The origin and population history of the endangered golden snub-nosed monkey (Rhinopithecus roxellana) remain largely unavailable and/or controversial. We here integrate analyses of multiple genomic markers, including mitochondrial (mt) genomes, Y-chromosomes, and autosomes of 54 golden monkey individuals from all three geographic populations (SG, QL, and SNJ). Our results reveal contrasting population structures. Mt analyses suggest a division of golden monkeys into five lineages: one in SNJ, two in SG, and two in QL. One of the SG lineages (a mixed SG/QL lineage) is basal to all other lineages. In contrast, autosomal analyses place SNJ as the most basal lineage and identify one QL and three SG lineages. Notably, Y-chromosome analyses bear features similar to mt analyses in placing the SG/QL-mixed lineage as the first diverging lineage and dividing SG into two lineages, while resembling autosomal analyses in identifying one QL lineage. We further find bidirectional gene flow among all three populations at autosomal loci, while asymmetric gene flow is suggested at mt genomes and Y-chromosomes. We propose that different population structures and gene flow scenarios are the result of sex-linked differences in the dispersal pattern of R. roxellana. Moreover, our demographic simulation analyses support an origin hypothesis suggesting that the ancestral R. roxellana population was once widespread and then divided into SNJ and non-SNJ (SG and QL) populations. This differs from previous mt-based “mono-origin (SG is the source population)” and “multiorigin (SG is a fusion of QL and SNJ)” hypotheses. We provide a detailed and refined scenario for the origin and population history of this endangered primate species, which has a broader significance for Chinese biogeography. In addition, this study highlights the importance to investigate multiple genomic markers with different modes of inheritance to trace the complete evolutionary history of a species, especially for those exhibiting differential or mixed patterns of sex dispersal.

**Key words:** golden snub-nosed monkey, population structure, gene flow, dispersal route, demographic simulation.

**Introduction**

The golden snub-nosed monkey (Cercopithecidae, Colobinae, Rhinopithecus roxellana) is an endangered flagship species, which is often referred to as China’s “No. 2 national treasure” (after the giant panda). Golden snub-nosed monkeys have been widely distributed in China throughout the late Pleistocene and early Holocene (Han 1982; Jablonski and Pan 1988; Gu and Jablonski 1989; Gu and Hu 1991; Jablonski 1998a, 1998b). However, dramatic diminution over the last 400 years due to human population expansion, wars, deforestation, and hunting (Li et al. 2002) has restricted this species to only three isolated mountainous regions of central China, that is, Sichuan/Gansu (SG), Qinling (QL), and Shennongjia (SNJ), with no more than 20,000 individuals in the wild (Quan and Xie 2002; Li and Zhao 2007; Luo et al. 2012).

Understanding the genetic diversity, population structure and migration pattern of R. roxellana is critical for implementing efficient management conservation strategies, yet this
information remains largely unavailable and/or controversial. Earlier studies using sequences of the mitochondrial (mt) D-loop region found differentiation among the three populations (SG, QL, and SNJ populations) and within the SG population (clades SG-A and SG-B) (Li et al. 2007; Pan et al. 2009). Based on the observations that QL and SNJ have limited gene flow, and that individuals from the two populations have a much closer genetic connection with those from the SG population, Pan et al. (2009) proposed two different hypotheses to explain the origin of the modern populations of golden monkeys: 1) a “mono-origin” hypothesis suggesting that the SG population was the source population (fig. 1A) and 2) a “multiorigin” hypothesis suggesting that the SG population was the result of migrations from both QL and SNJ populations (fig. 1B). Further analyses of the entire D-loop region including more individuals by Luo et al. (2012) found additional population subdivision within QL (clades QL-A and QL-B) and suggested significant gene flow from SG to QL and QL to SNJ, thus supporting the mono-origin hypothesis (fig. 1A).

All these findings, however, are based on mtDNA and it is unclear whether they are affected by sex-biased dispersal. This is particularly relevant for *R. roxellana* because this species has a prominent and unusual social system exhibiting a multilevel sociality and multiple dispersal patterns. Based on behavioral and genetic data, male-biased (Ren et al. 2000; Zhao et al. 2008; Yao et al. 2011; Chang et al. 2014; Huang et al. 2017), female-biased (Guo et al. 2007; Zhang et al. 2008), and bisexual dispersal (Qi et al. 2009; Yan 2012) have been reported.

The results of a recent nuclear genome analysis of 27 golden monkey individuals contradicted the population history as inferred from mtDNA data (Zhou et al. 2016). Instead, the study suggested differentiation only between the SNJ population and the other two populations (SG and QL). This contradiction strengthens the view that integrated analyses of multiple markers with different modes of inheritance (i.e., patrilineal, matrilineal and biparental markers) are crucial to test the effect of sex-biased dispersal and to obtain an unbiased and complete understanding of the origin and population history of golden monkeys. Moreover, no analysis of the patrilineal history of snub-nosed monkeys has been undertaken so far.

In this study, we analyzed three genomic markers including mt genomes, Y-chromosomes, and autosomal genomes of 54 golden monkey individuals from all three geographic populations (SG, QL, and SNJ). Based on the integrated analyses of maternally, paternally, and biparen tally inherited markers, we provide a detailed and refined scenario for the origin and population history of the endangered golden snub-nosed monkey.

**Results**

**Sampling, Sequencing, and Single Nucleotide Polymorphism Calling**

In total, 54 *R. roxellana* individuals from all three geographic populations (25 individuals from SG, 16 from QL, and 13 from SNJ; fig 1; supplementary table S1, Supplementary Material online) were used in this study. All sequences are at a depth of 10-fold except for two individuals (QLS03 and SNJ01) with 24-fold coverage. Single nucleotide polymorphism (SNP) calling and stringent quality-filtering retained 9.36 million high-quality SNPs and a total of 8.93 million autosomal SNPs for the analyses. We estimated the number of pairwise nucleotide mismatches using in-house Perl scripts and pairwise genetic distance using MEGA 7 for all 54 individuals. The number of pairwise nucleotide mismatches is from 250,792 to 3,282,599 (2,579,642 on average), and the pairwise genetic distance is 0.1–14.1% (9% on average), confirming that all samples derived from different individuals.

**Nucleotide Diversity, Population Differentiation, and Linkage Disequilibrium**

Nucleotide diversity (*π*) and population differentiation (*Fst*) of *R. roxellana* populations from three geographic distributions (SG, QL, and SNJ populations) in the context of **mt**
genome, Y-chromosome, and nuclear autosomes are shown in figure 2A and 2B. We find that the SNJ population has the lowest nucleotide diversity compared with SG and QL populations (fig. 2A). In addition, we find a slower linkage disequilibrium (LD) decay in SNJ than in SG and QL (fig. 2C). These findings are consistent with the fact that SNJ is the smallest population of *R. roxellana* with an extremely low effective population size (<100; Chang et al. 2012). Intriguingly, mt genome estimate of π is notably lower in SNJ compared with SG and QL, while autosomal estimate of π is similar in all three populations (fig. 2A). We speculate that this is due to a very small SNJ founder population, which was later introgressed by males from the other populations. In addition, our analyses indicate that mean nucleotide diversity of the Y-chromosomes is much lower than of mt genomes and autosomes (π = 0.3 × 10^{-4} for Y-chromosome vs. 5.22 × 10^{-3} for mt and 1.109 × 10^{-3} for autosomes).

Fst analyses from autosomes, mt genomes, and Y-chromosomes suggest the highest differentiation between SNJ and QL and the lowest between SNJ and SG, are twice as high as between SG and QL. These results suggest that among all three populations SNJ is most genetically differentiated. In addition, the Fst values between populations for autosomal SNPs are all lower than for those corresponding to mt genomes and Y-chromosomes, with the Y-chromosomes exhibiting the highest (fig. 2B).

**Discordance of Uniparental and Biparental Phylogenetic Trees**

The phylogenetic relationships of *R. roxellana* individuals based on mt genomes, Y-chromosomes, and autosomes are shown in figure 3. In the mt neighbor-joining (NJ) tree (fig. 3A), *R. roxellana* is divided into five lineages (SG/QL-mt-mixed, SG-mt, QL-mt-1, QL-mt-2, and SNJ-mt) that cluster into two major clades (Clade-mt-1 and Clade-mt-2). Clade-mt-1 corresponds to the SG/QL-mt-mixed lineage, consisting of four individuals from SG (MS05, QLA03, GS03, and MS12) and two individuals from QL (QLN02 and QLN03).
Fig. 3. Phylogenetic analyses of mt genomes from all individuals based on NJ method (A) and ML method (B). The phylogenetic analyses of Y-chromosome data from male individuals based on NJ/ML methods (C) and of mt genomes from the same set of male individuals based on NJ and ML methods (D). Different relationship between NJ and ML methods was indicated as a dotted line. Phylogenetic analyses of autosomal SNPs from all individuals based on NJ, ML, and ASTRAL methods (E). Node support is presented above nodes. Results from admixture analyses (F) and PCAs (G) for autosomal SNPs.
The other four lineages form Clade-mt-2. In Clade-mt-2, QL-mt-1 (11 individuals from QL) and SNJ-mt (13 individuals from SNJ) group together. QL-mt-2 (three individuals from QL) and SG-mt lineages (21 individuals from SG) are closely related. The monophyly of these five lineages and their phylogenetic relationships are strongly supported with 100% bootstrap support (BS), except for the sister-group relationships between SNJ-mt and QL-mt-1 (56% BS). Maximum likelihood (ML) analyses yield a slightly different tree topology with QL-mt-1 and QL-mt-2/SG-mt grouping together and SNJ-mt representing the most basal lineage within Clade-mt-2 (fig. 3B). However, the statistical support for this relationship is low (62% BS). Thus, SNJ, QL-mt-1, and QL-mt-2/SG-mt lineages form an unresolved trichotomy in mt genome analyses.

All phylogenetic analyses (NJ and ML) based on Y-chromosome sequences from 27 R. roxellana males yield the same tree topology (fig 3C). For comparison, we also reconstructed mt genome trees from the same 27 males (fig. 3D). The Y-chromosome gene tree divides individuals into four lineages (SG/QL-y-mixed, SG-y, QL-y, and SNJ-y) in two major clades (Clade-y-1 and Clade-y-2). Clade-y-1 corresponds to the SG/QL-y-mixed lineage, including individuals from the SG/QL-mt-mixed (i.e., Clade-mt-1) and another two from SG (QLA02 and MS17). These two individuals cluster in the SG-mt lineage in the mt genome tree (fig. 3D). The other three lineages form Clade-y-2. In Clade-y-2, SNJ-y and SG-y are more closely related to each other than they are to the QL-y lineage. BS values for the monophyly of the four lineages are 92–100% (NJ) and 98–100% (ML), while the branching pattern among the four lineages is supported by 82–100% (NJ) and 88–100% (ML).

All analyses based on autosomal SNPs retrieved identical tree topologies (fig. 3E). The autosomal tree divides R. roxellana individuals into five lineages (SNJ-snp, QL-snp, SG/QL-snp-mixed, SG-snp-1, and SG-snp-2) in two clades (SNJ-snp and non-SNJ-snp). The SNJ individuals form their own population-specific clade (SNJ-snp), and the other four lineages form the non-SNJ-snp clade. This pattern is consistent with the Fst results illustrating the deep genetic divergence and differentiation of SNJ from SG and QL (fig. 2B). With the exception of the three SG individuals in the SG/QL-snp-mixed lineage (MS05, QLA02, and QLA03), the other SG individuals segregate into two lineages. One lineage contains the majority of SG individuals (SG-snp-1), while the other (SG-snp-2) comprises eight SG individuals (MS06, MS07, MS09, MS10, MS15, MS16, QLA01, and QLA05) and forms a sister lineage to QL-snp. The supports for the monophyly of the four lineages and the relationships among them are 100% in all analyses.

Our phylogenetic analyses revealed obvious discordances among trees derived from mt genome, Y-chromosome, and autosome data. Our mt genome analyses (fig. 3A, B, and D) indicate the presence of two major lineages within both SG and QL, and one in the SG lineage (SG/QL-mixed) diverged earliest in the tree. In sharp contrast, in the autosomal tree (fig. 3E) individuals from the two major QL lineages identified in the mt genome analyses cluster into a monophyletic QL clade without further subdivisions, while individuals of the SG population are divided into three distinct lineages. Furthermore, the autosomal tree places SNJ as the most basal lineage, rather than SG/QL-mixed as indicated in the mt genome analyses. The results of the Y-chromosome analyses (fig. 3C) bear features similar to mt genome analyses in placing SG/QL-mixed lineage as the first diverging lineage and dividing the SG population into two major lineages, while resembling the autosomal results in merging most individuals from QL into a major QL clade. Besides at the population level, incongruence among markers was also found at the individual level (fig. 3). For example, QLA02, an individual from SG, clusters with other SG individuals in the SG-mt lineage in the mt genome tree, while it clusters within the SG/QL-mixed lineages in the Y-chromosomal and autosomal gene trees; GS03 and MS12, two individuals from SG, are nested within the SG/QL-mixed lineages on the mt and Y-chromosomal gene trees, but in the autosomal tree they cluster with most other SG individuals in the SG-snp-1 lineage.

Population Structure and Admixture Analyses

The results of the admixture analysis and principal component analysis (PCA) of autosomal SNP data are shown in figure 3F and G, which provide additional corroborative evidence for the phylogenetic results (fig. 3E). In the admixture analysis (fig. 3F), K = 2 has the lowest cross-validation error. When partitioning individuals into two groups (K = 2), SNJ individuals are separated from all other individuals, reflecting a clear divergence of SNJ among R. roxellana populations. At K = 3, the non-SNJ individuals are divided into 50% QL and 50% SG populations. For some individuals, including all eight individuals from the SG-snp-2 lineage and one individual from QL-snp lineage (QL09), admixture between both populations is indicated. In addition, one individual from SNJ-snp-2 (MS16) and two individuals from QL-snp (QL07 and QLS04) are observed in individuals admixed with little SNJ population component as well (fig. 3F). Further partitioning into four groups (K = 4) separates SG into two clusters, corresponding to SG-snp-1 and SG-snp-2. In addition, the mixture of some SG-snp-1 individuals with SG-snp-2 component (e.g., MS04, MS05, and QLA03) and the mixture of some QL individuals with SG and/or little SNJ component (e.g., QL07, QL09 and QLS04) are observed.

In the PCA (fig. 3G), the first component (PC1) separates SNJ from non-SNJ (QL and SG), while the second component (PC2) fails to separate QL and SG. The SG population exhibits two substructures (corresponding to SG-snp-1 and SG-snp-2 lineages), and the SG-snp-2 lineage is closer related to the QL-snp lineage than to the SG-snp-1 lineage. Two QL and three SG individuals form the SG/QL-snp-mix lineage. Altogether these findings suggest genetic admixture between QL and SG populations.

Intriguingly, the two QL individuals (QLN02 and QLN03) that are consistently intermixed with several SG individuals observed in all phylogenetic analysis (fig. 3A–E) and PCA (fig. 3G) based on the three markers analyzed (within SG/QL-mt-mixed, SG/QL-y-mixed, and SG/QL-snp-mixed lineages) show the whole SG genetic components in all the
admixtures from $K = 2$ to $4$ (fig. 3F). Moreover, f3-statistics found no sign of admixture from SG and QL in these two individuals (Z score > 5). We speculate that these two individuals migrated from SG to the northern QL Mountain very recently and that the migration time is too recent to mix with resident individuals. As previous reports suggested complete geographical isolation between these populations since 400 years (Li et al. 2002), the recent migration event might be due to human translocation.

Incomplete Lineage Sorting Test and Migration Events of mt and Y Loci

Conflicting gene trees resulting from different genomic data can result from either incomplete lineage sorting (ILS) and/or gene flow. First, we assessed whether the phylogenetic conflicting signal could be explained by ILS. Using a strategy similar to that used by Sequeira et al. (2011), we simulated 10,000 mtDNA and Y-chromosome trees separately using a model with no gene flow and the population history and parameter estimates from the preferred demographic model based on the simulation analyses of autosomal data. We tested the possibility of recovering the mt genome and Y-chromosome phylogenies under the speciation scenario inferred from the autosomal genome data. If the same population structuring and relationships as shown in the mt genome analyses and Y-chromosome analyses were recovered in <5% of the simulated trees, ILS was rejected. The coalescent simulation results show that no mt and Y phylogenies (the tree topologies observed with empirical mt and Y-chromosome data sets) were inferred in the simulated trees, indicating that the discrepancy between the genetic markers is less likely to be caused by ILS.

We estimated gene flow among populations for mt genome and Y-chromosome data sets. The mt genome results suggest a directional gene flow from SG to QL, and bidirectional gene flow between QL and SNJ, but no gene flow between SG and SNJ. The Y-chromosome results suggest directional gene flow from SG to QL and from QL to SNJ, and no gene flow between SG and SNJ (fig. 4B and C and supplementary table S2, Supplementary Material online).

Demographic History and Migration Events of Autosomal Genomes

According to the results above, we simulated the demographic models to investigate the demographic history of R. roxellana based on autosomal DNA polymorphism data. First, we established six demographic models for SNJ and non-SNJ clades covering a large range of demographic possibilities during the Pleistocene (supplementary fig. S1A–F, Supplementary Material online). These six demographic models were specified as representative scenarios based on past climate and geographical environment with which R. roxellana used to live (Ren and Beug 2002; Sun and An 2005; Zhan et al. 2011), and previous studies of the demographic history of R. roxellana (Li et al. 2003; Zhou et al. 2014, 2016; Yu et al. 2016). The Akaike information criterion (AIC) results identify the preferred model for SNJ with two sequential bottlenecks accompanied by a recent expansion (supplementary fig. S1E, Supplementary Material online) and for non-SNJ a model with one bottleneck accompanied by a recent reduction (supplementary fig. S1D, Supplementary Material online).

Based on the preferred models for SNJ and non-SNJ clades, we then explored the demographic relationships between the two clades by comparing three demographic scenarios (supplementary fig. S2, Supplementary Material online). Models A and B are founder-migration models, which treat SNJ (supplementary fig. S2A, Supplementary Material online) and non-SNJ (supplementary fig. S2B, Supplementary Material online) as the ancestral population, respectively, while Model C is an isolation-with-migration (IM) model, which assumes that ancestral R. roxellana was once widespread and then split into SNJ and non-SNJ (supplementary fig. S2C, Supplementary Material online). The AIC results manifest that Model C is the preferred demographic scenario. We also compared Model C with two alternative IM models taking different population sizes, migration rates, and divergence times into consideration (supplementary figs. S3B and S4D, Supplementary Material online), and with the isolation with no migration model (supplementary fig. S3A, Supplementary Material online) as well as with isolation-after-migration model (supplementary fig. S3C, Supplementary Material online). The results still prefer Model C as the best model for expounding the origin of R. roxellana.

To further clarify the demographic relationships between SG and QL within non-SNJ, we compared three demographic scenarios (supplementary fig. S4, Supplementary Material online), while fixing the demographic (bottleneck) scenario of an ancestral population based on the result of supplementary figure S2C, Supplementary Material online. Models A and B represent scenarios which assume SG (supplementary fig. S4A, Supplementary Material online) and QL (supplementary fig. S4B, Supplementary Material online) as respective ancestral populations. Model C represents an IM scenario (supplementary fig. S4C, Supplementary Material online). According to AIC results, Model A is the most likely, indicating that QL derived from SG with a small founder population size. To explore the migration period from SG to QL and the subsequent population size of QL, four migration demographic models were compared (Models A–D in supplementary fig. S5, Supplementary Material online). Model D was preferred, suggesting that QL derived from SG with a small founder size after SG experienced a recent population size reduction.

The above preferred demographic scenarios (Model C in supplementary fig. S2, Supplementary Material online, and Model D in supplementary fig. S5, Supplementary Material online) were combined to form the final optimal model for the demographic history of R. roxellana (fig. 5). The point estimations of all parameters, including divergence time, effective population size, and migration rate, were inferred from 1,300 independent runs (fig. 5). The final demographic scenario indicates that the ancestral R. roxellana population was once widespread and then experienced a bottleneck between 205.9 and 177.9 thousand years ago (ka). Around 24.5 ka, the widespread ancestral population, according to the fossil record (Han 1982; Jablonski and Pan 1988; Gu and Jablonski 1989; Gu and Hu 1991; Pan 1995; Jablonski 1998a, 1998b)
most likely occurring in south and central China, was split into two populations, that is, SNJ and non-SNJ. For non-SNJ, a population size reduction occurred 18.9 ka and a small founder population further migrated from SG to QL around 13.5 ka. The QL population subsequently experienced an expansion around 9.3 ka, while the SNJ population was faced with a severe reduction during the Holocene around 1.3 ka.

The inferred migration rates shown in the final demographic model (fig. 4A and supplementary table S2, Supplementary Material online) indicate continuous gene flow among the three populations after their initial divergence. Further investigations demonstrate that the migration rates between SG and QL are larger than those between SNJ and SG and between SNJ and QL, which is consistent with the results of phylogenetic (fig. 3E) and admixture analyses (fig. 3F). This supports the differentiation of SNJ from the other two populations and genetic admixture between SG and QL. In addition, the migration rate from SG to QL (1.079 individual per generation) was found to be nearly eight times higher than from QL to SG (0.169 individual per generation), which is consistent with the inference from the preferred demographic model that QL originated from SG. We further validated our results by site frequency spectrum (SFS) analyses (supplementary fig. S7, Supplementary Material online) and the genetic diversity comparison between simulation and observed data (supplementary table S3, Supplementary Material online).

Discussion

By investigating genome-wide data from both uniparental and biparental loci from a total of 54 individuals, we provide a detailed and refined scenario of the demographic and biogeographic history of the golden snub-nosed monkey. Phylogenetic analyses of mt genomes, Y-chromosome, and autosomal genomes showed contrasting population relationships and structuring, which can be explained by the effects of ILS and/or gene flow. Our coalescent simulations showed that the observed gene tree discordance between genetic markers
is less likely to be caused by ILS. We interpret the significant differences as a result of secondary gene flow, which occurred more extensively than previously recognized by a single genetic marker (mt D-loop region; Pan et al. 2009; Luo et al. 2012). Indeed, the inferred migration rates from our demographic model simulations of autosomal data (fig. 4A and supplementary table S2, Supplementary Material online) indicated continuous gene flow among all three populations after their initial separation. Extensive gene flow can also be reflected in the admixture analyses (fig. 3F). Some individuals are observed to have genetic components from all three populations, and notably, the SG-snp-2 lineage is the product of admixture between QL and the SG-snp-1 lineage.

Gene flow among all three populations was detected at autosomal loci, while asymmetric gene flow was found at mt and Y-chromosome loci (fig. 4B and C and supplementary table S2, Supplementary Material online). The lower level of gene flow for mt genomes and Y-chromosomes compared with autosomes is also reflected by the greater differentiation of mt genomes and Y-chromosome compared with autosomes (fig. 2B). The multilevel social system and multiple dispersal patterns in R. roxellana could be an explanation for inconsistent differentiation patterns and heterogeneous signal for the observed gene flow. According to previous studies on golden monkeys, one-male units (consist of an adult male, several adult females, and offspring) form the basic social unit, with many one-male units forming a band, and a number of bands, including all-male units (consist of three to five adult and subadult males) forming a troop (Kirkpatrick 1995; Ren et al. 2000; Qi et al. 2009). This multilevel sociality results in multiple dispersals (Huang et al. 2017), with three different kinds of dispersal patterns reported in R. roxellana, that is, male-biased (Ren et al. 2000; Zhao et al. 2008; Yao et al. 2011; Chang et al. 2014; Huang et al. 2017), female-biased (Guo et al. 2007; Zhang et al. 2008), and bisexual dispersal (Qi et al. 2009; Yan 2012). Therefore, different gene flow scenarios among the three genomic markers appear at least partly to be the result of sex-linked differences in the dispersal patterns of R. roxellana. As can be seen in figure 4 and supplementary table S2, Supplementary Material online, gene flow from SG to QL and from QL to SNJ are consistently significant in all three genomic markers; however, gene flow from SNJ to QL is identified in mt genome and autosomes but not in Y-chromosomes. Other signals of gene flow between populations (e.g., QL to SG, SG to SNJ, and SNJ to SG) are detected in autosomes but not in mt genomes and Y-chromosomes. The present observation of gene flow signals in all genetic markers from QL to SNJ is consistent with previous findings of male and female dispersals in the QL population based on mt D-loop analyses (Yan 2012) and field observations (Qi et al. 2009), supporting bisexual dispersal. In comparison, for SNJ, male dispersal has been reported from microsatellite analyses (Chang et al. 2014) and long-term field studies (Ren et al. 2000; Yao et al. 2011). However, we found mt and autosomal gene flow from SNJ to QL, and did not find any Y-chromosomal gene flow from SNJ to other populations. We explain these discrepancies by the fact that earlier studies investigated dispersal patterns in bands and troops within a population (SNJ), while we investigated dispersal patterns from the view of among three populations with a larger geographic range. Moreover, the 1:1.57 sex ratio of male/female reported in SNJ (Chang et al. 2014) may lead to relaxed intrasexual, that is, male–male competition for mates and the reduced possibility of male dispersal. For SG, little information about sex-biased dispersal is available, and we hypothesize the existence of bisexual dispersal in SG based on our results. Therefore, our approach based on the integrated analyses of uniparentally and biparentally inherited genetic markers not only reveals sex-linked differences in the dispersal patterns of R. roxellana but also provides evidence for bisexual dispersal in SG for the first time from a genetic perspective.

Our study shows that genomics provides a useful tool to understand sex-biased dispersal of species. This is an important basis to extend prediction-based genetic studies to empirically grounded field studies. In addition, our study highlights the importance to investigate patrilineally, matrilineally, and biparentally inherited markers to trace the complete evolutionary history of a species. This approach can be particularly insightful to study nonhuman primates, which exhibit differential (male dispersal or female dispersal) or mixed patterns of sex dispersal (bisexual dispersal) (Eriksson et al. 2006; Douadi et al. 2007; Nater et al. 2011; Schubert et al. 2011; Nietlisbach et al. 2012; Inoue et al. 2013; Langergraber et al. 2014). By analyzing mtDNA, Y-chromosomal, and autosomal data, a detailed and refined migration pattern was obtained which would have not been possible if only one or two of these markers have been investigated.

Our results provide a detailed and refined scenario into the origin and dispersal of R. roxellana. Previous studies based on mt D-loop suggested two hypotheses concerning the origin and dispersal of R. roxellana (Li et al. 2007; Pan et al. 2009; Luo et al. 2012). According to the mono-origin hypothesis, SG was the ancestral population, followed either by a stepwise dispersal with a first colonization event from SG to QL and then from QL to SNJ (Luo et al. 2012) or that QL and SNJ originated separately from SG (Pan et al. 2009) (fig. 1). In contrast, the multiorigin hypothesis suggests that SG derived from QL and SNJ (Pan et al. 2009). Our mt genome and Y-chromosome analyses (fig. 4B and C) imply dispersal routes from SG to other populations, supporting the mono-origin hypothesis, whereas the autosomal analyses revealed a different pattern. Our demographic simulations and analyses of migration routes support the presence of a once-widespread ancestral R. roxellana population which then further split into populations. We found no evidence for the multiorigin hypothesis based on any of our analyses. Moreover, if SG is the product of a fusion between QL and SNJ, as suggested by the multiorigin hypothesis, SG should show lower genetic diversity and fewer private variations as well as higher LD than QL and SNJ, which is not the case (fig. 2A and C).

We further compared different hypotheses using demographic simulations. Our results support the origin model suggesting that the ancestral R. roxellana population was once widespread and then split into SNJ and non-SNJ populations as demographic scenario preferred over the alternative models of mono-origin and multiorigin hypotheses.
(supplementary fig. S6, Supplementary Material online). Indeed, based on the fossil record, *R. roxellana* had a wide distribution in China during the Pleistocene and early Holocene (Han 1982; Jablonski and Pan 1988; Gu and Jablonski 1989; Gu and Hu 1991; Pan 1995; Jablonski 1998a, 1998b), despite that *R. roxellana* populations now survived only in three isolated mountainous regions. This implies that the ancestral *R. roxellana* population was once widespread and became then fragmented with ongoing gene flow, which corroborates the preferred hypotheses. The demographic scenario inferred from our model simulations showed that the ancestral bottleneck (177.9–205.9 ka) occurred during the Penultimate Glaciation. The split into SNJ and non-SNJ (24.5 ka), the non-SNJ population reduction (18.9 ka), and the split between SG and QL within non-SNJ (13.5 ka) all took place around the beginning of the Last Glacial Maximum. These observations indicate that extreme cold climate could have caused several rounds of severe population reductions. The subsequent Holocene brought warmer climate which may have promoted migration and expansion of QL (9.31 ka). The very recent severe reduction of SNJ (1.3 ka) may be related with human activities. Overall, our study provides insights into the processes that shaped species distribution across China and has a broader significance for Chinese biogeography. The climate changes resulted from repeated glacial and interglacial period occurred in Pleistocene and Holocene (e.g., the Penultimate Glaciation and the Last Glacial Maximum) as well as the anthropogenic activities (habitat loss due to human commercial activities including logging and hunting; Li et al. 2002; Wang et al. 2014) have shaped the population demography of the golden snub-nosed monkey. A similar pattern has been found for the endangered giant panda and red panda, which geographically largely overlaps with *R. roxellana* (Zhu et al. 2010; Hu et al. 2011; Zhao et al. 2013). Therefore, the herein inferred demographic and biogeographic scenario can be used as a hypothesis to be tested in other species in the same region as they most likely have experienced the same changes in climate and geology.

### Materials and Methods

#### Sampling and Sequencing

We analyzed genome sequences of 54 *R. roxellana* individuals from all three geographic distributions (SG, QL, and SNJ). These include 26 resequencing genomes obtained from Zhou et al. (2016), 23 from our recently published study (Yu et al. 2016), the short-read library reads from 1 de novo *R. roxellana* genome (Rrox_v1; Zhou et al. 2014), and 4 newly resequenced individuals in this study (supplementary table S1, Supplementary Material online). Overall, the present study includes 25 individuals from SG, 16 from QL, and 13 from SNJ. In addition, the published de novo *R. bieti* genome (Rb0; Yu et al. 2016) was used as outgroup for rooting (supplementary table S1, Supplementary Material online).

Genomic DNA of *R. roxellana* individuals collected in this study was extracted from tissue samples using the QIAamp DNA Blood Mini Kit (Qiagen). A 500-bp paired-end insert library was constructed according to the Illumina protocol. All libraries were sequenced on an HiSeq 2000 platform. The newly sequenced short read data have been deposited in the NCBI Short Read Archive under the accession number SRP137287.

#### SNP Calling and Filtering

We used Burrows–Wheeler Aligner (BWA) version 0.6.2 (Li and Durbin 2009) to align the clean paired-end reads from each of the *R. roxellana* individuals and the *R. bieti* outgroup to the de novo assembled *R. roxellana* reference genome (Rrox_v1). BAM alignment files were generated using SAMtools (Li et al. 2009). Polymerase chain reaction duplicates were removed using PICARD (http://picardsourceforge.net). Indel (inserts and deletions) realignment was performed using the IndelRealigner algorithm implemented in the Genome Analysis Toolkit (McKenna et al. 2010).

We identify SNPs using SAMtools mpileup (Li et al. 2009). The raw SNPs were filtered to obtain high-quality variants using a custom script using the following criteria: 1) SNPs located within 5 bp of an InDel, 2) SNPs with >20% missing data, and 3) SNPs with a phred-scaled quality score <30. To obtain autosomal SNPs for subsequent analyses, we excluded the SNPs mapping to the *R. roxellana* scaffolds that show more than 70% of sequence similarity with the rhesus macaque sex chromosomes (160.9 Mb in length; rheMac3; ftp://hgdownload.cse.ucsc.edu/ goldenPath/rheMac3bigZips/rheMac3.fa.gz).

#### Identification of Mt Genomes and Y-Chromosome Sequence Data

Besides the autosomal SNPs, the mt genomes and Y-chromosome sequence data were also identified in order to test the effect of sex-biased dispersal and to get an unbiased and complete understanding of the origin and population history of golden monkeys. The mt genomes of all individuals were determined according to the consensus sequences between the mapped short reads and a published 16,549-bp-long mt genome sequence of *R. roxellana* (Sterner et al. 2006) using SAMtools mpileup (Li et al. 2009). We then filtered the consensus sequences using VCFtools v0.1.12a (Danecek et al. 2011) and an in-house Perl scripts (see supplementary notes, Supplementary Material online). The consensus sequences of all individuals were then aligned using Muscle software (Edgar 2004), and the ambiguous sites were removed using Gblocks v0.91b (gapall) (Castresana 2000; Talavera and Castresana 2007).

The Y-chromosome sequences of all male individuals were determined according to the consensus sequences between the mapped short reads and the annotated ~0.49-Mb-long Y-linked genes (SRY, ZFY, UTY, USPY9, DDX3Y, and RPS4Y1) of the published *R. roxellana* reference genome (Zhou et al. 2014) using the same SAMtools parameters and filtering criteria (see supplementary notes, Supplementary Material online). For individuals without sex information, we compared genomic coverage on the X-chromosome and Y-chromosome, respectively, with the genomic coverage on autosomal chromosomes to identify their sex (see supplementary notes,
**Supplementary Material** online). The reliability of the results was also confirmed by applying it to other individuals with known sex information, which obtained consistent results. Finally, 27 male and 27 female individuals of *R. roxellana* were identified (see supplementary table S1, Supplementary Material online).

**Nucleotide Diversity, Population Differentiation, and LD**

The nucleotide diversity (π) within a population and the population differentiation index (Fst) between populations were calculated based on autosomal SNPs with 100-kb sliding windows and with 10-kb step size using VCFtools v0.1.11 (Daneczek et al. 2011). For mt genomes and Y-chromosomes, π and Fst were calculated using DNASP 5.0 software (Librado and Rozas 2009). LD was calculated using Haploview (Barrett et al. 2005). We calculated the r-square statistic (r²), which is the correlation coefficient between two focal loci of interest (see supplementary notes, Supplementary Material online).

**Phylogenetic Tree Reconstruction**

We used *R. bieti* as the outgroup for all phylogenetic analyses. For the autosomal SNP data set, we constructed an NJ tree using MEGA7.0 (Kumar et al. 2016) with the p-distance model and an ML tree using FastTree (Price et al. 2010) with the GTRGAMMA model. Due to large number of autosomal SNPs, the NJ tree using MEGA can only be reconstructed with the p-distance model, whereas the tree reconstruction with other models exceeds the computational limitation of MEGA. In total, 1,000 bootstrap replications were performed. In addition, a coalescent-based species tree was constructed using ASTRAL-II (Mirarab and Warnow 2015). Individual gene trees were obtained using RAxML version 8 (Stamatakis 2014) based on different nonoverlapping window sizes (100, 200, and 300 kb) under the GTRGAMMA model with 500 bootstrap replicates. We restricted this analysis on scaffold with a length >100 kb and to windows that contained more than 200 SNPs.

For both mt genomes and Y-chromosome sequences, trees were constructed using MEGA for NJ analyses and RAxML for ML analyses. For NJ analyses, the Kimura’s two-parameter model was used. For ML analyses, the best-fit substitution models were selected using AIC (Akaike 1974; Posada and Buckley 2004) with jModeltest 2.1.1 (Darriba et al. 2012). The GTRGAMMA model was selected as the best-fit model for both mt genomes and Y-chromosome sequences. In total, 1,000 bootstrap replicates were performed for both analyses.

**Population Structure and Admixture Analyses**

The autosomal SNPs were thinned by LD values using PLINK v1.01, resulting in a set of ~1.01-Mb SNPs for population structure and admixture analyses. A PCA was carried out using the smartPCA program from the Eigensoft package (Patterson et al. 2006) without outlier iterations. Eigenvectors were generated with the R function region. The Tracy–Widom test was used to determine the significance level of the eigenvectors. The admixture analysis was done with ADMIXTURE1.23 (Alexander et al. 2009) with default settings. We performed independent ADMIXTURE runs from K = 1 to K = 7 with 1,000 bootstrap replicates. The optimal K value was determined using the lowest cross-validation values of the number assumptions. Patterns of recent mixture of the population were carried out using f3 test (Reich et al. 2009) using the ADMIXTOOLS package (Patterson et al. 2012).

**ILS Test**

The ILS effect was assessed by testing the possibility of recovering the phylogenetic pattern inferred from mt and Y-chromosome data under the speciation scenario inferred from autosomal data. In the absence of estimates for *R. roxellana* Y-chromosome mutation rate, we used human Y-chromosome rate of 7.6 × 10⁻¹⁰ mutations per base per year (Jobling and Tyler-Smith 2017). For *R. roxellana* mt mutation rate, we used BEAST 2.1.3 software (Bouckaert et al. 2014) based on three data sets, that is, whole mt genome data set (Data set 1), protein-coding data set (Data set 2), and non-protein-coding data set (Data set 3), given that different mt genome regions have different mutation rates. Nine published complete mt sequences from *Rhinopithecus* species (Sterner et al. 2006; Yu et al. 2011; Liedigk et al. 2012; Hong et al. 2017) and two outgroups from *Pygathrix nemaeus* (Roos et al. 2011) and *Semnopithecus entellus* (Sterner et al. 2006) were used. The divergence time between *Rhinopithecus* and *P. nemaeus* of 6.9 (6.25–7.55) Ma (Sterner et al. 2006) and the origin time for genus *Rhinopithecus* of 2.5 (1–4) Ma (Delsom 1994) were applied as the calibration points. We used a GTR+G nucleotide substitution model under a lognormal relaxed clock. As a result, the mt mutation rates for Data set 1, Data set 2, and Data set 3 were 6.0 ± 0.27 × 10⁻⁶, 2.12 ± 0.266 × 10⁻⁸, and 3.064 ± 0.376 × 10⁻⁸, respectively.

Testing for ILS was performed with 10,000 simulated trees of mt and Y-chromosome using ms, a program for generating samples under a variety of neutral models (Hudson 2002). We tested under a model with no gene flow and the population history and parameter estimates from the preferred demographic model based on the simulation analyses of autosomal data (see supplementary notes, Supplementary Material online). Three independent runs were performed to check whether the same population structuring and relationships as shown in mt genome analyses and Y-chromosome analyses were recovered in the simulated trees. If they were recovered in <5% of the simulated trees, ILS was rejected.

**Mt Genomes and Y-Chromosome Gene Flow Analyses**

We used MIGRATE-N 3.6 (Beerli and Felsenstein 2001; Beerli 2006) to estimate gene flow among populations based on mt genome and Y-chromosome. Transition/transversion ratio was set to 10.62 for mt genome and 2.02 for Y-chromosome as calculated with MEGA. Three long chains with 50,000 recode steps and ten short chains with 5,000 recode steps at a sampling increment of 200 were set, and burn in was 10,000 for each chain. We used a default static heating
scheme with four chains, and swapping interval was 1. Ten independent replicates were run.

Demographic History Modeling
To reveal the demographic history of *R. roxellana*, we used the maximum composite likelihood approach based on the joint SFS implemented in fastsimcoal2 (*Excoffier* et al. 2013) to assess the fit of various demographic models and to infer the final optimal demographic scenario for golden monkeys.

First, the joint SFS of different *R. roxellana* populations was inferred by assuming the allele state of *R. bieti* as the ancestral allele state and removing the heterozygote allele sites in the *R. bieti* genome. Considering the calculation capability of fastsimcoal2, the whole joint SFS was scaled down to one-tenth in the subsequent demographic analysis. A total of 13,152,183 SNPs were used for this analysis. The mutation rate per generation was set as $1.36 \times 10^{-9}$, and we applied a generation time of 10 years (*Zhou* et al. 2014). In each parameter estimation procedure, 100,000 coalescent simulations and at least 20 expectation-conditional maximization cycles, up to a maximum of 40, were used. To get a reliable global maximum estimation for each scenario and avoid local maximum, we ran between 200 and 2,000 replicates. The AIC was used to compare the fit of different models and the model with the smallest AIC value was selected as the preferred model, which indicates that AIC differed significantly if AIC model $A + 2 < AIC$ Model B (*Excoffier* et al. 2013). The preferred point estimations of all parameters were chosen by ML method from 1,300 independent runs, and their confidence interval was calculated by simulating 200 independent SFSs conditional on the preferred demographic scenario selected. Re-estimation procedures were implemented for each simulated data set.

To further assess the fit of the optimal demographic model, we compared the expected and the observed SFS and genetic diversity. The expectations were obtained from 200 replicates.

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments
We thank Guo-Dong Wang and Wei-Wei Zhai for providing valuable suggestions on data analyses and helpful comments on this article. This work was supported by the grant from the National Nature Science Foundation of China to L.Y. ([91731311](http://natnscf.gov.cn/)), the grant from the National Youth Talent Support Program to L.Y. ([8113006](http://nyts.org.cn/)), the grant from Yunnan University’s Research Innovation Fund for Graduate Students ([YDY17097](http://yunnan.edu.cn/)), and the grant from Hubei Provincial Technological Innovation Major Project ([2018AcA132](http://hubei.gov.cn/)). This work was supported by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (Large Research Infrastructure Funding).

References


