'. The psmc file was visualized with the mutation rate (μ) being set to 3.8×10^{-9} per site per generation (generation time = 2 years) based on the mean nuclear substitution rate of avian (Zhang et al. 2014).

Results

Identification of hybrids

The Bayesian tree demonstrated substantial support for mutual monophyletic clades of the two pheasant species with the exception of incongruent phylogenetic position of one male Golden Pheasant (i.e. SCPW-02) collected near the suture zone (Pingwu, Sichuan) using different kind of molecular markers. The alleles of the male were clustered with the *C. amherstiae* when the autosome nuclear loci were employed to reconstruct the Bayesian phylogenetic relationship (Fig. S1), while were clustered with

the *C. pictus* clade reconstructed by the mitochondrial cyt b gene (Fig. S2). When the Z-link alleles were applied, the male alleles were found to be clustered with the two clades, respectively (Fig. S3).

The phylogenetic incongruence was highly supported by the inference from the NewHybrids analyses that the male was detected as the hybrid of F2 (posterior probability = 0.493) or backcross (posterior probability = 0.493) (Fig. 1).

Evolutionary history with ABC estimation

Since we performed ABC analyses on the two datasets separately, the results of estimation will be presented respectively in the following paragraphs.

For allopatric population of *Chrysolophus*, the best-supported model corresponds to an IM model (posterior probability = 0.97, Fig. 2b), indicating that ongoing gene flow have been existing between *C. pictus* and *C. amherstiae* throughout the whole divergence process. The divergence time of *C. pictus* and *C. amherstiae* was estimated as 2.23 million years ago (Mya), and the gene flow from *C. pictus* to *C. amherstiae* ($M_2 = 2.00$) was higher than that from *C. amherstiae* to *C. pictus* ($M_1 = 1.054$) during the *Chrysolophus* divergence. The N_e of *C. amherstiae* ($N_{ca} = 1.191 \times 10^5$) was much larger than that of *C. pictus* ($N_{cp} = 0.617 \times 10^5$) and both of them increased compared with the ancestor population size (0.35×10^5 , Table S5). All of estimations mentioned above were conducted under the IM model.

For all sampled individuals of *Chrysolophus*, the best-supported model is the same IM model (posterior probability = 0.55, Fig. 2f), however the posterior probability of the secondary contact model significantly increased compared with allopatric population estimation (posterior probability = 0.45, Fig. 2h). The divergence time of *C.*

pictus and *C. amherstiae* under the best-supported IM model was estimated as 1.86 Mya, and the gene flow of the two pheasants was still asymmetric that gene flow from *C. pictus* to *C. amherstiae* remains higher (Table 1). We further estimated the parameters of population dynamics under secondary contact model since the posterior probabilities of secondary contact model and isolation with migration model were very close. The initial secondary contact time of the two pheasants is estimated to be about 25,800 years ago, indicating the new hybrid zone was formed after the last glacial maximum. Gene flow remains the same trend with the estimation under IM model but migration has increased in both directions compared with the IM model (Table 1). Consequently, Medium to high gene flow maintained in the speciation of the two pheasants in both categories with all models.

Historical demography with PSMC

We sequenced and assembled the genome of *C. amherstiae* to serve as the reference. The genome assembly of *C. amherstiae* was 980Mb long (genomic coverage = 118.8×) and contained 21,996 scaffolds (N50 = 399Kb) (Table S8).

We conducted PSMC analyses to reconstruct the historical demography in both species of Phasianidae spanning the whole Pleistocene period, extending as far back as 5 million years (Fig. 3). The populations of two species of *Chrysolophus* were very similar in their estimated N_e history around 2 Mya, which is potentially consistent with the divergence time (2.23 Mya for allopatric population, 1.86 Mya for all sampled individuals) estimated by ABC analyses. At the start of the last ice age (110 thousand years ago, kya), the N_e of *C. pictus* increased slightly, and gradually decreased during the Last Interglacial (LIG), and ended with a sharp increase before the Last Glacial

Maximum (LGM). From a peak about 30 kya, the N_e of *C. pictus* dropped by a factor of ten between 30 and 20 kya, throughout the LGM. Interestingly, the N_e of *C. amherstiae* showed a opposite trend to the N_e of *C. pictus*, decreasing from the last ice age to the beginning of LIG. We then saw a gradually increase in estimated N_e of *C. amherstiae* during LIG, and also a slightly growth in LGM.

Discussion

Gene flow is conventionally considered as a homogenizing force that can moderate population divergence and impede speciation in the strict allopatric model of speciation. In this study, with multi-loci analysis by Approximate Bayesian Computation, we detected ancient and recurrent gene flow during the speciation process of the Golden pheasant and Lady Amherst's pheasant, suggesting that the incomplete reproductive isolation has long existed between the two sexual dichromatic species. With genomic data-based PSMC analysis, we also show that the population expansion of Golden pheasant since Last Interglacial may played an essential impact in shaping the current suture zone and inducing hybridization in the wild.

Ancient and recurrent gene flow

The evolution of reproductive isolation that underpins the Biological Species Concept allows us to reconsider the role of gene flow in speciation. Much of the studies implies that the types of reproductive isolation have been classified into two main categories: pre-zygotic isolation (e.g., assortative mating, sexual selection) and post-zygotic isolation (i.e., low hybrid fitness) (Coyne and Orr 2004; Nosil 2012). Gene flow during speciation may strengthen the barrier of reproductive isolation through reinforcement acts on pre-zygotic selection, as long as selection is not swamped by migration

(Servedio and Kirkpatrick 1997; Kirkpatrick and Servedio 1999; Servedio 2000; Servedio 2004).

According to the results of ABC analyses for the allopatric populations of *Chrysolophus*, there has been long-term gene flow since the initial divergence of *C. pictus* and *C. amherstiae* two million years ago (Table 1). The ancient (or early) gene flow with speciation is not unusual in east Asia especially in the species complex distributed across the continental Taiwan Island and mainland China, e.g., the Hwamei (Li et al., 2010), bush warblers (Wei et al. 2019) and bamboo-partridges (Wang et al. 2021) or the closely related species diverged in the east edge of Qinghai-Tibet Plateau, e.g., the long-tailed tits (Wang et al. 2014), eared pheasants (Wang et al. 2017) and leaf-warblers (Zhang et al. 2019). The ecological divergent selection for local adaptation is often considered as the important factor that promotes ecological-driven speciation in the face of gene flow between these species without sexual dichromatic traits. Our results provided a new case that there was medium-level gene flow (Table 1) between the ancient populations of the two *Chrysolophus* species with significantly different ecological niches. However, the ecological segregation that limited the gene flow during the glacial time (see Lyu et al. 2015) may not be the major determinants for the separation of these two sexual dichromatic species.

The occasional hybrid records of Golden pheasants and Lady Amherst's pheasants from the suture zone (see Deng 1974, He et al 1993, and Shi et al. 2018) and the less inviability of F₁ hybrids or backcrosses in captivity (Arrieta et al. 2013) showed the probability that the current gene flow could be maintained due to the incompletely developed pre- and post-zygotic isolation. Our genetic data analysis revealed that when sympatric samples were included, the secondary contact model was better supported than when just allopatric samples were used. And the current high-level gene flow (Fig.

2) between the two pheasant species in the suture zone may have occurred after LIG (95% HPD: $0.13-1.206 \times 10^5$ years before) which is consistent with the estimation of expansion time based on the ecological niche model (see Lyu et al. 2015) and our PSMC estimation (Fig. 3). On the other side, we could not refuse the prediction of isolation with migration model with or without the sympatric samples (Table 1) allowing long-term gene flow since their divergence, which demonstrated the scenario of mixing-isolation-mixing process with recurrent gene flow should occur during their speciation. Since the two categories of factors that contributes to reproductive isolation, i.e., sexual dichromatic plumage which were usually attributed to the consequences of sexual selection that benefit to maintain the pre-zygotic isolation (Servedio and Noor 2003), and ecological segregation which may contribute more to the post-zygotic isolation (Funk et al. 2006), were both not completely developed between the Golden and Lady Amherst's pheasant, with the fact that they were still able to maintain the independence of species level with divergent male plumage color and recurrent gene flow, we believe that the reinforcement selection on both sexual and ecological traits may play a significant role in the divergence and speciation of the two pheasant species. Further research into the fitness and attractiveness of the backcross or F_1 hybrids in the suture zone would be helpful to undercover the mechanisms on the selection against the gene flow.

Heterogeneous pattern of population demographic in LGM

It is widely perceived that climate change in the Pleistocene provided conditions for speciation (Hewitt 2000; Hewitt 2001). Gene flow might increase or decrease with climate change, depending on how climate change affects the habitat (Wasserman et al. 2010; Olsen et al. 2011). The cyclical climate changes in quaternary period (beginning

2.4 Mya), especially the LIG (between 130 kya and 70 kya) and LGM (between 26 kya and 18 kya), strongly affected the formation and distribution of species through “expansion-contraction” of their ranges (Hewitt 2000; Hewitt 2004; Provan and Bennett 2008). In East Asia, the impact of glacial period on the effective population size illustrated a different trend from that of high-latitude regions such as Europe and North America: For example, advancing ice sheets in the drove the expansion of polar biota and the contraction of temperate biota to refugia in glaciated regions such as Europe (Hewitt 2004; Provan and Bennett 2008), North America (Shafer et al. 2010) and Antarctica (Fraser et al. 2012). However, east Asia in the subtropical zone was not covered by ice sheets during the entire Pleistocene (Zhou et al. 2004), and the effective population size showed a contraction during the LIG and expansion during the LGM (Dong et al. 2017). According to the estimation of PSMC, the population dynamics of *C. pictus* and *C. amherstiae* were opposite, which can be inferred that LGM and LIG both had different effects on the two species, and the expansion of *C. pictus* during the recent secondary contact may provide opportunities for gene exchange (Fig. 3).

Potential evolutionary processes contribute to reproductive isolation of the two sexual dichromatic pheasants

The evolutionary pattern of *Chrysolophus* belongs to the isolation with migration model, which means *C. pictus* and *C. amherstiae* have chances to encounter and hybridize during the evolutionary history. However, they haven’t fused into one species, instead maintained differentiation that significantly differ from phenotypes and ecological niches in the presence of long-term gene flow. Since the hybrids of the two pheasants were able to interbreed and produce fertile offspring, indicating that the post-zygotic isolation of *C. pictus* and *C. amherstiae* has not been fully established, and the barrier

of reproductive isolation might be mainly improved by the promotion of pre-zygotic isolation (shown as sexual selection in our system). Each occurrence of gene flow is equivalent to a secondary contact, which increasingly strengthened the pre-zygotic isolation by reinforcement not only in the divergent male plumage color between the two pheasant species, but also may act on the cryptic divergence of female preference in male's color which was beneficial to maintaining speciation with the recurrent gene flow.

Conclusion

The results of the study support the scenario of speciation between Golden pheasant (*C. pictus*) and Lady Amherst's pheasant (*C. amherstiae*) with cycles of mixing-isolation-mixing due to the dynamics of natural selection and sexual selection in late Pleistocene. This research provided a good research system as an evolutionary model to test reinforcement selection in speciation and to explore the genetics mechanisms for the origin and maintenance of reproductive isolation between sexual dichromatic species.

Acknowledgments

We are grateful to Wanglang National Nature Reserve, Yanyun Zhang, Chao Zhao, Zhixin Zhou and Yiqiang Fu for their help in collecting samples. We thank Bowen Zhang, Yu Hao and Jindan Guo for assistance in statistics.

Authors' contributions

L.D. and J.Z. conceived the study; L.D. and W.L. collected the key samples; Z.L., J.Z. and M.G. performed the study and analyzed data; Z.L., J.Z. and L.D. drafted the manuscript with help from M.G. and W.L. All authors approved the final version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No.31471987).

Availability of data and materials

The dataset used in the present study will be deposited to Genbank after the manuscript is accepted.

Ethical note

The research protocol was approved by College of Life Sciences, Beijing Normal University: No. CLS-EAW-2013-007

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interests.

Reference

Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 2002;160(3):1217–1229.

Arrieta RS, Lijtmaer DA, Tubaro PL. Evolution of postzygotic reproductive isolation in galliform birds: Analysis of first and second hybrid generations and backcrosses. *Biol. J. Linn. Soc.* 2013;110(3):528–542.

Beaumont MA. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* 2010;41(1):379–406.

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics*. 2002;162(4):2025–35.

Beaumont O, Bonichon N, Duchon P, Eyraud-Dubois L, Larchevêque H. A distributed algorithm for resource clustering in large scale platforms. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 2008;5401 LNCS:564–7.

487 Bertorelle G, Benazzo A, Mona S. ABC as a flexible framework to estimate
488 demography over space and time: Some cons, many pros. *Mol. Ecol.* 2010;19(13):
489 2609–25.

490 Blum MGB, François O. Non-linear regression models for Approximate Bayesian
491 Computation. *Stat. Comput.* 2010;20(1):63–73.

492 Bolnick DI, Fitzpatrick BM. Sympatric Speciation: Models and Empirical Evidence.
493 *Annu. Rev. Ecol. Evol. Syst.* 2007;38(1):459–87.

494 Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, et al.
495 ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Res.*
496 2008;18(5):810–20.

497 Cornejo-Romero A, Vargas-Mendoza CF, Aguilar-Martínez GF, Medina-Sánchez J,
498 Rendón-Aguilar B, Valverde PL, et al. Alternative glacial-interglacial refugia
499 demographic hypotheses tested on *Cephalocereus columna-trajani* (Cactaceae) in the
500 intertropical Mexican drylands. *PLoS One.* 2017;12(4):e0175905.

501 Coyne JA, Orr HA. Speciation. Sinauer Associates, Sunderland, Massachusetts; 2004.

502 Csilléry K, François O, Blum MGB. Abc: An R package for approximate Bayesian
503 computation (ABC). *Methods Ecol. Evol.* 2012;3(3):475–9.

504 Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: More models, new
505 heuristics and parallel computing. *Nat. Methods.* 2012;9(8):772.

506 Deng QX. A natural hybrid of *Chrysolophus Amherstiae* and *C. picus*. *Curr Zool.*
507 1974;20(1):107-06(in Chinese).

508 Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. “Touchdown” PCR to
509 circumvent spurious priming during gene amplification. *Nucleic Acids Res.*
510 1991;19(14):4008.

511 Dong F, Hung CM, Li XL, Gao JY, Zhang Q, Wu F, et al. Ice age unfrozen: severe
512 effect of the last interglacial, not glacial, climate change on East Asian avifauna.
513 *BMC Evol. Biol.* 2017;17(1):1–8.

514 Dong L, Heckel G, Liang W, Zhang Y. Phylogeography of Silver Pheasant (*Lophura*
515 *nycthemera* L.) across China: Aggregate effects of refugia, introgression and riverine
516 barriers. *Mol. Ecol.* 2013;22(12):3376–90.

517 Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with
518 BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 2012;29(8):1969–73.

519 Duchon P, Živković D, Hutter S, Stephan W, Laurent S. Demographic inference
520 reveals African and European admixture in the North American *Drosophila*
521 *melanogaster* population. *Genetics*. 2013;193(1):291–301.

522 Fraser CI, Nikula R, Ruzzante DE, Waters JM. Poleward bound: Biological impacts
523 of Southern Hemisphere glaciation. *Trends Ecol. Evol.* 2012;27(8):462–71.

524 Funk DJ, Nosil P, Etges WJ. Ecological divergence exhibits consistently positive
525 associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci.*
526 *U. S. A.* 2006;103(9):3209–13.

527 He JL, Wei L. Note on some natural hybrids of *Chrysolophus Amherstiae* and *C. picus*
528 from Dayi County, Sichuan. *Zool Res.* 1993;14(3):240–39 (in Chinese).

529 Hewitt G. The genetic legacy of the quaternary ice ages. *Nature*.
530 2000;405(6789):907–13.

531 Hewitt GM. Speciation, hybrid zones and phylogeography - Or seeing genes in space
532 and time. *Mol. Ecol.* 2001;10(3):537–49.

533 Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. *Philos.*
534 *Trans. R. Soc. B Biol. Sci.* 2004;359(1442):183–95.

535 Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on
536 nucleotide data. *Genetics*. 1987;116(1):153–9.

537 Jean-Marie C, Filipe S, Mark AB, Christian PR, Jean-Michel M, David JB, et al.
538 Inferring population history with DIY ABC: a user-friendly approach to approximate
539 Bayesian computation. *Bioinformatics*. 2008;24(23):2713–9.

540 Kirkpatrick M, Servedio MR. The reinforcement of mating preferences on an island.
541 *Genetics*. 1999;151(2):865–84.

542 Lambeck K, Chappell J. Sea level change through the last glacial cycle. *Science*.
543 2001;292(5517):679–86.

544 Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A,
545 et al. Evolution of Darwin’s finches and their beaks revealed by genome sequencing.
546 *Nature*. 2015;518(7539):371–5.

547 Li H. [Heng Li - Compares BWA to other long read aligners like CUSHAW2]
548 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
549 arXiv Prepr. arXiv. 2013;

550 Li H, Durbin R. Inference of human population history from individual whole-
551 genome sequences. *Nature*. 2011;475(7357):493–6.

552 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
553 Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.

554 Li JW, Yeung CKL, Tsai PW, Lin RC, Yeh CF, Yao C Te, et al. Rejecting strictly
555 allopatric speciation on a continental island: Prolonged postdivergence gene flow
556 between Taiwan (*Leucodioptron taewanus*, Passeriformes Timaliidae) and Chinese
557 (*L. canorum canorum*) hwameis. *Mol. Ecol*. 2010;19(3):494–507.

558 Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA
559 polymorphism data. *Bioinformatics*. 2009;25(11):1451–2.

560 Liu Y, Liu S, Yeh CF, Zhang N, Chen G, Que P, et al. The first set of universal
561 nuclear protein-coding loci markers for avian phylogenetic and population genetic
562 studies. *Sci. Rep*. 2018;8(1):1–12.

563 Lyu N, Päckert M, Tietze DT, Sun YH. Uncommon paleodistribution patterns of
564 *Chrysolophus* pheasants in east Asia: Explanations and implications. *J. Avian Biol*.
565 2015;46(5):528–37.

566 Mayr E. Systematics and the Origin of Species from the Viewpoint of a Zoologist.
567 Columbia University Press, New York; 1942.

568 Nosil P. Ecological speciation. Oxford University Press, Oxford, U.K; 2012.

569 Nosil P. Speciation with gene flow could be common. *Mol. Ecol*. 2008;17(9):2103–6.

570 Olsen JB, Crane PA, Flannery BG, Dunmall K, Templin WD, Wenburg JK.
571 Comparative landscape genetic analysis of three Pacific salmon species from
572 subarctic North America. *Conserv. Genet*. 2011;12(1):223–41.

573 Pavlidis P, Laurent S, Stephan W. MsABC: A modification of Hudson’s ms to
574 facilitate multi-locus ABC analysis. *Mol. Ecol. Resour*. 2010;10(4):723–7.

575 Peng MS, Wu F, Murphy RW, Yang XJ, Zhang YP. An ancient record of an avian
576 hybrid and the potential uses of art in ecology and conservation. *Ibis (Lond. 1859)*.
577 2016;158(2):444–5.

578 Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, et al. The genomic
579 landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*.
580 2014;344(6190):1410–4.

581 Provan J, Bennett KD. Phylogeographic insights into cryptic glacial refugia. *Trends*
582 *Ecol. Evol.* 2008;23(10):564–71.

583 Rambaut A, Drummond AJ. Tracer v 1.5. <http://beast.bio.ed.ac.uk/Tracer>. 2009.

584 R Core Team. A language and environment for statistical computing. R Foundation
585 for Statistical Computing, Vienna, Austria; 2020. URL <https://www.R-project.org/>.

586 Rice WR, Salt GW. Speciation via disruptive selection on habitat preference:
587 experimental evidence. *Am. Nat.* 1988;131(6):911–7.

588 Rohling EJ, Fenton M, Jorissen FJ, Bertrand P, Ganssen G, Caulet JP. Magnitudes of
589 sea-level lowstands of the past 500,000 years. *Nature*. 1998;394(6689):162–5.

590 Servedio MR. Reinforcement and the genetics of nonrandom mating. *Evolution* (N.
591 Y). 2000;54(1):21–9.

592 Servedio MR. The what and why of research on reinforcement. *PLoS Biol.*
593 2004;2(12):e420.

594 Servedio MR. Limits to the evolution of assortative mating by female choice under
595 restricted gene flow. *Proc. R. Soc. B Biol. Sci.* 2011;278(1703):179–87.

596 Servedio MR. The relationship between sexual selection and speciation. *Curr. Zool.*
597 2012;58(3):413–5.

598 Servedio MR. Geography, assortative mating, and the effects of sexual selection on
599 speciation with gene flow. *Evol. Appl.* 2016;9(1):91–102.

600 Servedio MR, Bürger R. The counterintuitive role of sexual selection in species
601 maintenance and speciation. *Proc. Natl. Acad. Sci. U. S. A.* 2014;111(22):8113–8.

602 Servedio MR, Kirkpatrick M. The effects of gene flow on reinforcement. *Evolution*
603 (N. Y). 1997;51(6):1764–72.

604 Servedio MR, Noor MAF. The role of reinforcement in speciation: theory and data.
605 *Annu. Rev. Ecol. Evol. Syst.* 2003;34:339–64.

- 606 Shafer ABA, Cullingham CI, Côté SD, Coltman DW. Of glaciers and refugia: A
607 decade of study sheds new light on the phylogeography of northwestern North
608 America. *Mol. Ecol.* 2010;19(21):4589–621.
- 609 Shi XY, Fu Q, Wang L, Jiang ZY, Shi XG, Li S. Note on the Natural Hybridization of
610 Golden Pheasant *Chrysolophus pictus* and Lady Amherst's Pheasant *C. amherstiae* in
611 Anzihe Nature Reserve, Sichuan Province. *Chinese J Zool.* 2018;53(4):660–63(in
612 **Chinese**).
- 613 Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA
614 polymorphism. *Genetics.* 1989;123(3):585–95.
- 615 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular
616 evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 2013;30(12):2725–9.
- 617 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of
618 progressive multiple sequence alignment through sequence weighting, position-
619 specific gap penalties and weight matrix choice. *Nucleic Acids Res.*
620 1994;22(22):4673–80.
- 621 Wang P, Liu Y, Liu Y, Chang Y, Wang N, Zhang Z. The role of niche divergence and
622 geographic arrangement in the speciation of Eared Pheasants (*Crossoptilon*, Hodgson
623 1938). *Mol. Phylogenet. Evol.* 2017;113:1–8.
- 624 Wang P, Yeh C, Chang J, Yao H, Fu Y, Yao C, et al. Multilocus phylogeography and
625 ecological niche modeling suggest speciation with gene flow between the two
626 Bamboo Partridges. *Avian Res.* 2021;12(1):1–10.
- 627 Wang W, Dai C, Alström P, Zhang C, Qu Y, Li SH, et al. Past hybridization between
628 two East Asian long-tailed tits (*Aegithalos bonvaloti* and *A. fuliginosus*). *Front. Zool.*
629 2014;11(1):1–13.
- 630 Wasserman TN, Cushman SA, Schwartz MK, Wallin DO. Spatial scaling and multi-
631 model inference in landscape genetics: *Martes americana* in northern Idaho. *Landsc.*
632 *Ecol.* 2010;25(10):1601–12.
- 633 Wei C, Dong L, Li SH, Alström P, Liu Y, Xia C, et al. From the Himalayas to a
634 continental Island: Integrative species delimitation in the Brownish-flanked Bush
635 Warbler *Horornis fortipes* complex. *Mol. Phylogenet. Evol.* 2019;131:219–27.
- 636 Weir JT, Schluter D. Calibrating the avian molecular clock. *Mol. Ecol.*
637 2008;17(10):2321–8.

- 638 Zhang D, Tang L, Cheng Y, Hao Y, Xiong Y, Song G, et al. “ghost Introgression” As
639 a Cause of Deep Mitochondrial Divergence in a Bird Species Complex. *Mol. Biol.*
640 *Evol.* 2019;36(11):2375–2386.
- 641 Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals
642 insights into avian genome evolution and adaptation. *Science.* 2014;346(6215):1311–
643 20.
- 644 Zhou SZ, Jijun L, Zhang SQ, Zhao JD, Cui JX. Quaternary glaciations in China. *Dev.*
645 *Quat. Sci.* 2004;2(Part C):105–13.

Tables and table captions

647

648

649 Table 1 Population parameters of two datasets under isolation with migration model

650 and secondary contact model estimated by ABC analysis. T_{div} : population divergence

651 time; T_2 : termination time of gene flow after secondary contact; M_1 : the number of

652 migrants from *C. amherstiae* to *C. pictus* in each generation. M_2 : the number of

653 migrants from *C. pictus* to *C. amherstiae* in each generation.

	Allopatric samples			All samples					
	Isolation with migration model			Isolation with migration model			Secondary contact model		
	T_{div}	M_1	M_2	T_{div}	M_1	M_2	T_2	M_1	M_2
2.5% HPD	4.26×10^5	0.184	0.254	4.57×10^5	0.116	0.473	0.031×10^5	0.668	2.14
Median	22.8×10^5	0.929	1.70	19.3×10^5	0.602	1.774	0.130×10^5	2.92	8.97
Mean	22.3×10^5	1.054	2.00	18.6×10^5	0.713	1.054	0.258×10^5	3.22	9.44
97.5% HPD	38.3×10^5	2.590	5.34	28.8×10^5	1.880	1.316	1.206×10^5	7.04	19.33

654

655 Figures and figure captions

656

657 Figure 1

658

659 **Fig. 1** The distribution and sampling locations of *C. pictus* (orange zone and red star)

660 and *C. amherstiae* (blue zone and red circle)

661 Figure 2

662

663 **Fig. 2** Demographic history scenarios tested in this study, which included posterior

664 probabilities for each model and parameters for divergence time (T_{div}), The time when

665 there is gene flow in the early period (T_1), the time when there is gene flow in the late

666 period (T_2), ancestral population size (N_A), *C. pictus* and *C. amherstiae* sizes (N_{cp} and

667 N_{ca} , respectively) and gene flow (M_1 and M_2) when applicable. For (a to d), samples

668 only include allopatric population of *Chrysolophus*. For (e to h), all sampled

669 individuals of *Chrysolophus*, including individuals collected from hybrid zone.

670

671 Figure 3

672

673 **Fig. 3** Inferred historical population sizes by Pairwise Sequential Markovian

674 Coalescent analysis using 19 whole-genome sequenced individuals, including thirteen

675 *C. pictus* and six *C. amherstiae* for re-sequencing data with an average of 35× depth.

676 A *de novo* genome of *C. amherstiae* was used as the reference genome. The numbers

677 of years per generation (g) for both species were set to be 2. The mutation rate per

678 generation (μ) was set to 3.8×10^{-9} per site. (Orange: *C. pictus* Black: *C. amherstiae*)

679