



# Hybrid Speciation and Introgression Both Underlie the Genetic Structures and Evolutionary Relationships of Three Morphologically Distinct Species of *Lilium* (Liliaceae) Forming a Hybrid Zone Along an Elevational Gradient

Yundong Gao<sup>1\*†</sup>, AJ Harris<sup>2†</sup>, Huaicheng Li<sup>1,3</sup> and Xinfen Gao<sup>1\*</sup>

<sup>1</sup> CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China, <sup>2</sup> Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, <sup>3</sup> Institut für Biologie, Freie Universität Berlin, Berlin, Germany

## OPEN ACCESS

### Edited by:

Nina Ronsted,  
National Tropical Botanical Garden,  
United States

### Reviewed by:

Adrian Christopher Brennan,  
Durham University, United Kingdom  
Jianquan Liu,  
Lanzhou University, China

### \*Correspondence:

Yundong Gao  
gaoyd@cib.ac.cn  
Xinfen Gao  
xfgao@cib.ac.cn

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 26 June 2020

**Accepted:** 02 November 2020

**Published:** 07 December 2020

### Citation:

Gao Y, Harris A, Li H and Gao X  
(2020) Hybrid Speciation  
and Introgression Both Underlie  
the Genetic Structures  
and Evolutionary Relationships  
of Three Morphologically Distinct  
Species of *Lilium* (Liliaceae) Forming  
a Hybrid Zone Along an Elevational  
Gradient. *Front. Plant Sci.* 11:576407.  
doi: 10.3389/fpls.2020.576407

We studied hybrid interactions of *Lilium meleagrinum*, *Lilium gongshanense*, and *Lilium saluenense* using an integrative approach combining population genetics, fieldwork, and phenological research. These three species occur along an elevational gradient, with *L. meleagrinum* occurring at lower elevations, *L. saluenense* at higher elevations, and *L. gongshanense* between them. The species show strong morphological differentiation despite there being no clear environmental barriers to gene flow among them. *Lilium gongshanense* is likely to have a hybrid origin based on our prior work, but its progenitors remain uncertain. We sought to determine whether gene flow occurs among these three parapatric species, and, if so, whether *L. gongshanense* is a hybrid of *L. meleagrinum* and/or *L. saluenense*. We analyzed data from multiple chloroplast genes and spacers, nuclear internal transcribed spacer (ITS), and 18 nuclear Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR) microsatellites for accessions of the three species representing dense population-level sampling. We also inferred phenology by examining species in the field and using herbarium specimens. We found that there are only two types of chloroplast genomes shared among the three species and that *L. gongshanense* forms two distinct groups with closest links to other species of *Lilium* based on ITS. Taken together, *L. gongshanense* is unlikely to be a hybrid species resulting from a cross between *L. meleagrinum* and *L. saluenense*, but gene flow is occurring among the three species. The gene flow is likely to be rare according to evidence from all molecular datasets, and this is corroborated by detection of only one putative hybrid individual in the field and asynchronous phenology. We suspect that the rarity of hybridization events among the species facilitates their continued genetic separation.

**Keywords:** divergence with gene flow, introgression, species boundary, *Lilium*, Hengduan Mountains, Southwestern China, *Nomocharis*, hybridization

## INTRODUCTION

Natural hybridization, which may lead to hybrid speciation and/or genomic reticulation, is now widely understood to be a major evolutionary mechanism in plants that can drive morphological change and adaptation, as well as shape biogeographic patterns (Soltis and Soltis, 2009). Fundamentally, hybridization can be defined as the interbreeding of individuals that are distinguishable on the basis of one or more heritable characters; often distinct species (Harrison, 1993; Harrison and Larson, 2014). Rates of hybridization are variable among kingdoms of life, but on average, around 25% of plant species are known to hybridize with at least one other species (Mallet, 2005) compared to 10% of animals. Thus, hybridization plays an important, frequent role in shaping the biodiversity of plants (Grant, 1981; Mallet, 2005; De Queiroz, 2007; Harrison and Larson, 2014; Taylor and Larson, 2019).

The potential outcomes of hybridization include hybrid speciation, introgression, reinforcement, extinction of rare species, and merger of two parental species (Abbott et al., 2013; Todesco et al., 2016). Of these, hybrid speciation and introgression via hybridization occur most frequently but are difficult to distinguish from one another using typical molecular genetics methods (Harrison and Larson, 2014; Taylor and Larson, 2019). However, methods that explicitly test rates of introgression or complex evolutionary hypotheses (e.g., Anderson and Thompson, 2002; Beerli and Palczewski, 2010; Cornuet et al., 2014) are increasingly available and help to disentangle the complicated evolutionary histories of hybridizing taxa. These methods are complemented by careful examinations of morphology (e.g., to detect the prevalence of intermediate forms) and use of ecological indicators (e.g., to determine the degree to which parental species and hybrid offspring are environmentally compatible or incompatible). Therefore, case studies, such as in ragworts (Brennan et al., 2019; Nevado et al., 2020), azaleas (Yan et al., 2019), chestnuts (Sun et al., 2020), butterflies (Capblancq et al., 2019), birds (Hooper et al., 2019), and turtles (Scott et al., 2019), continue to provide valuable insights into the complex roles of hybridization and gene flow in evolution and speciation.

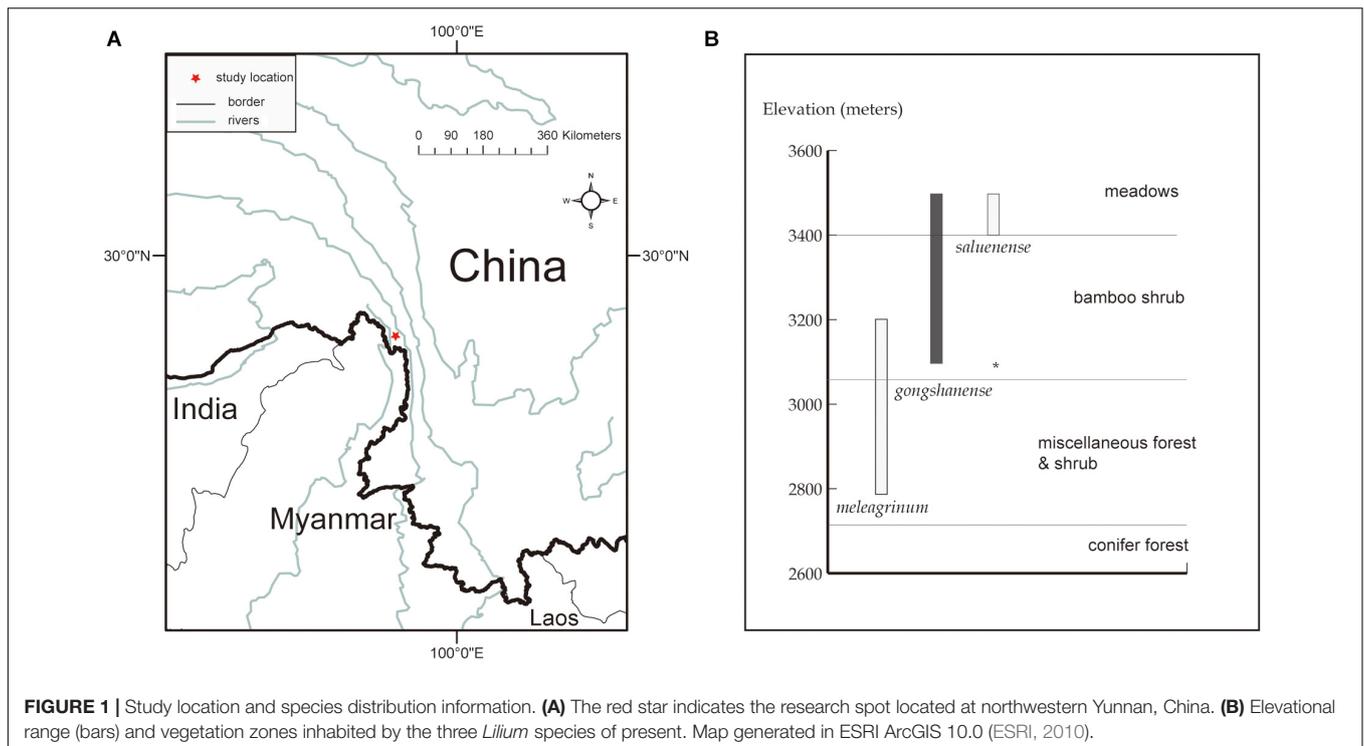
The species *Lilium gongshanense* (Y.D. Gao et X.J. He) Y.D. Gao (hereafter LG) also comprises a potentially valuable model system for these kinds of complex, reticulate evolutionary processes. This species may have hybrid origins based on prior studies (Gao et al., 2012a, 2013, 2015; Gao and Gao, 2016), and it currently occurs in close proximity to two other species of *Lilium*: *L. meleagrimum* (Franchet) Y.D. Gao, and *Lilium saluenense* (Balf. f.) S.Y. Liang (hereafter LM and LS, respectively). These three species were formerly treated within the genus *Nomocharis* Franchet, which is now known to represent a polyphyletic, but morphologically cohesive, group within *Lilium* (Gao et al., 2015) comprising ca. 15 species that are endemic in eastern Asia and mostly occur at high elevations, often in alpine areas. *Nomocharis* is polyphyletic for including a monophyletic group of *Lilium* that do not share the unique morphology of *Nomocharis* but also occur at high altitudes,

especially in alpine areas (i.e., non-*Nomocharis* lilies, or N-N lilies in Gao et al., 2015). Thus, although *Lilium* is a large genus of ca. 130 species, all three of these species of *Lilium*, LG, LS, and LM, are closely related to one another. In a prior study, Gao et al. (2012a) suggested that LS may be a progenitor of LG based on chloroplast DNA, while more recently, internal transcribed spacer (ITS) revealed that two distinct clades of LG that were sister to large clades of *Nomocharis* and N-N species (Gao et al., 2015).

LG, LS, and LM occur parapatrically along an elevational gradient, and it remains unclear whether there is gene flow among them, especially between LG and LM and LS. All three of these focal species are endemic to China and co-occur within a small region in northwestern Yunnan Province of southwestern China (Figure 1A). Along the elevational gradient, LG occurs between LS, which grows at the highest elevations in alpine meadows, and LM, which is found among lower elevation alpine shrubs (Table 1 and Figure 1B). In the case of parapatrically distributed species, a transition zone, sometimes bearing hybrids, usually forms between them (e.g., Streisfeld and Kohn, 2005; Butlin et al., 2008; Ortiz et al., 2018), such as in the position occupied by LG. Thus, it seems plausible that LG represents a hybrid of LS and LM. However, transitional zones typically represent sharp environmental contrasts, contain few hybrids, and are generally not reported to result in hybrid speciation (e.g., Hull et al., 1996). Moreover, LG is not an intermediate of LS or LM, and we have observed few or no intermediate forms representing any of the pairs of species. All three species are morphologically distinct.

The morphology of the three species has been fully discussed and illustrated in previous work (see Gao et al., 2012a, 2015; Gao and Gao, 2016). In short, LS has alternate leaves and pink tepals that have a dark purple blotch at the base and crimson blotches or spots over the rest of the tepal surfaces. LG has alternate leaves and pale yellow tepals and only few dark purple spots over the rest of the tepal surfaces. LM has opposite or whorled leaves and white or pale pink tepals. LM also differs by having fleshy, swollen, cylindrical filaments abruptly narrowing to filiform. The distinct appearances of these species make them easy to recognize in the wild (Figures 2a–f).

In this study, we sought to elucidate possible gene flow among LS, LG, and LM where they occur parapatrically and, consequently, better understand the evolutionary origins of LG. Specifically, we conducted population- and species-level analyses of these three taxa with the aim to resolve the following questions: (1) Is there ongoing gene flow among populations of LS, LG, and LM such that they form a hybrid zone or show evidence for historical gene flow? (2) If there is gene flow, do the genetic structures among them result from hybrid speciation or introgression? (3) How did the genetic structures form and how are they maintained? And (4) how do these species sustain their unique morphology if they experience gene flow? We expect that our case study of *Lilium gongshanense* can not only help to resolve the origins and evolution of the species but also help to inform increasingly more realistic approaches to integrating hybrid processes into evolutionary inference.



## MATERIALS AND METHODS

### Study Location

All three focal species in this study are endemic to China and co-occur in a small area within the Gaoligong Mountain Range in northwestern Yunnan Province of southwestern China ( $\sim N 27.80^\circ$ ,  $E 98.48^\circ$ , about  $100 \text{ km}^2$ ) along an elevational gradient of 2,800 to 3,500 m above sea level (Figure 1). There are limited specimen records indicating that LM and LS may also occur in adjacent areas. On the other hand, LG is only found in this area, within which we have never observed species of *Lilium* other than these three. Our study specifically focuses on a site within the Gaoligong Mountain Range that is centered around Mt. Heipu, which is divided into a northern and southern valley.

### Sampling

In 2009, we began an investigation at the study location. Following several field seasons, we performed comprehensive, population-level sampling of species in July 2016. We sampled populations along the gradient from the peak of Mt. Heipu into both the northern and southern valleys. Within populations, our sampling of individuals comprised only leaves in order to limit the ecological impacts of our study, except that we collected one specimen per population as a voucher, which we deposited in herbarium of CDBI. The sampled leaves represented individuals growing at least 20 m apart, and we stored the leaves in silica gel at  $-80^\circ\text{C}$  until use. We documented each sampled individual with a photograph and by obtaining geocoordinates. In total, we sampled nine populations and 139 individuals according to population size. The populations comprised LMn ( $n = 9$ ), LMXn

( $n = 10$ ), LGXn ( $n = 7$ ), LGm ( $n = 12$ ), LSs ( $n = 24$ , and an individual of putative hybrid origin, i.e., LSn,  $n = 1$ ), LGs ( $n = 25$ ), LGXs ( $n = 13$ ), LMXs ( $n = 10$ ), and LMs ( $n = 25$ ) (Figure 3). The population names consist of the abbreviation of species name (LS, LG, or LM), the distribution on the northern or southern slope of Heipu (n and s), and cases where populations of two species overlap (X) (Table 1 and Figure 3).

### DNA Extraction, Sequencing, and Microsatellite Genotyping

We extracted total genomic DNA from the dried leaf tissue via a modified cetyltrimethyl ammonium bromide method (Doyle and Doyle, 1987). We amplified and sequenced ITS and the cpDNA regions *trnL-F*, *rbcl*, *matK*, and *rpl32-trnL* to reconstruct phylogenetic and haplotype networks. We chose ITS because it occurs in multiple copies throughout the genome. Multiple copies may reveal a biparental genealogical signature, especially by showing two distinct copies, one from either parent. These signals can be easily be detected by observing double peaks in sequencing chromatograms. We selected the four cpDNA makers because two of them have been proposed as DNA barcodes for their high variability and amplification success (Shaw et al., 2007), and the other two showed suitable variation in prior studies of *Lilium* (Gao et al., 2013, 2015). We amplified all five markers using the primers from our previous study, Gao et al. (2015), via polymerase chain reaction (PCR) with 50 ng of genomic DNA in 20  $\mu\text{L}$  reactions in a GeneAmp PCR System 9700 (Applied Biosystems, United States). We performed the ITS reactions using the following thermocycler protocol:  $94^\circ\text{C}$  initial denaturation and enzyme activation for 2 min; 35

**TABLE 1** | General information of species and population composition of hybrid zone under study.

Population (size)	Species	Elevation range investigated (meters)	Plastid haploid	EST-SSR Type	Growth stage when sampling (year 2016)	Habitat
LSn*	Putative hybrid (LS × LG)	3100	H1	IV, V	Late bloom stage	Alpine bamboo shrub
LSs (24)	<i>Lilium saluenense</i>	3400–3500	H1	V	In bud and vegetation phase	Alpine meadow
LNGs (25)	<i>Lilium gongshanense</i>	3300–3400	H1, H2	III	Early bloom stage	Alpine bamboo shrub
LGXs (13)	<i>Lilium gongshanense</i>	3100–3300	H1	III	Full bloom stage	Alpine miscellaneous shrub
LMXs(10)	<i>Lilium meleagrimum</i>	3000–3200	H3, H5	II, IV	Full bloom stage	Alpine miscellaneous shrub
LGm(12)	<i>Lilium gongshanense</i>	3400–3500	H1	II, IV	Full bloom stage	Alpine bamboo shrub
LGXn(7)	<i>Lilium gongshanense</i>	3100–3200	H1, H6	II, IV	Full bloom stage	Alpine miscellaneous shrub
LMXn(10)	<i>Lilium meleagrimum</i>	3100–3200	H3, H4, H5	II, IV	Late bloom stage	Alpine miscellaneous shrub
LMn(9)	<i>Lilium meleagrimum</i>	2800–3000	H6	I, II, III	Late bloom stage	Alpine miscellaneous shrub
LMs(25)	<i>Lilium meleagrimum</i>	2800–3000	H2, H3, H4, H5	I, III	Late bloom stage	Alpine miscellaneous shrub

Population codes were made base on abbreviations of species names, as well as their relative distribution in the research spot (“n” as north, “s” as south and “m” for median). When two population are overlapped, a “X” symbol was applied. \*LSn contains only one individual; code in bold character indicating overlapping populations (two pairs adjacently).

cycles of 94°C denaturation for 30 s, 55°C primer annealing for 30 s, and 72°C extension for 60 s and a final extension of 72°C for 10 min. For the plastid markers, the amplification conditions were the same except that primer annealing was performed at 52°C for 45 s each cycle. We sent our amplified PCR products to Invitrogen Biotech Co., Ltd. (Shanghai, China) for purification and sequencing, which was done on an ABI-3730XL DNA sequencer. For each accession, forward and reverse sequencing reactions were performed for increased coverage.

For the resulting sequences, we performed manual editing in SeqMan Pro 7.1 (implemented in DNASTAR, Lasergene) and sequence alignment in MEGA 4 using the ClustalX method (Thompson et al., 1997) under default settings. We adjusted alignments by eye within MEGA4.0 and trimmed the alignments to the limits of the ITS and the plastid regions, respectively, by comparing with examples deposited in GenBank.

We also selected 18 primer pairs representing highly polymorphic EST-SSR markers (**Supplementary Table 1**) for amplification to assess demographic structures among the three species. To select the 18, we initially amplified 50 markers for a subset of individual collections of *Lilium* using primers designed for *Lilium regale* (Yuan et al., 2013), and we visualized results on a polyacrylamide gel electrophoresis. From among the 50, we selected 18 that consistently yielded good-quality amplicons. We separated PCR amplicons on a MegaBACE 1000 (GE Healthcare Biosciences, Sunnyvale, CA, United States) and scored alleles manually using GENETIC PROFILER software (version 2.2; GE Healthcare Biosciences). Final genotyping of EST-SSRs was conducted by Sangon Biotech (Shanghai) Co., Ltd. (Shanghai, China) using our amplifications.

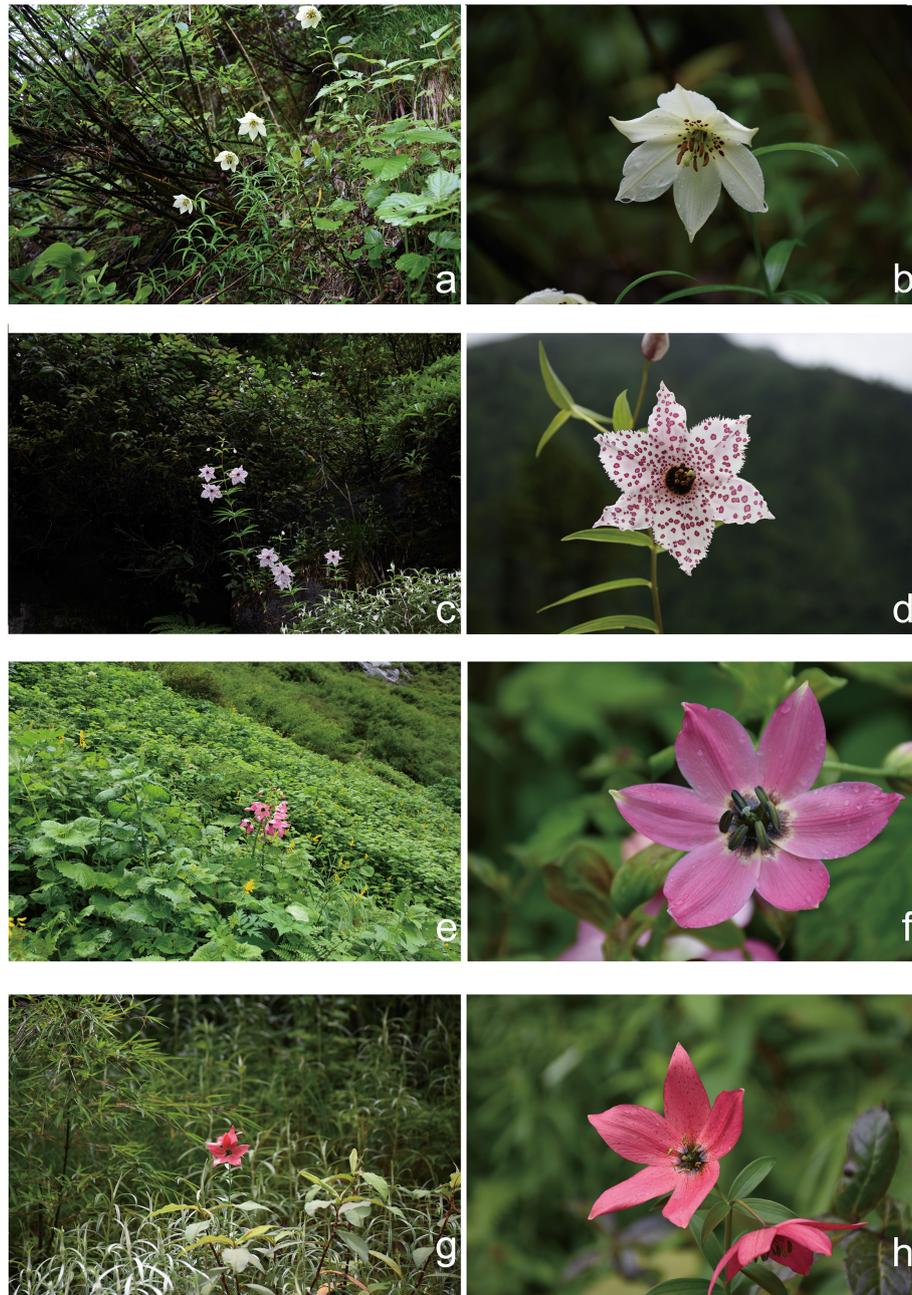
## Population Genetic and Phylogeny Data Analyses

We first verified congruence in phylogenetic signal among the four cpDNA fragments by carrying out a partition-homogeneity test (Farris et al., 1994) using the software PAUP\* 4.0b (Swofford, 2003). Based on the results, we used the combined cpDNA

dataset for all subsequent analyses. For the combined cpDNA, we analyzed sequence variation in MEGA4.0 and calculated nucleotide ( $\pi$ ) and haplotype ( $h$ ) diversity (Nei, 1987) in DnaSP 5 (Librado and Rozas, 2009). We also compared population differentiation for phylogenetically ordered ( $N_{ST}$ ) and unordered ( $G_{ST}$ ) haplotypes and for all populations using PERMUT1.0 (Pons and Petit, 1996) and performed 1,000 permutations to determine if  $N_{ST} > G_{ST}$ , which constitutes a test for phylogeographic structure. Additionally, we conducted analyses of molecular variance (AMOVA) (Excoffier et al., 1992) in the software package ARLEQUIN 3.5 (Excoffier et al., 2005) using all populations under the assumption that three species exist. We also applied AMOVA to the EST-SSR data with the same objectives, and in both cases we performed 1,023 permutations. We also investigated the phylogenetic relationships among populations represented by haplotypes using the statistical parsimony procedure for phylogenetic network estimations implemented in TCS 1.21 (Clement et al., 2000) with a 95% criterion for parsimonious connections.

To generate a phylogenetic network based on ITS, we used all new sequences generated in this study plus all available sequences of species formerly classified in *Nomocharis* and the N-N lilies (Gao et al., 2013, 2015). In total, we utilized 152 sequences of ITS, which we aligned using ClustalX in MEGA4.0 (Tamura et al., 2007) under default settings. Based on the alignment, we reconstructed an unrooted phylogenetic network in SplitsTree4 (Huson and Bryant, 2006) using the NeighborNet algorithm with Kimura 2-parameter (K2P) distances and ordinary least-squares inference of branch lengths. We performed a bootstrap analysis of the network in SplitTree4 with 1,000 replicates. We anticipated that the phylogenetic networks based on both ITS and cpDNA would help to reveal possible historical or ongoing gene flow among these three species and (in the case of ITS) their close allies that cannot be as readily detected using a bifurcating phylogenetic tree.

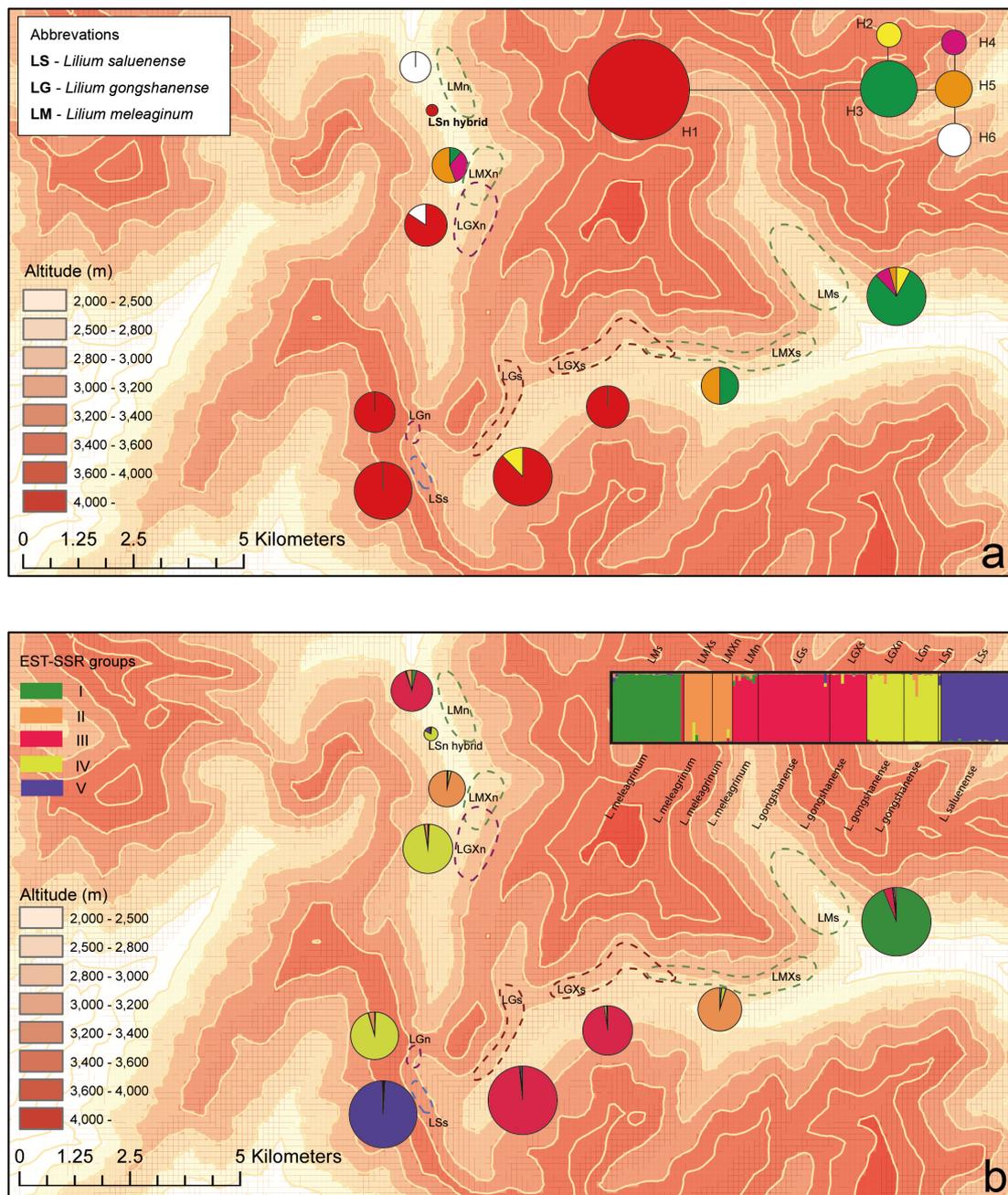
We primarily obtained population genetics statistics from the EST-SSR data in GenALEX version 6.501



**FIGURE 2** | Typical habits of taxa in present study. **(a,b)** *Lilium gongshanense*; **(c,d)** *L. meleagrinum*; **(e,f)** *L. saluenense*; **(g,h)** possible F2 hybrid between *L. saluenense* and *L. gongshanense*.

(Peakall and Smouse, 2012) and GENEPOP version 4.0 (Raymond and Rousset, 1995) except as otherwise noted. These methods enabled us to summarize genetic structures within and among populations. In particular, we calculated differentiation between populations using  $F_{ST}$  (Weir and Cockerham, 1984), for which we determined significance at each locus and overall using 1,000 permutations. Additionally, using GenALEX, we performed a principal coordinates analyses (PCoA) to assess the clustering of genotypes.

We assessed historical gene flow based on EST-SSR datasets using MIGRATE v. 4.2.1 (Beerli and Palczewski, 2010) on CIPRES (Miller et al., 2010). Using the continuous Brownian motion model, we started five independent Markov Chain Monte Carlo (MCMC) chains, each with 100,000,000 generations and sampling every 100 steps under a constant mutation model. We discarded the first 20% of sampled generations as “burn-in.” We checked for convergence among analyses, combined the results, and obtained the mode and 95% high probability density. Based



**FIGURE 3** | Geographic distributions of genotypes among 10 populations based on chloroplast and nuclear DNA. **(a)** Geographic distribution of six cpDNA haplotypes (H1–H6) and network of cpDNA haplotypes constructed by TCS 1.21. Sizes of circles in network are proportional to observed frequencies of haplotypes. **(b)** Geographic distribution of genotypes based on EST-SSR loci. Histogram of the STRUCTURE assignment test for the populations (139 individuals) shown at top. LM, *Lilium meleagrinum*; LG, *L. gongshanensis*; LS, *L. saluenense*. Map generated in ESRI ArcGIS 10.0 (ESRI, 2010).

on these results, we inferred historic gene flow between each pair of species.

We also used the complete EST-SSR dataset (i.e., nine populations comprised 139 individuals, with LSn and LSs combined) to infer population structure in STRUCTURE version 2.3 (Pritchard et al., 2009). We ran STRUCTURE under the admixture model with independent allele frequencies

for 100,000 MCMC generations following 10,000 burn-in generations. We performed 10 replications each in STRUCTURE for  $K = 1-20$ , and then we selected the optimal  $K$  according to Evanno et al. (2005) for downstream analyses.

We tested for specific levels of introgression among LS, LM, and LG and the hybrid status of the later using the EST-SSR markers within NewHybrids software

(Anderson and Thompson, 2002) version 2.0 + Developmental<sup>1</sup>. NewHybrids comprises a Bayesian framework for explicitly testing hypotheses of introgression, including the existence of F1 hybrids, where the hypotheses are represented as expectations about genotype compositions. For example, F1 hybrids theoretically have 50% of loci in which allele one arises from parent one, and allele two arises from parent two. In the other 50% of loci, allele one arises from parent two, and allele two from parent one. Other levels of parental contributions to hybrids can be set as desired as well as the condition of no hybridization, and all models are evaluated simultaneously. Therefore, we tested 25 alternative hypotheses for parental contribution, ranging from none, to the existence of F1 hybrids, to very small contributions from potential donors involving only 5% of loci (**Supplementary File 1**). Within an analysis, hybridization can be evaluated between only two lineages or species (see also Liu et al., 2019). Thus, we performed the tests of the 25 hypotheses on three datasets: a dataset of all sampled accessions of the three species of *Lilium*, a dataset comprising only LG and LS (in total 49 individuals with population codes LGXn, LGm, LSs, including the putative hybrid LSn) (Hypothesis 1, hereafter), and a third consisting of LG and LM (in total 64 individuals with population codes LMXn, LMXs, LMn, LGs, LGXs) (Hypothesis 2, hereafter). We chose the latter two datasets based on results from analyses in STRUCTURE and our personal observations in the field. NewHybrids is performed for individuals, so we obtained average posterior probabilities for each model for each population represented within our datasets.

## Field Observations and Phenology

We visited the study location four times in 2009, 2010, 2016, and 2018. During these visits, we sought to detect any recent (i.e., in generations) hybrids, such as F1 hybrids or recent backcrosses, in the field and to observe possible mechanisms that might prevent these unique species from homogenizing. Initially, we set coarse observational transects (0.5~1 km) and recorded the taxonomic turnover of species and habitat changes along the elevational gradient, growth stages of each population (i.e., early, full, and late bloom stages), and monitored for recent hybrids (**Table 1**).

Based on our initial observations, it appeared that the highest elevation species, LS, bloomed later than the other two, which largely overlapped in phenology. Thus, we compiled a robust phenological dataset for the three species using our observations at the study location and elsewhere in the Gaoligong Mountain Range and additional herbarium specimens. We realize that phenology may be strongly affected by environment, so that using samples from sites outside of the study location could introduce errors into our phenological estimations. However, all three species appear to have very narrow geographic distributions limited to a few peaks with the northern Gaoligong Range and appear to have narrow habitat (including elevation) preferences. Therefore, we expect that the additional samples are unlikely to show differing phenology based on habitat, and including them facilitated a larger sample size of these relatively rare species.

<sup>1</sup><https://github.com/eriqande/newhybrids>

To obtain phenological information from herbarium specimens, we visited the herbaria of PE, CDBI, and KUN in China and used digital images provided online by A, E, K, and P in Europe and North America. From among available herbarium specimens, we recorded the collection dates of specimens with flowers. For the final phenological dataset, we generated boxplots and performed independent *t*-tests for flowering times using Origin 8.0 (OriginLab Corporation, United States).

## Analysis of Ecological Differences Among Species

We sought to detect ecological boundaries among these three species of *Lilium* because such boundaries may play a role in the maintenance of their unique species identities. To attempt to detect ecological boundaries, we applied PCA and ecological niche modeling (ENM) to 68 georeferenced accessions of the three species based on both our field collections and specimen records from Chinese Virtual Herbarium<sup>2</sup>. These comprised 20 occurrences of LG and 24 occurrences each of LS and LM. To conduct these analyses, we obtained ecological data for abiotic and biotic features of the environment from the locations of the georeferenced accessions. The abiotic environmental variables comprised all 19 temperature and precipitation variables of the bioclim dataset (Hijmans et al., 2005)<sup>3</sup>, elevation (Gesch et al., 1999)<sup>4</sup>, cloudiness (or sunshine hours per year), water limitation on plant growth (Churkina and Running, 1998)<sup>5,6</sup> and gravel, sand, silt, and clay concentrations within top and subsoil (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). The biotic variables consisted of plant species richness by ecoregion (Kier et al., 2005)<sup>7</sup>, net primary productivity (Imhoff et al., 2004)<sup>8</sup>, percent forest and grassland cover by terrestrial ecoregion (Hoekstra et al., 2010)<sup>9,10</sup>, mammalian and bird species richness (Jenkins et al., 2013)<sup>11,12</sup> and soil microbial diversity (Serna-Chavez et al., 2013)<sup>13</sup>. In all, we obtained data for 39 abiotic and biotic environmental variables. Although not all variables may be directly related to species ecological tolerances (e.g., mean annual temperature, mammalian, and bird species richness), they are well-known proxies for many aspects of the environment that are likely determinates of where species occur (Jackson et al., 2009; Hatfield et al., 2013). We resampled all variables at 30-arc-second resolution and extracted their values for the georeferenced data points using the Spatial Analyst extension of Global Information Systems (GIS) layers in ArcGIS v. 10.4.1 (ESRI, 2010).

<sup>2</sup><http://www.cvh.ac.cn/>

<sup>3</sup><https://www.worldclim.org/>

<sup>4</sup><https://databasin.org/datasets/366a1bef53344c02bcd7d7611d5f61f7>

<sup>5</sup><https://databasin.org/datasets/6fcba40c990c483484a1d51dacf57d3c>

<sup>6</sup><https://databasin.org/datasets/321550b8dfeb459ca5856508621bb1bc>

<sup>7</sup><https://databasin.org/datasets/43478f840ac84173979b22631c2ed672>

<sup>8</sup><https://databasin.org/datasets/e416442a16ae4bb5b176515f2c0a6fff>

<sup>9</sup><https://databasin.org/datasets/ff10a966216a4995a036a2cf63f74ce2>

<sup>10</sup><https://databasin.org/datasets/735a3581a4634cc5bcc57f914f276b9>

<sup>11</sup><https://databasin.org/datasets/9d77254b6cea4b51b54e4e25d5f08360>

<sup>12</sup><https://databasin.org/datasets/6820867bbee4edba14f33f75d07c629>

<sup>13</sup><https://esdac.jrc.ec.europa.eu/content/global-soil-biodiversity-maps-0>

We performed PCA analyses with all variables, which we transformed using  $z$  scores, for all 68 georeferenced data points. Thereafter, we used PCA axes one and two for detecting clusters of populations of the three species of *Lilium* with a  $k$ -means analysis performed using the cluster v. 2.0.8 library (Maechler et al., 2009) in R. Within  $k$ -means, we allowed up to 10 possible clusters, and we detected the optimal number of clusters by comparing each with 500 Monte Carlo simulations in the factoextra v. 1.0.5 (Kassambara and Mundt, 2017) library for R and taking the first local optimum.

We sought to utilize ENMs as multivariate models against which to test the ecological similarities among the three species of *Lilium*. Thus, we followed methods similar to Raxworthy et al. (2007). In brief, we assumed that if individuals of the same species are ecologically distinct, then models generated using individuals of one species should show a better fit when tested using randomly selected individuals of the same species compared to using individuals of a different species. Although the limitations of using coarse-scale environmental data for ecological assessments of species that occur in such close proximity is suspicious, our intention is to recover the preliminary niche difference among species and explore the key environmental variables to favor the selection of field data in the future.

Prior to ENM, we reduced the number of environmental variables by removing all but one from among those with strong covariance. To detect covariance, we generated a Euclidean distance matrix of the  $z$ -transformed environmental variables for the populations of *Lilium* and used that to construct a UPGMA tree of dissimilarity. Using the UPGMA, we identified clusters of variables with dissimilarities roughly in units of standard deviation on account of the  $z$ -transformation of the variables. Within each cluster, we retained the variable with the highest absolute value for loading on PCA axis 1. This yielded 16 variables for ENM analysis (**Supplementary Figure 1**).

We carefully devised a geographic extent for the ENM based on our knowledge of the geographic range of the three focal species of *Lilium* (Liang, 1995; Liang and Tamura, 2000) and of natural geological boundaries within the region. To accomplish this, we first designed a bounding box generously inclusive of the known range of LS, LG, and LM. The bounding box consisted of corners at N 31°12'34.35", N 22°56'24.40", E 94°13'.99", and E 103°6'1.59" (**Supplementary Figure 2**). We then obtained the geographic extent encompassing all geological provinces that contacted the bounding box and that were partially or completely terrestrial. We obtained the GIS layers representing regional geological provinces from the US Department of Interior (datasets prv3al, prv3bl, and geo8ag)<sup>14</sup>. The amalgamation of all geological provinces contacting the bounding box served as the geographic extent for ENM, and we clipped all environmental variables to that geography.

We performed ENM in Maxent 3.4.1 (Phillips et al., 2004; Phillips and Dudík, 2008). Within Maxent, we generated models for LS and LM and tested them using all 20 samples of *L. gongshanense*. For comparison, we also modeled each species

independently using 25% subsampling and, alternatively, cross-validation for model testing. We built all models using a random seed and 20,000 points for background sampling. We made comparisons between models using the area under the curve (AUC) metric based on model testing.

## Isolation by Distance and Isolation by Environment Test

To further understand possible drivers of genetic boundaries among the three species, we calculated isolation by distance (IBD) and isolation by environment (IBE). For genetic distances, which comprised the response variable, we used pairwise  $F_{st}$  values for populations resulting from the cpDNA data and Nei's genetic distances for the nuclear EST-SSR data and performed tests using both datasets independently. For IBD, we integrated elevation into distance by using surface distances instead of standard Euclidean distances. To obtain surface distances, we converted the latitude, longitude, and elevation of our 10 sampled, georeferenced populations to  $x$ ,  $y$ , and  $z$  coordinates representing a point on the surface of the earth within an Earth-centered, Earth-fixed Cartesian coordinate system (Swinger et al., 2003). For conversion, we used a web portal<sup>15</sup> that assumed an ellipsoid earth represented by WGS 84 (Decker, 1986). Following the conversion, we generated a matrix of pairwise surface distances in R v. 3.5.3 ("Great Truth") using the dist() function of the core library. For IBE, we obtained environmental distances by extracting environmental values for the populations representing the same variables and GIS sources used for ENM. We extracted the variables using the raster library in R (Hijmans et al., 2013) and used these data to generate a distance matrix. We generally followed Jiang et al. (2019) in transforming the predictor variable (i.e., environment in IBE and geographic distance in IBD) by the other, potentially nuisance variable (i.e., geographic distance for IBE and environment for IBD). We performed the transformation by weighting (multiplying) distances of the variable of interest by the distances of the other variable. We performed Mantel tests to assess IBE and IBD in the vegan library of R with 9,999 permutations (Jiang et al., 2019).

## RESULTS

### Sequence Data and Tests for Gene Flow and Genetic Structures

The aligned cpDNA data represent 139 individuals from 10 populations and comprised 2,989 bp including indels or 2,965 bp when indels were excluded. The data consisted of nine single-nucleotide polymorphisms, of which all were parsimony informative. From among 139 individuals, we identified six haplotypes (H1–H6, GenBank accession numbers MN635282 to MN635305), which were divided into two major clusters by 10 steps based on single nucleotide polymorphisms (**Figure 3** and **Supplementary Figure 3**). Among the haplotypes, H1 comprises two species (LS and LG) with six populations and 83

<sup>14</sup><https://data.doi.gov/>

<sup>15</sup><https://www.oc.nps.edu/oc2902w/coord/llhxyz.htm>

individuals and constituted one major cluster. The other major cluster comprised H2–H6 and included two species (LM and LG) with five populations and 56 individuals (**Supplementary Figure 3**). Within this cluster, we found that LG possesses only the haplotypes H2 and H6. Overall, we found that LG has two major types of chloroplast DNA and that all three species share only these two major types. These data support the existence of gene flow between LG and LM and between LG and LS.

Genetic statistics representing the cpDNA generally revealed high diversity values. In particular, the total nucleotide ( $\pi$ ) and haplotype ( $H_d$ ) diversity across the metapopulation were 0.00113 and 0.600, respectively, and genetic diversity for the metapopulation ( $H_T = 0.673$ ) was higher than the average diversity within populations ( $H_S = 0.206$ , **Supplementary Table 2**). Moreover, the AMOVA using cpDNA showed a very high level of variation among groups defined according to morphology and classical taxonomy ( $PV = 83.77\%$ ,  $F_{CT} = 0.8798$ ,  $p < 0.001$ , **Supplementary Table 3**). Taken together, these data from cpDNA support genetic distinctiveness of the three species.

The ITS sequences consisted of 152 accessions [139 newly generated and the GenBank accession numbers MN636494 to MN636632 (numbered as SUB6504140)<sup>16</sup>] including the three species of *Lilium* that are the subjects of this study, other *Lilium* formerly classified as *Nomocharis*, and the N-N lilies with accession numbers list in **Supplementary Table 4**. We excluded the ITS2 region of ITS because we were unable to convincingly align it during preliminary analyses due to high levels of polymorphism. The remaining characters consisted of ITS1 and 5.8S and contained 61 variable sites, of which 45 were parsimony informative.

The phylogenetic network based on nuclear ITS resolved LS with its close allies, consistent with the phylogenetic framework developed in prior studies (Gao et al., 2013, 2015). However, the species LG were poorly resolved and divided into two clusters, separated by species of N-N lilies (Gao et al., 2015), although unexpectedly, species of LM comprised two clusters separated by several former species of *Nomocharis* (**Figure 4**). The branch lengths for individuals of LG in the network are also much longer, on average, than for other species, suggesting high intraspecific variation (**Figure 4**). These data suggest that both of the parental species of LG may be N-N lilies and that it originated recently enough so that concerted evolution has not yet eliminated one parental type of ITS. More ITS sequences from other former *Nomocharis* [especially *Lilium sealyi* Y.D.Gao, *L. basilissum* (W.E. Evans) Y.D. Gao, and *L. pardanthinum* (Franchet) Y.D.Gao, formerly named *Nomocharis farreri* (W.E.Evans) Harrow, *N. basilissa* W.E.Evans and *N. pardanthina* Franchet, respectively] is needed to resolve the hybridization interactions of LM. Overall, the ITS data show that, although gene flow is occurring between LG and LM and LS, respectively, LM and LS are likely not progenitors of LG.

We detected a total of 173 alleles from among the 18 EST-SSR loci surveyed (**Supplementary Table 5**). We used these results and STRUCTURE to evaluate the genetic distinctiveness

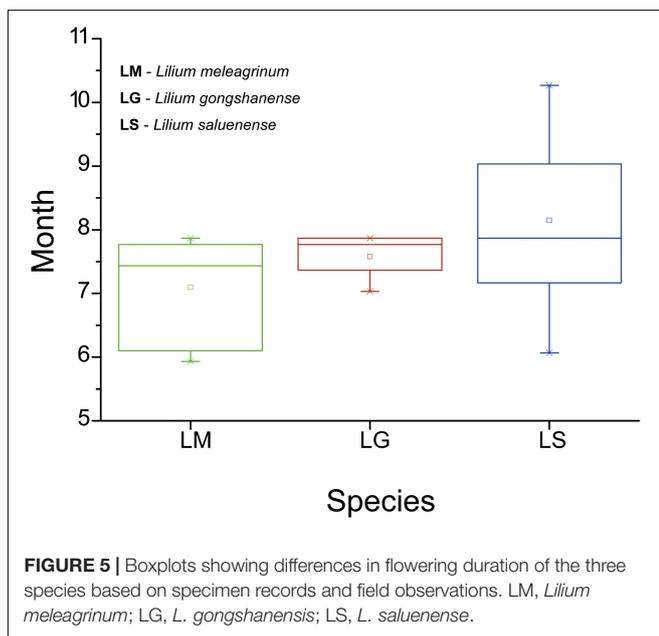
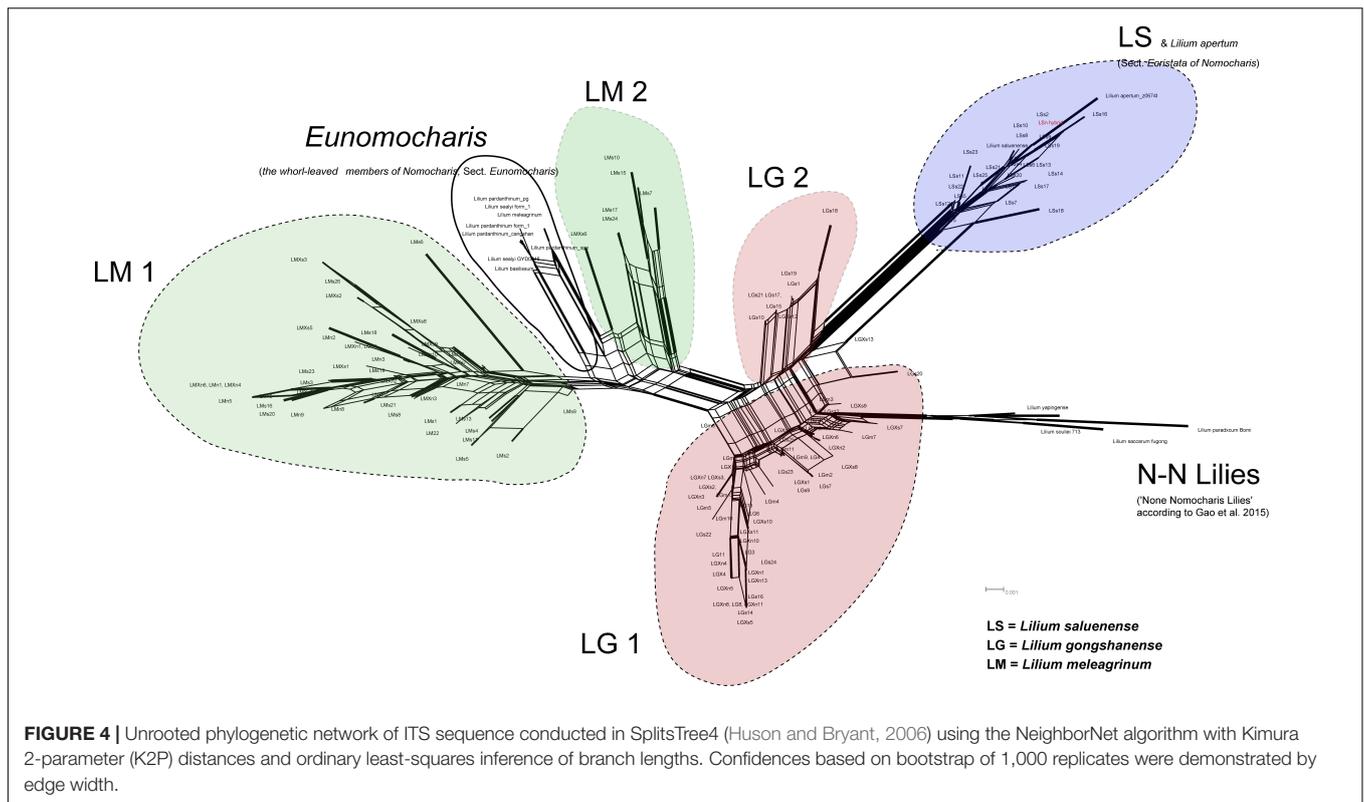
of the populations and three species. We tested three alternative possibilities setting structure to evaluate three different numbers of populations ( $K$ ): (1) only two species exist, and LG is hybrid of the other two ( $K = 2$ ); (2) there are three independent species as indicated by morphology ( $K = 3$ ); and (3) there are five populations, which is the most optimal number according to the outcome of the statistical analysis to determine cluster number (Evanno et al., 2005) ( $K = 5$ , **Supplementary Figure 4**). When  $K = 2$ , LG was genetically isolated from LS (except LS<sub>n</sub>, which we believe may be a single hybrid occurrence between LG and LS; see below) but not isolated from LM (**Supplementary Figure 5**). At  $K = 3$ , all three species are resolved as genetically distinct. At  $K = 5$ , LG and LM were independently divided into two groups each. Comparing the results of different values of  $K$  (**Supplementary Figure 5**), it is clear that LS has greater genetic distinctiveness from LM and LG, which, together, show greater genetic associations. Moreover, the PCoA showed a clear cluster of LS and an intergrading cluster of LM and LG when plotted according to axes 1 and 2 (**Supplementary Figure 6**). However, the plot of PCoA axes 2 and 3 revealed two clusters of NG consistent with the ITS network analysis in SplitsTree and linked one cluster more strongly with NS. In contrast, the historic gene flow based on the EST-SSR dataset using MIGRATE showed no significant effective population exchanges among these three species (**Supplementary Table 6**), indicating relatively stable boundaries have existed among them.

Based on EST-SSRs, the NewHybrids analyses for Hypothesis 1 and Hypothesis 2 showed that the most probable explanation for the genotypic data was that each population consists largely of a single genetically distinct species without much hybridization (**Supplementary Table 7**). Under Hypothesis 1, all LS populations belonged to one genotype [0.98 posterior probability (pp)], whereas all LG populations were resolved as a second genotype (0.41–0.70 pp). Under Hypothesis 2, the LMXs and LMX<sub>n</sub> populations were resolved as belonging to one genotype (0.89 and 0.88 pp, respectively), and all other populations, including LMs, were resolved as a second genotype (0.96–0.99 pp). However, in the NewHybrids analysis with all species, all populations were resolved as having 95% of genes arising from only one species, whereas the other 5% include introgressed alleles (0.55–0.80 pp). In the absence of detecting any F1 hybrids or populations with average levels of introgression suggesting abundant F1 hybrids, hybridization events are likely to occur only rarely among the three species.

## Field Observations and Phenology

In total (including our field observations), we collected phenological data from 44 records (LM = 15, LG = 5, LS = 24) with flowers. Our results showed that flowering time differs among the three species (**Figure 5**). However, only the difference between LM and LS is significant (LM vs. LS,  $t = -3.137$ ,  $p = 0.003$ ; LM vs. LG,  $t = -1.414$ ,  $p = 0.174$ ; LS vs. LG,  $t = 1.068$ ,  $p = 0.295$ ). The flowering time of LG overlaps partially with the flowering times of the other two species. These results suggest that, while some opportunities for gene exchange especially between LM and LG and between LS and LG exist, they may be temporally limited and therefore rare.

<sup>16</sup><https://submit.ncbi.nlm.nih.gov/subs/genbank/SUB6504140/overview>



Via our observations of species in the field, we observed a single, isolated individual resembling LS at 3,100 m where populations of LG occur. The morphology of this individual comprised open-flat pink tepals with tiny spots, linear alternate leaves, and a dwarf habit. This individual comprised a single sample, which we treated as a distinct “population,” LS<sub>N</sub>

(Figures 2g,h), in our analyses, and it was supported as being a hybrid of LG and LS based on analyses of EST-SSR in STRUCTURE (Figure 3 and Supplementary Figure 5) but not supported by NewHybrids. However, in the field, it was morphologically similar to LS and did not show intermediate features.

### Ecological Distance and Isolation

In the PCA analyses, we detected three groups of individuals using k-means (Supplementary Figure 7). These were separated on axis 1, representing primarily abiotic variables, and on axis 2, primarily consisting of biotic variables (Supplementary Figure 2). The three clusters did not correspond to existing species boundaries. AUCs from ENM analyses revealed that models for LS and LM were equally accurate when tested using LG as when tested using intraspecific cross-validation or subsampling (Supplementary Figure 8). The IBD tests using the cpDNA and nuclear EST-SSR showed no significant relationship between genetic and non-Euclidean geographic distance (Table 2; for

**TABLE 2 |** The results of Isolation by Distance (IBD) and Isolation by Environment (IBE) based on nuclear SSR and chloroplast (cp) datasets.

Test	Genome	Stat (Pearson's <i>r</i> )	<i>p</i> -value
IBD	Nuclear	−0.07718	0.7118
IBD	Cp	0.1822	0.0909
IBE	Nuclear	0.789	0.0001
IBE	Cp	0.8074	0.0001

the distance matrices, see **Supplementary File 3**). However, IBE based on each of the cpDNA and nuclear EST-SSRs showed a highly significant relationship to genetic distance. Although these species visibly occur in different environments (e.g., along an elevational gradient, within different habitats), many dimensions of the ecological niche may be shared among them, and differences in their ecological tolerances may be slight, overall.

## DISCUSSION

### Gene Flow Occurs Rarely Among the Three Species

Prior work suggested that LG is a hybrid species based on possessing two types of ITS and on morphological intermediacy (Gao et al., 2012a). Within the prior work, we inferred that the maternal progenitor of LG was LS because sampled individuals of both species shared the same chloroplast genome. However, the results here show that LG possesses two types of cpDNA genomes: one consistent with LS and another consistent with LM. Although we detected six cpDNA haplotypes (H1–H6, **Figure 3** and **Supplementary Figure 3**), it is clear that H1 is highly distinct from the others (10 changes), whereas H2–H6 are more closely related and show only one change each. Our results suggest that both the maternal and overall history of the species is more complex than we reported in our prior study.

While the presence of gene flow among these species is supported by both the cpDNA and EST-SSRs (**Supplementary Figures 3, S5**), especially between LM and LG, most lines of evidence suggest that gene flow is likely rare. In the field, we observed only one individual, LS<sub>n</sub> (**Figures 2g,h**), which may represent an F1 hybrid or other relatively recent hybridization event. We suspect that the LS<sub>n</sub> individual represents a rare hybrid of LS and LG because of its morphology, which is LS-like, and its genetic composition according to STRUCTURE (**Figure 3** and **Supplementary Figure 5**) and the PCoA. The STRUCTURE analysis linked LS<sub>n</sub> with LG, whereas the PCoA shows a clear relationship of this individual to LM. Taken together, this suggests that an LG individual with a background of introgression with LM hybridized with an LS individual. This is consistent with the finding from multiple lines of evidence (**Figure 3b** and **Supplementary Figures 3, S5**) that introgression plays a comparatively greater role in the evolutionary dynamic between LM and LG and also highlights the complicated nature of the evolutionary histories of these species, especially LG. Notably, NewHybrids does not support that LS<sub>n</sub> has substantial genetic contributions from either LS or LM (**Supplementary Table 7**; Hypothesis 1 in NewHybrids showing that LS<sub>n</sub> is a “pure” LG but with negligible support). Overall, additional field work is needed to detect what may be rare instances of hybridization facilitating gene flow and explaining introgression.

### Gene Flow Yields Biased Introgression, but Not Hybrid Speciation

Based on the ITS phylogenetic network (**Figure 4**), LG was divided into two major clusters representing two distinct types

of ITS. Possession of two types of ITS has long been considered strong evidence for a hybrid origin, including in several studies on species of Liliaceae (e.g., in *Gagea*) (Peterson et al., 2009, 2011; Peruzzi et al., 2011). Typical patterns in the evolution of ITS have been widely studied, and normally, the multiple copies of ITS within a genome undergo concerted evolution so that all copies are inevitably identical (Zimmer et al., 1980; Koch et al., 2003). However, concerted evolution may have two outcomes in hybrid species: (1) loss of one copy and fixation of the second or (2) evolution of a new ITS type that represents a mixture of the two original ITS sequences. In cases of recent hybrid speciation, both ITS copies may be present, because concerted evolution is incomplete (reviewed in Koch et al., 2003). Therefore, in the case of an ancient hybrid speciation event, we would expect concerted evolution to have led to a single copy of ITS in the modern species. In the case of recent hybridization, the two copies of ITS should show the closest relationships with each parental species, or, as we found, with a large group of species (**Figure 4**), LG may result from a recent hybridization event.

Overall, our genetic results seem to suggest low, ongoing introgression between the three parapatric species under study and therefore support the existence of a hybrid zone among them. Gene flow between species comprises hybridization followed by backcrossing of the hybrid offspring into one or both parental species. Gene flow between LG and parapatric species, LM and LS, compounds the already-complex evolutionary history of the hybrid taxon. Notably, LG and LM appear to be experiencing greater gene flow than LG and LS, or, alternatively, LG and LM bear a remaining signature of ancient gene flow in addition to gene flow in the present. In either case, the outcome is that introgression between LG and LM vs. LG and LS has been asymmetrical.

### Formation and Maintenance of Genetic Structures

Our results show asymmetrical introgression between LS to LG such that backcrossing occurs more often with LG than with LS. Thus, genes of LS are more commonly introduced into populations of LG and less often the other way around. This is consistent with our observation that gene flow appears more often downslope and that there are upslope barriers (**Figure 3**). This observed pattern of biodiversity may result from differences in phenology in concert with dichogamy in *Lilium*, that is, the maturation of pollen before maturation of the stigma. LS has a later flowering time than either LG or LM based on our observations in the field and of museum specimens (**Figure 5**). Because of this, the mature pollen of LS may reach a mature stigma of LG, but pollen from LG has already dispersed by the time stigmas of LS are mature. This bias in pollination due to phenology may explain the limited gene flow we detected between LG and LS based the nuclear EST-SSRs (**Supplementary Table 7, Figure 3a**, and **Supplementary Figure 6**) showed an independent identity that LS was largely genetically distinct. In contrast with LS, LG, and LM have a more highly overlapping phenology so that interbreeding between populations of these species may be more bidirectional. Nevertheless, the asymmetry in introgression

is also apparent with LG and LM, so that LM is more often the recipient of LG alleles instead of a donor to LG (**Figure 3** and **Supplementary Figure 5**). Thus, here too, the pattern of gene flow is largely downslope, at least for the nuclear genome. More broadly, differences in phenology have also been identified as the cause of asymmetrical introgression in several genera including *Populus* (Lexer et al., 2005) and *Iris* (Martin et al., 2007). However, we still cannot exclude that other biotic factors such as pollinator interactions also affect the direction of gene flow.

Asymmetry of introgression is also apparent in the plastid genomes, such that LG and LS share more haplotypes in common than LG and LM. The plastid introgression may be best explained by seed dispersal. All species of *Lilium* produce many seeds that are encircled by flat, narrow wings and are arranged like a pile of coins inside of a capsule fruit (Liang and Tamura, 2000). Seeds of *Lilium* are thought to have very low dispersal abilities because the fruits are dry and lack animal rewards and, despite that they have wings, the seeds can only glide over short distances in most cases (Minami and Azuma, 2003; Horning and Webster, 2009). Thus, seed dispersals over larger distances, such as between populations, may occur by gravity where the capsule is the dispersal propagule for unreleased seeds. In this case, seeds of LS are likely to be distributed to populations of LG, which are downslope and situated between LS and LM, which occur further downslope.

## How Species Maintain Their Identities Despite Gene Flow

Overall, there was limited gene flow detected among species, although gene flow appears more common between LG and LM, and only low levels of introgressed nuclear genetic material are present as evidenced by our results in NewHybrids (**Supplementary Table 7**). Elimination of introgressed loci may occur if alleles are locally maladaptive or genetically linked with weakly deleterious alleles, e.g., as in *Heliconius* butterflies (Edelman et al., 2019). Notably, elimination of introgressed DNA from a genome takes longer when genome sizes are large (Du et al., 2017), such as in *Lilium*, which has among the largest genomes known in flowering plants [44.88–167.58 pg (Du et al., 2017);  $38 \pm 7.28$  pg]<sup>17</sup>. Thus, the rather small amounts of introgressed DNA detected among the species seem to corroborate our observations in the field that hybrid speciation events are rare, perhaps even more so among LG and LS. The rarity of hybridization events between the species probably plays a crucial role in the ability of each to maintain its distinctive morphology and genetic identity (Gao et al., 2015).

One commonly invoked mechanism by which species retain their identities despite introgression is by ecological filtering, whereby the abiotic or biotic niche of a parental species may select for individuals, including hybrids, which have the traits of that parent and against the establishment and persistence of individuals with traits of the other parent or that exhibit intermediacy (Noble and Slatyer, 1980; Kraft et al., 2015). Thus, a hybrid offspring may survive and reproduce in the population of a parent with which it is morphologically

similar, while intermediate prodigy or those bearing traits of the other parent may have constrained survivability and fitness. While the environmental niches of these three species are remarkably overlapping, based on our ENM results (**Supplementary Figure 8**), the IBE test (**Table 2**) using both the cpDNA and nuclear EST-SSR data shows that environment may play a significant role in maintaining species boundaries. Therefore, environmental filtering may be critical to the maintenance of these entities as distinct species.

The occurrence of parental-like hybrids and the exclusion of intermediates could be facilitated by genetic linkages on the large chromosomes of *Lilium* (Peruzzi et al., 2009; Gao et al., 2012b). Huge genome's low recombination rate in plant also might slow the interspecific gene exchange (Tiley and Burleigh, 2015). Linkages may enable locally adaptive suites of traits that are connected to species defining morphologies to occur together much more often than we would expect for unlinked meiotic events (Rieseberg et al., 2000). Thus, environmental filtering and genetic linkage on large chromosomes may operate simultaneously to enable species to maintain unique identities. Genetic linkage merits further exploration in *Lilium*, including its relationship to ecological filtering, using whole genomic data (e.g., as in Edelman et al., 2019).

## CONCLUSION

Through our integrative approach using field observations, biotic and abiotic features of the environment, and population genetics based on the nuclear and chloroplast markers, we have elucidated a rather complex history for LG involving both ancient hybridization and subsequent gene flow within a hybrid zone involving species that are not progenitors. We believe this species comprises an exceptional model for assessing the robustness of methods, such as ABBA-BABA (Martin et al., 2015), which seek to infer evolutionary histories of species given confounding events. Robust methods should be able to infer multiple hybridization and introgression events at different points in the evolution of a lineage, and new models may be developed around well-understood, natural cases, such as LG. Genomic data are also needed to yield the most robust insights into the genomes involved in the origins of LG, as well as of the putative hybrid interactions of LM.

Despite the complexity of the evolutionary history of LG, we show that it is a cohesive, distinct species. Moreover, the boundaries between it and LM and LS are actively maintained, possibly by phenology and aspects of the environment, including (but likely not limited to) elevation. Boundaries between species in parapatry are maintained by intrinsic barriers to gene exchange. However, the concept of species boundary implies that species are evolutionarily independent populations that are reproductively isolated from other such species, such as according to the “biological species concept,” despite recognition of hybridization as an evolutionary process (Mallet et al., 2016). In reality, species boundaries are unlikely to be uniform in space, in time, or across

<sup>17</sup><https://cvalues.science.kew.org/>

the genome (Harrison and Larson, 2014; Larson et al., 2014), so that hybridization can occur while, simultaneously, the distinct character of species can be maintained. New concepts such as “genomic islands of divergence” and “divergence hitchhiking” (Bay and Rugg, 2017; Larson et al., 2017; Ma et al., 2018) are emerging to help explain how species can sustain their identities while gene flow is active. Fundamental to these concepts is that species distinct identities may arise from selection for or against a relatively few genes and adjacent, chromosomally linked regions. The complex evolutionary history of LG makes it a suitable model for development and testing of new approaches to resolving reticulate evolutionary origins.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

YG and XG conceived and designed the experiments. HL and YG performed the experiments. YG and AH analyzed the data and contributed to the reagents, materials, analytical tools, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by grants from the National Natural Science Foundation of China (NSFC Grant Nos. 31500163 and 31770406) and the Biological Resources Programme, Chinese Academy of Sciences (Grant No. KFJ-BRP-017-17) to YG.

## ACKNOWLEDGMENTS

We appreciate Dr. Meng Li, Mr. Wenbin Ju, and Ms. Juan Xue, as well as the Gongshan Bureau of Gaoligongshan Natural Preserve for assistance with collection of materials. We are also grateful for the staff and curators of PE, SZ, and KUN herbaria for help on investigating specimens.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.576407/full#supplementary-material>

**Supplementary Figure 1** | UPGMA tree showing similarity of z-transformed environmental variables among species occurrences used for ecological niche modeling. We selected one variable from each clade (black circles) with height greater than six (red line) to use in modeling. The heights are roughly in units of

standard deviation due to the z-transformation. Within clades, we selected the variable with the highest loading on PC axis one according to our PCA analysis.

**Supplementary Figure 2** | Geographic extent used for modeling shown in light blue based on geological units (dark blue) contacting a bounding box generously encompassing the range of the three focal species (black) and excluding marine areas and lakes.

**Supplementary Figure 3** | Network of cpDNA haplotypes presented by all populations (**A**) and 3 species (**B**), respectively. Size of circles in network are proportional to observed frequencies of haplotypes, and line distance indicated the mutations.

**Supplementary Figure 4** | The optimal value of  $K$  ( $K = 5$ ) based on the method of Evanno et al. (2005) using EST-SSRs of STRUCTURE analysis.

**Supplementary Figure 5** | Histogram of the STRUCTURE assignment test for all populations (139 individuals) inferred by eighteen nuclear microsatellite (EST-SSR) loci with different scenarios (explained in the main text).

**Supplementary Figure 6** | PCoA of scoring for EST-SSRs for individuals representing the three species of *Lilium*. Populations of LG are shown in brown/orange/yellow, populations of LS are shown in reds, and populations of LM are shown in blues. Percentage of variation explained is presented along with axis labels.

**Supplementary Figure 7** | PCA analysis of ecological boundaries among putative species of *Lilium* including three groups of individuals using k-means. Axis one representing primarily abiotic variables, and on axis two, primarily consisting of biotic variables (**Supplementary File 2**).

**Supplementary Figure 8** | Comparisons between models using the area under the curve (AUC) metric based on model testing. ENM performed in Maxent 3.4.1 (Phillips et al., 2004; Phillips and Dudík, 2008). Models generated for *Lilium saluenense* and *L. meleagrinum* and tested them using all 20 samples of *L. gongshanense*. Each species modeled independently using 25% subsampling and crossvalidation for model testing.

**Supplementary Table 1** | EST-SSR marker and primer sequences accordingly. Eighteen primer pairs for highly polymorphic EST-SSR for amplification to assess demographic structures among the three species were based on primers designed for *Lilium regale* (Yuan et al., 2013).

**Supplementary Table 2** | Genetic diversity parameters for all 10 populations of *Lilium*.

**Supplementary Table 3** | Hierarchical analysis of molecular variation (AMOVA) based on data from the three chloroplast DNA (cpDNA) spacers and EST-SSRs for all 9 populations of *Lilium* separated by taxa.

**Supplementary Table 4** | ITS accessions used in present to build the network (**Figure 4**).

**Supplementary Table 5** | Genotype data of EST-SSR including 18 locus of 139 individuals used in present. Missing data were marked as “-9” and specified in STRUCTURE analysis.

**Supplementary Table 6** | Posterior distribution of effective population sizes ( $\theta$ ) for each species, and effective migration rates ( $M$ ) between each pair of the species based on analyses in MIGRATE.

**Supplementary Table 7** | Table of population averages for NewHybrids models tested for all data, and two hypotheses involving a smaller number of populations. Bold type indicates the most highly supported model of parental contributions for each population.

**Supplementary File 1** | The 25 models tested using NewHybrids for the full dataset as well as Hypothesis 1 and 2 (explained in text).

**Supplementary File 2** | Loadings of all sampled environmental variables on all axes of PCA analysis.

**Supplementary File 3** | Pairwise population  $F_{st}$  values based on Nei genetic distance using EST-SSRs and chloroplast makers, as well as the earth surface distances instead of standard Euclidean distances used in IBD and IBE test.

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246.
- Anderson, E., and Thompson, E. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160, 1217–1229.
- Bay, R. A., and Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proc. R. Soc. B Biol. Sci.* 284:20162414. doi: 10.1098/rspb.2016.2414
- Beerli, P., and Palczewski, M. (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185, 313–326. doi: 10.1534/genetics.109.112532
- Brennan, A. C., Hiscock, S. J., and Abbott, R. J. (2019). Completing the hybridization triangle: the inheritance of genetic incompatibilities during homoploid hybrid speciation in ragworts (*Senecio*). *AoB Plants* 11:ply078.
- Butlin, R. K., Galindo, J., and Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2997–3007. doi: 10.1098/rstb.2008.0076
- Capblancq, T., Mavárez, J., Rioux, D., and Després, L. (2019). Speciation with gene flow: evidence from a complex of alpine butterflies (*Coenonympha*, Satyridae). *Ecol. Evol.* 9, 6444–6457. doi: 10.1002/ece3.5220
- Churkina, G., and Running, S. W. (1998). Contrasting climatic controls on the estimated productivity of global terrestrial biomes. *Ecosystems* 1, 206–215. doi: 10.1007/s100219900016
- Clement, M., Posada, D., and Crandall, K. A. (2000). TCS, a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x
- Cornuet, J.-M., Pudlo, P., Veysier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763
- De Queiroz, K. (2007). Species concepts and species delimitation. *Syst. Biol.* 56, 879–886. doi: 10.1080/10635150701701083
- Decker, B. L. (1986). *World Geodetic System 1984*. St Louis, MO: Defense Mapping Agency Aerospace Center, 24.
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Du, Y. P., Bi, Y., Zhang, M. F., Yang, F. P., Jia, G. X., and Zhang, X. H. (2017). Genome size diversity in *Lilium* (Liliaceae) is correlated with karyotype and environmental traits. *Front. Plant Sci.* 8:1303. doi: 10.3389/fpls.2017.01303
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science* 366, 594–599. doi: 10.1126/science.aaw2090
- ESRI (2010). *ArcMap*. Redlands, CA: ESRI.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294x.2005.02553.x
- Excoffier, L., Laval, G., and Schneider, S. (2005). ARLEQUIN (version 3.0), an integrated software package for population genetic data analysis. *Evol. Bioinform. Online* 1, 47–50.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes, application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- FAO/IIASA/ISRIC/ISSCAS/JRC (2012). *Harmonized World Soil Database (version 1.2)*. Rome: FAO.
- Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. (1994). Testing significance of incongruence. *Cladistics* 10, 315–319. doi: 10.1111/j.1096-0031.1994.tb00181.x
- Gao, Y. D., and Gao, X. F. (2016). Accommodating *Nomocharis* in *Lilium* (Liliaceae). *Phytotaxa* 277, 205–210. doi: 10.11646/phytotaxa.277.2.8
- Gao, Y. D., Harris, A. J., and He, X. J. (2015). Morphological and ecological divergence of *Lilium* and *Nomocharis* within the Hengduan Mountains and Qinghai-Tibetan Plateau may result from habitat specialization and hybridization. *BMC Evol. Biol.* 15:147. doi: 10.1186/s12862-015-0405-2
- Gao, Y. D., Harris, A., Zhou, S. D., and He, X. J. (2013). Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the Q-T plateau and the Hengduan Mountains. *Mol. Phylogenet. Evol.* 68, 443–460. doi: 10.1016/j.ympev.2013.04.026
- Gao, Y. D., Hohenegger, M., Harris, A., Zhou, S. D., He, X. J., and Wan, J. (2012a). A new species in the genus *Nomocharis* Franchet (Liliaceae): evidence that brings the genus *Nomocharis* into *Lilium*. *Plant Syst. Evol.* 298, 69–85. doi: 10.1007/s00606-011-0524-1
- Gao, Y. D., Zhou, S. D., He, X. J., and Wan, J. (2012b). Chromosome diversity and evolution in tribe Liliae (Liliaceae) with emphasis on Chinese species. *J. Plant Res.* 125, 55–69. doi: 10.1007/s10265-011-0422-1
- Gesch, D. B., Verdin, K. L., and Greenlee, S. K. (1999). New land surface digital elevation model covers the Earth. *Eos Trans. Am. Geophys. Union* 80, 69–70. doi: 10.1029/99eo00050
- Grant, V. (1981). *Plant Speciation*. New York, NY: Columbia University Press.
- Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process*. New York, NY: Oxford University Press.
- Harrison, R. G., and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *J. Hered.* 105(Suppl. 1), 795–809. doi: 10.1093/jhered/esu033
- Hatfield, J. S., Myrick, K. E., Huston, M. A., Weckerly, F. W., and Green, M. C. (2013). *Vascular Plant and Vertebrate Species Richness in National Parks of the Eastern United States*. Report No. 2013/002. Washington, DC: National Capital Region Office.
- Hijmans, R., Cameron, S., Parra, J., Jones, P., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276
- Hijmans, R. J., van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J., Lamigueiro, O., et al. (2013). *Raster Package for R. version*.
- Hoekstra, J. M., Molnar, J. L., Jennings, M., Revenga, C., Spalding, M. D., Boucher, T. M., et al. (2010). *The Atlas of Global Conservation: Changes, Challenges, and Opportunities to Make a Difference*, ed. J. L. Molnar (Berkeley, CA: University of California Press), 272.
- Hooper, D. M., Griffith, S. C., and Price, T. D. (2019). Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Mol. Ecol.* 28, 1246–1262. doi: 10.1111/mec.14874
- Horning, M. E., and Webster, M. S. (2009). Conservation genetics of remnant *Lilium philadelphicum* populations in the Midwestern United States. *Am. Midl. Nat.* 161, 286–301. doi: 10.1674/0003-0031-161.2.286
- Hull, S. L., Grahame, J., and Mill, P. J. (1996). Morphological divergence and evidence for reproductive isolation in *Littorina saxatilis* (Oliva) in northeast England. *J. Molluscan Stud.* 62, 89–99. doi: 10.1093/mollus/62.1.89
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harris, R., and Lawrence, W. T. (2004). Global patterns in human consumption of net primary production. *Nature* 429, 870–873. doi: 10.1038/nature02619
- Jackson, S. T., Betancourt, J. L., Booth, R. K., and Gray, S. T. (2009). Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19685–19692. doi: 10.1073/pnas.0901644106
- Jenkins, C. N., Pimm, S. L., and Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2602–E2610.
- Jiang, S., Luo, M.-X., Gao, R.-H., Zhang, W., Yang, Y.-Z., Li, Y.-J., et al. (2019). Isolation-by-environment as a driver of genetic differentiation among populations of the only broad-leaved evergreen shrub *Ammopiptanthus mongolicus* in Asian temperate deserts. *Sci. Rep.* 9:12008.
- Kassambara, A., and Mundt, L. (2017). *Factoextra v1.0.5. Extract and Visualize the Results of Multivariate Data Analyses*. Available online at: <https://www.rdocumentation.org/packages/factoextra/versions/1.0.5> (accessed August 30, 2019).
- Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Krefl, H., et al. (2005). Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* 32, 1107–1116. doi: 10.1111/j.1365-2699.2005.01272.x
- Koch, M. A., Dobes, C., and Mitchell-Olds, T. (2003). Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Mol. Biol. Evol.* 20, 338–350. doi: 10.1093/molbev/msg046

- Kraft, N. J., Adler, P. B., Godoy, O., James, E. C., Fuller, S., and Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29, 592–599. doi: 10.1111/1365-2435.12345
- Larson, E. L., White, T. A., Ross, C. L., and Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Mol. Ecol.* 23, 1668–1678. doi: 10.1111/mec.12601
- Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., and Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Mol. Ecol.* 26, 554–570. doi: 10.1111/mec.13933
- Lexer, C., Fay, M. F., Joseph, J. A., Nica, M. S., and Heinze, B. (2005). Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol. Ecol.* 14, 1045–1057. doi: 10.1111/j.1365-294x.2005.02469.x
- Liang, S. Y. (1995). Chorology of Liliaceae (*s. str.*) and its bearing on the Chinese flora. *Acta Phytotaxonomica Sinica* 33, 27–51.
- Liang, S. Y., and Tamura, M. (2000). “*Lilium L.*” in *Flora of China*, eds Z. Y. Wu and P. H. Raven (Beijing: Science Press & Missouri Botanical Garden Press), 135–161.
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Liu, Y., Harris, A., Gao, Q., Su, X., and Ren, Z. (2019). A population genetics perspective on the evolutionary histories of three clonal, endemic, and dominant grass species of the Qinghai–Tibet Plateau: *Orinus* (Poaceae). *Ecol. Evol.* 9, 6014–6037. doi: 10.1002/ecs3.5186
- Ma, T., Wang, K., Hu, Q. J., Xi, Z. X., Wan, D. S., Wang, Q., et al. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc. Natl. Acad. Sci. U.S.A.* 115, E236–E243.
- Maechler, M., Struyf, A., Hubert, M., and Hornick, K. (2009). *Cluster analysis Extended Rousseeuw et al.* Available online at: <https://CRAN.R-project.org/package=cluster> (accessed August 27, 2019).
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010
- Mallet, J., Besansky, N., and Hahn, M. W. (2016). How reticulated are species? *Bioessays* 38, 140–149. doi: 10.1002/bies.201500149
- Martin, N. H., Bouck, A. C., and Arnold, M. L. (2007). The genetic architecture of reproductive isolation in Louisiana irises: flowering phenology. *Genetics* 175, 1803–1812. doi: 10.1534/genetics.106.068338
- Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi: 10.1093/molbev/msu269
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES science gateway for inference of large phylogenetic trees,” in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8.
- Minami, S., and Azuma, A. (2003). Various flying modes of wind-dispersal seeds. *J. Theor. Biol.* 225, 1–14. doi: 10.1016/s0022-5193(03)00216-9
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.
- Nevado, B., Harris, S., Beaumont, M., and Hiscock, S. J. (2020). Rapid homoploid hybrid speciation in British gardens: the origin of Oxford ragwort (*Senecio squalidus*). *Mol. Ecol.* 29, 4221–4233. doi: 10.1111/mec.15630
- Noble, I. R., and Slatyer, R. (1980). The use of vital attributes to predict successional changes in plant communities subject to recurrent disturbances. *Vegetatio* 43, 5–21. doi: 10.1007/978-94-009-9200-9\_2
- Ortiz, D. A., Lima, A. P., and Werneck, F. P. (2018). Environmental transition zone and rivers shape intraspecific population structure and genetic diversity of an Amazonian rain forest tree frog. *Evol. Ecol.* 32, 359–378. doi: 10.1007/s10682-018-9939-2
- Peakall, R., and Smouse, P. E. (2012). GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Peruzzi, L., Leitch, I. J., and Caparelli, K. F. (2009). Chromosome diversity and evolution in Liliaceae. *Ann. Bot.* 103, 459–475. doi: 10.1093/aob/mcn230
- Peruzzi, L., Peterson, A., Tison, J.-M., and Harpke, D. (2011). New light on phylogeny and taxonomy of the Eurasian *Gagea villosa*–*G. fragifera* complex (Liliaceae). *Nord. J. Bot.* 29, 722–733. doi: 10.1111/j.1756-1051.2011.01187.x
- Peterson, A., Harpke, D., Peruzzi, L., Levichev, I. G., Tison, J. M., and Peterson, J. (2009). Hybridization drives speciation in *Gagea* (Liliaceae). *Plant Syst. Evol.* 278, 133–148. doi: 10.1007/s00606-008-0102-3
- Peterson, A., Levichev, I. G., Peterson, J., Harpke, D., and Schnittler, M. (2011). New insights into the phylogeny and taxonomy of Chinese species of *Gagea* (Liliaceae)—speciation through hybridization. *Org. Divers. Evol.* 11, 387–407. doi: 10.1007/s13127-011-0059-x
- Phillips, S. J., and Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175. doi: 10.1111/j.0906-7590.2008.5203.x
- Phillips, S. J., Dudík, M., and Schapire, R. E. (2004). “A maximum entropy approach to species distribution modeling,” in *Proceedings of the Twenty-First International Conference*, Banff, 655–662.
- Pons, O., and Petit, R. J. (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144, 1237–1245.
- Pritchard, J. K., Wen, X., and Falush, D. (2009). *Documentation for Structure Software: version 2.3*. Available online at: [http://pritch.bsd.uchicago.edu/structure\\_software/release\\_versions/v2.3.2/structure\\_doc.pdf](http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.2/structure_doc.pdf) (accessed September 10, 2018).
- Raxworthy, C. J., Ingram, C. M., Rabibisoa, N., and Pearson, R. G. (2007). Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Syst. Biol.* 56, 907–923. doi: 10.1080/10635150701775111
- Raymond, M., and Rousset, F. (1995). Genepop (version-1.2) - population-genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249. doi: 10.1093/oxfordjournals.jhered.a111573
- Rieseberg, L. H., Baird, S. J., and Gardner, K. A. (2000). Hybridization, introgression, and linkage evolution. *Plant Mol. Evol.* 42, 205–224. doi: 10.1007/978-94-011-4221-2\_11
- Scott, P. A., Glenn, T. C., and Rissler, L. J. (2019). Formation of a recent hybrid zone offers insight into the geographic puzzle and maintenance of species boundaries in musk turtles. *Mol. Ecol.* 28, 761–771. doi: 10.1111/mec.14983
- Serna-Chavez, H. M., Fierer, N., and Van Bodegom, P. M. (2013). Global drivers and patterns of microbial abundance in soil. *Glob. Ecol. Biogeogr.* 22, 1162–1172. doi: 10.1111/geb.12070
- Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms, the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275
- Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Streisfeld, M. A., and Kohn, J. R. (2005). Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59, 2548–2559. doi: 10.1554/05-514.1
- Sun, Y., Lu, Z., Zhu, X., and Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat. Commun.* 11:3375.
- Swinger, K. K., Lemberg, K. M., Zhang, Y., and Rice, P. A. (2003). Flexible DNA bending in HU–DNA cocrystal structures. *EMBO J.* 22, 3749–3760. doi: 10.1093/emboj/cdg351
- Swofford, D. L. (2003). *PAUP\* Phylogenetic Analysis Using Parsimony (\* and other Methods)*. Sunderland, MA: Sinauer Associates.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4, molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Taylor, S. A., and Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3, 170–177. doi: 10.1038/s41559-018-0777-y
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL\_X windows interface, Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Tiley, G. P., and Burleigh, J. G. (2015). The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evol. Biol.* 15:194. doi: 10.1186/s12862-015-0473-3
- Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hubner, S., et al. (2016). Hybridization and extinction. *Evol. Appl.* 9, 892–908.

- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- Yan, L. J., Burgess, K. S., Zheng, W., Tao, Z. B., Li, D. Z., and Gao, L. M. (2019). Incomplete reproductive isolation between *Rhododendron* taxa enables hybrid formation and persistence. *J. Integr. Plant Biol.* 61, 433–448.
- Yuan, S. X., Ge, L., Liu, C., and Ming, J. (2013). The development of EST-SSR markers in *Lilium regale* and their cross-amplification in related species. *Euphytica* 189, 393–419.
- Zimmer, E., Martin, S., Beverley, S., Kan, Y., and Wilson, A. C. (1980). Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 77, 2158–2162.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JL declared a past co-authorship with one of the authors XG to the handling editor.

Copyright © 2020 Gao, Harris, Li and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.