

# Tertiary Origin and Pleistocene Diversification of Dragon Blood Tree (*Dracaena cambodiana*-Asparagaceae) Populations in the Asian Tropical Forests

Jian-Li Zhao<sup>1,2,3</sup>, Lu Zhang<sup>1,2,3</sup>, Selvadurai Dayanandan<sup>3</sup>, Shivaprakash Nagaraju<sup>4</sup>, Dong-Mei Liu<sup>1,2</sup>, Qiao-Ming Li<sup>1,2\*</sup>

**1** Laboratory of Plant Phylogenetics and Conservation, Xishuangbanna Tropical Botanical Garden, The Chinese Academy of Sciences, Kunming, Yunnan, People's Republic of China, **2** Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, The Chinese Academy of Sciences, Menglun, Yunnan, People's Republic of China, **3** The Graduate University of The Chinese Academy of Sciences, Beijing, People's Republic of China, **4** Department of Biology, Concordia University, Montreal, Quebec, Canada

## Abstract

**Background:** The origin of extraordinarily rich biodiversity in tropical forests is often attributed to evolution under stable climatic conditions over a long period or to climatic fluctuations during the recent Quaternary period. Here, we test these two hypotheses using *Dracaena cambodiana*, a plant species distributed in paleotropical forests.

**Methods:** We analyzed nucleotide sequence data of two chloroplast DNA (cpDNA: *atpB-rbcL* and *trnD-trnT*) regions and genotype data of six nuclear microsatellites from 15 populations (140 and 363 individuals, respectively) distributed in Indochina Peninsular and Hainan Island to infer the patterns of genetic diversity and phylogeographic structure. The population bottleneck and genetic drift were estimated based upon nuclear microsatellites data using the software programs BOTTLENECK and 2MOD. The lineage divergence times and past population dynamics based on cpDNA data were estimated using coalescent-based isolation-with-migration (IMa) and BEAST software programs.

**Results:** A significant phylogeographic structure ( $N_{ST}=0.876$ ,  $G_{ST}=0.796$ ,  $F_{ST-SSR}=0.329$ ,  $R_{ST}=0.449$ ;  $N_{ST}>G_{ST}$ ,  $R_{ST}>F_{ST-SSR}$ ,  $P<0.05$ ) and genetic differentiation among populations were detected. Bottleneck analyses and Bayesian skyline plot suggested recent population reduction. The cpDNA haplotype network revealed the ancestral populations from the southern Indochina region expanded to northward. The most recent ancestor divergence time of *D. cambodiana* dated back to the Tertiary era and rapid diversification of terminal lineages corresponded to the Quaternary period.

**Conclusions:** The results indicated that the present distribution of genetic diversity in *D. cambodiana* was an outcome of Tertiary dispersal and rapid divergence during the Quaternary period under limited gene flow influenced by the uplift of Himalayan-Tibetan Plateau and Quaternary climatic fluctuations respectively. Evolutionary processes, such as extinction-recolonization during the Pleistocene may have contributed to the fast diversification in *D. cambodiana*.

**Citation:** Zhao J-L, Zhang L, Dayanandan S, Nagaraju S, Liu D-M, et al. (2013) Tertiary Origin and Pleistocene Diversification of Dragon Blood Tree (*Dracaena cambodiana*-Asparagaceae) Populations in the Asian Tropical Forests. PLoS ONE 8(4): e60102. doi:10.1371/journal.pone.0060102

**Editor:** Giovanni G Vendramin, CNR, Italy

**Received:** August 24, 2012; **Accepted:** February 25, 2013; **Published:** April 1, 2013

**Copyright:** © 2013 Zhao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by the funds from the Chinese Academy of Sciences (CAS) for Key Topics in Innovation Engineering (grant number KSCX2-YW-Z-002) and from the CAS 135 program (XTBG-T01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: lqm@xtbg.ac.cn

These authors contributed equally to this work.

## Introduction

Understanding the origin of extraordinary biological diversity in tropical forests remains as one of the greatest challenges in evolutionary biology. In general, two contrasting hypotheses have been put forward to explain the origin of high biological diversity in tropical ecosystems. The persistence hypothesis stipulates that the gradual evolution during the long-term survival of organisms and low extinction rates under stable climatic conditions may have lead to the present day high biological diversity in the tropics. Alternatively, the refugia hypothesis suggests that the current diversity in the tropics could be attributable to successive isolation

and subsequent expansion of populations in response to frequent oscillations of the climate during the Quaternary period [1,2,3,4,5,6]. Thus, the persistence hypothesis suggests that the origin of high biodiversity in the tropics should date back beyond the Quaternary period, whereas the refugia hypothesis stipulates that the biological diversification in the tropics is more recent and coincides with the Quaternary climatic fluctuations. Numerous studies suggest that the high biological diversity in the tropics may have originated during the pre-Quaternary era [1,2,4,6,7]. However, several examples, particularly from the Neotropics, support the refugia hypothesis suggesting a recent diversification

during the Quaternary period [1,2,7,8]. In contrast to the Neotropics, the evolution of biological diversity in the Asian Paleotropics may have been influenced by major geological events including the collision of the Deccan plate with Eurasia and associated topographical alterations during the Tertiary [6,9,10,11,12,13]. Our present understanding of the post-glacial time evolution of plants and animals is mostly based on studies focusing on temperate and sub-tropical organisms, and limited data are available on the post-glacial evolution in paleotropical ecosystems.

The limited understanding of the evolution of tropical plants during the Quaternary period could be attributable to paucity of reliable data from the tropical regions. The fossil records of vascular plants in the tropics representing the Quaternary period are rare [14] and pollen data may not accurately identify the responses of plants to changes in the climate during the Quaternary period [15,16]. As a means of overcoming these limitations, DNA based molecular genetic approaches have become an integral tool in understanding the genetic diversification and historical plant demographics. DNA-based genomic data of extant species are being used to infer genetic diversification and biogeographical history over the past several million years and provide a basis for inferring historical population demographics [17]. In particular, molecular dating based on chloroplast DNA (cpDNA) sequences and nuclear microsatellites provides a unique opportunity to explore the historical demography and chronology of evolutionary changes [15,16,18,19,20]. In Angiosperms, cpDNA is generally maternally inherited and dispersed through seeds, and therefore cpDNA markers provide a means to study seed-mediated species migration patterns [14,21]. The biparentally nuclear microsatellite (nSSR) data provide a means to investigate fine-scale genetic structure attributable to the combined effects of both seed and pollen flow.

*Dracaena cambodiana* Pierre ex Gagnepain (Asperagaceae), commonly known as the dragon blood tree is a vulnerable species and generally distributed in the escarpments of island-like limestone mountains in paleotropical Southeastern Asia. Current dragon trees mainly distributed in Asia and Africa [22,23]. *D. cambodiana* was the representative of dragon trees in Asia. Dispersal of tropical elements between Asia and Africa has two possible ways [6]. One was megathermal rainforest expansion during warm Paleocene and Miocene. The other one was Indian plate collision to Eurasian land during Miocene. Whatever the ways for *D. cambodiana*, this species should exist in Asian Paleotropics during Tertiary. Thus, *D. cambodiana* serves as an ideal model species to test the biodiversity-original hypotheses of plant species in Paleotropics. The objectives of this study are (i) to illustrate the genetic diversity, gene flow and past population demographics using cpDNA and nSSR; (ii) to determine the ancestral populations and past migration patterns through phylogeographic analysis of cpDNA; (iii) to date the genetic divergence under different molecular clock methods.

## Materials and Methods

### Ethics Statement

*Dracaena cambodiana* is national second class protected plants in China. This species is listed in the Inventory of Rare and Endangered Plants of China and the Key Protected Inventory of Wild Plants of China (<http://db.kib.ac.cn/eflora/View/plant/ZXBWSpecies.aspx>), but it is not evaluated by the International Union for Conservation of Nature (IUCN) and it is also not recognized as the endangered or protected species in Thailand, Cambodia, Laos and Vietnam. Furthermore, the locations in

Thailand, Cambodia, Laos and Vietnam are not privately-owned or protected in any way except the Pumat National Park of Vietnam. We got the permission of the Wildlife Protection and Administration Office under the Forestry Department of Yunnan, Guangxi and Hainan Province in China and the permission of the Administration of Pumat National Park of Vietnam. We also got the permission of local forestry department in other locations of Thailand, Cambodia, Laos and Vietnam. The sampling process was under the guidance of local rangers. This plant was solely used for scientific research and our sampling will not affect the regular growth of *D. cambodiana*.

### Sample Collection and DNA Extraction

Leaf samples of *D. cambodiana* plants from 15 localities representing four geographical regions covering most of the natural distribution range of the species in Asia (Figure 1a, Figure 1d, Table 1) were collected, and dried using silica gel. The total genomic DNA was extracted from about 100 mg of silica-gel-dried leaf materials following a CTAB based DNA extraction protocol [24].

### Molecular Markers

**Chloroplast DNA (cpDNA) sequences.** Ten individuals representing different geographic regions were surveyed for the nucleotide sequence variation at four cpDNA regions (*tmL-tmF*, *psbA-tmH*, *atpB-rbcL*, and *tmD-tmT*). Based upon sequence variability, *atpB-rbcL* [25] and *tmD-tmT* [26] regions were chosen for further analysis. These regions were PCR amplified, purified and sequenced (see online supplement for details). The DNA sequences were aligned using CLUSTALW [27] as implemented in MEGA V5 [28] and manually edited to improve the alignment. Insertions and deletions (indels) in the repeated regions were excluded from the analyses as these regions are known to evolve rapidly and prone to homoplasy [29,30]. A total of 140 individuals were sequenced and representative cpDNA haplotype sequences of *D. cambodiana* and outgroups were deposited in the GenBank (accessions: JF784389–JF784419).

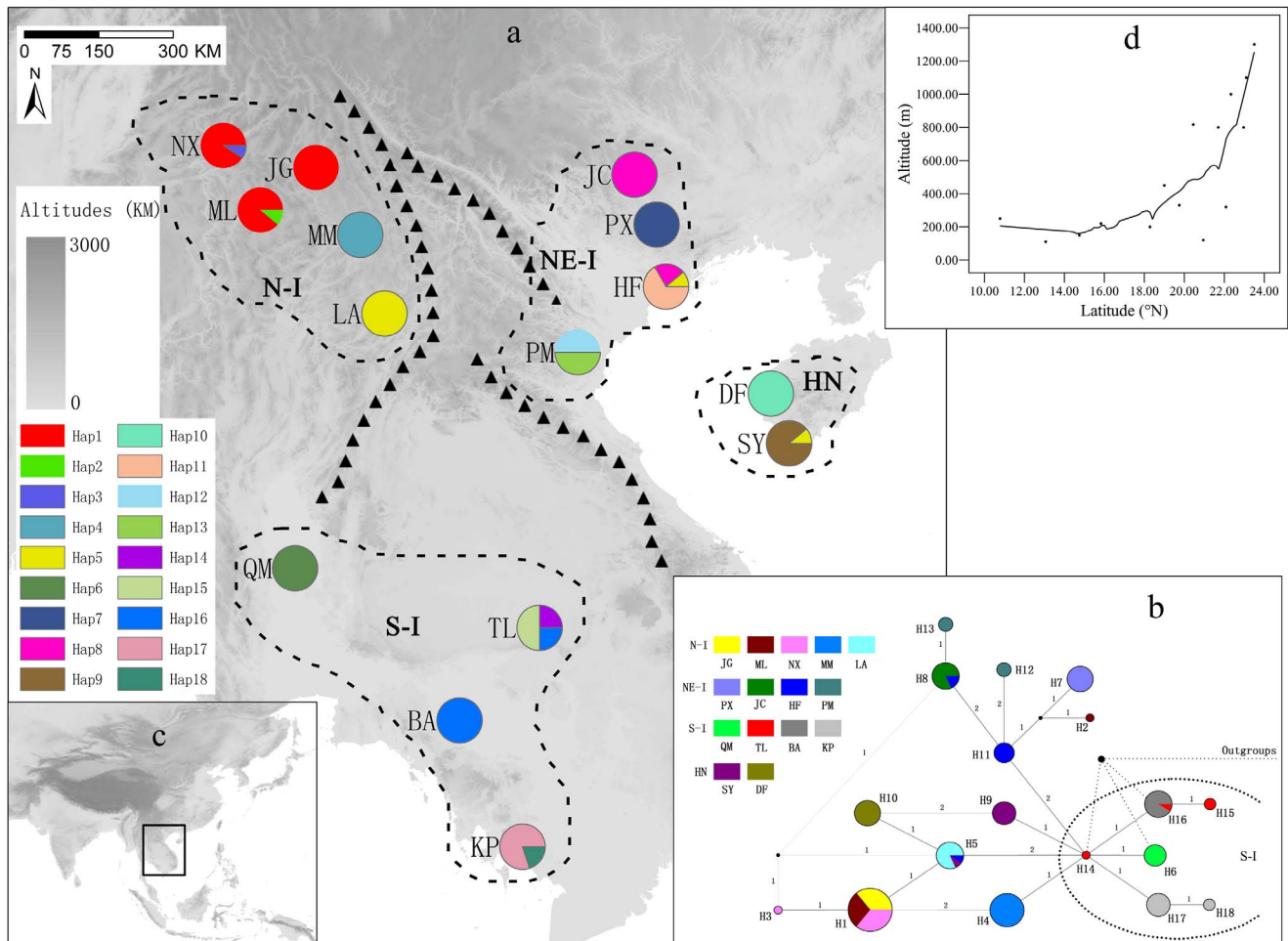
**Nuclear microsatellite genotyping.** Six polymorphic nuclear microsatellite loci, *DC003*, *DC006*, *DC140*, *DC437*, *DC460*, *DC501* were used for genotyping a total of 363 individuals following the protocol of [31].

### Genetic Analyses

**Diversity parameters.** The sampled *D. cambodiana* populations were clustered into four groups based upon their geographical distribution (Figure 1).

For cpDNA data, the haplotype diversity ( $h$ ) and nucleotide diversity ( $\pi$ ) were calculated for each group and overall level using the program DNASP V5.10 [32]. The total genetic diversity ( $H_T$ ) and within population genetic diversity ( $H_S$ ) were calculated using PERMUTCPSSR V2.0 [33]. The genetic differentiation ( $F_{ST-cp}$ ) among all fifteen populations and among four groups was calculated using ARLEQUIN V3.1 [34].

For microsatellite data, the presence of null alleles, large allele dropout and scoring errors due to stuttering were tested using MICRO-CHECKER V2.2.3 [35] and anomalies were rectified following Brookfield [36]. The heterozygosity (observed  $H_O$  and expected  $H_E$ ) and test for linkage-disequilibrium were calculated using GENEPOP V3.3 [37]. The deviations from Hardy-Weinberg equilibrium values were tested using FSTAT V2.9.3.2 [38] by calculating  $F_{IS}$  values for each population and each locus. The statistical significance of deviations was assessed by randomization for 1000 times. The proportion of genetic variation partitioned among populations and among groups of populations



**Figure 1. Geographic distribution and evolutionary relationships of 18 cpDNA haplotypes in *Dracaena cambodiana*.** **a:** Different colors in pie charts represent haplotypes. The dashed lines indicate the division of fifteen populations into four groups based upon their geographic location. Black triangles represent main geographical barriers originated during the Tertiary era (modified from Lacassin *et al.*, 1993; Huchon *et al.*, 1994; Searle, 2006). **b:** Colors represent populations. The black dots are the missing haplotypes or missing samples. The numbers on the branches represent number of mutations between two connected haplotypes. The gray connected lines among haplotypes were resolved based on the methods of Crandall & Templeton (1993). Segmented-elliptic-dashed lines represent the cluster of S-I populations. Straight-dashed line is the connection to outgroups. The detailed connections of outgroups is given in Figure S2. **c:** general position of map A in Southeast Asia. **d:** The trend of altitudes shifting along latitudes.  
doi:10.1371/journal.pone.0060102.g001

were quantified using analyses of molecular variance (AMOVA) as implemented in ARLEQUIN and the statistical significance was tested with 10000 permutations.

**Population differentiation and phylogeographic structure.** The presence of phylogeographic structure of cpDNA was tested by comparing the level of population differentiation based upon the infinite allele model ( $G_{ST}$ ) with an estimate of population subdivision based on phylogenetically ordered alleles ( $N_{ST}$ ) using the program PERMUTCPSSR. Because  $G_{ST}$  may be problematic for the interpretation of genetic differentiation, we employ the standardized genetic differentiation ( $G'_{ST}$ ) to measure genetic differentiation among populations according to the formula  $G'_{ST} = G_{ST}(1 + H_S)/(1 - H_S)$  [39].

Using nuclear microsatellite data, we assessed the genetic structure based on two differentiation indices namely  $F_{ST-SSR}$  [40] and  $R_{ST}$  [41]. The  $F_{ST-SSR}$  value is based on infinite allele model and  $R_{ST}$  estimate is based on the stepwise mutation model. The significant difference between these two estimates is an indication of the existence of phylogeographic structure. The genetic

structure was inferred using SPAGEDI software [42]. We tested for the presence of isolation-by-distance (IBD) by regression analysis between genetic differentiation  $F_{ST}/(1-F_{ST})$  and the natural logarithm of geographic distance for all pairs of populations [43]. The statistical significance of IBD was tested using the Mantel test with 1000 permutations as implemented in NTSYS V2.10e [44]. We tested the phylogeographic signal using nuclear microsatellite data following Hardy *et al.* [45] as implemented in SPAGEDI. This approach is based on comparison between  $F_{ST-SSR}$  and  $R_{ST}$  which are expected to be equal under the null hypothesis of no phylogeographic signal.  $G'_{ST}$  of nSSR was estimated by SMOGD [46].

**Detection of genetic clusters.** We used nuclear microsatellite data to test for the existence of cryptic population structure employing the Bayesian model based software program STRUCTURE V2.3.3 [47]. An admixture model with correlated allele frequencies was used for estimating the historical population admixture and number of genetic clusters ( $k$ ) ranging from 1 to 20 [48,49]. All analyses employing the STRUCTURE program were

**Table 1.** Sampling information of *Dracaena cambodiana* for cpDNA and nuclear microsatellite (nSSR) analyses.

Code	Population	Latitude (N)/Longitude (E)	Altitude (m)	N cpDNA/nSSR	cpDNAhaplotypes (no. of individuals)
<b>S-I Southern Indochina</b>				<b>31/84</b>	
QM	Chiang Mai, Thailand	15°50'/100°12'	220	7/–	H6(7)
TL	Ubon, Thailand	14°45'/104°38'	150	4/28	H14(1), H15(2), <b>H16(1)</b>
BA	Batdambang, Cambodia	13°04'/103°11'	110	10/28	<b>H16(10)</b>
KP	Kampot, Cambodia	10°47'/104°19'	250	10/28	H17(8), H18(2)
<b>NE-I Northeastern Indochina</b>				<b>34/78</b>	
PX	Pingxiang, Guangxi Province, China	22°05'/106°45'	320	10/24	H7(10)
JC	Jingxi, Guangxi Province, China	22°58'/106°21'	800	9/24	<b>H8(9)</b>
HF	Haiphong, Vietnam	20°57'/106°57'	120	9/30	<b>H5(1), H8(2), H11(6)</b>
PM	Pumat National Park, Vietnam	19°45'/105°19'	330	6/–	H12(3), H13(3)
<b>N-I Northern Indochina</b>				<b>56/156</b>	
JG	Jinggu, Yunnan Province, China	23°06'/100°35'	1100	10/23	<b>H1(10)</b>
ML	Menglian, Yunnan Province, China	22°20'/99°34'	1000	9/42	<b>H1(8), H2(1)</b>
NX	Zhenkang, Yunnan Province, China	23°30'/98°54'	1300	11/30	<b>H1(10), H3(1)</b>
MM	Mengyuan, Yunnan Province, China	21°42'/101°22'	800	17/30	H4(17)
LA	LouangPhrabang, Laos	20°27'/101°49'	817	9/31	<b>H5(9)</b>
<b>HN Hainan Island</b>				<b>19/45</b>	
DF	Dongfang, Hainan Province, China	19°00'/108°49'	450	10/24	H10(10)
SY	Sanya, Hainan Province, China	18°17'/109°09'	200	9/21	<b>H5(1), H9(8)</b>
Overall		–	–	<b>140/363</b>	H1-18 (140)

The shared haplotypes are indicated in ***bold-italic***.

(–) = no individuals were available for nuclear microsatellite analysis.

doi:10.1371/journal.pone.0060102.t001

performed with 20 replicates, each with a burnin period of 20 000 and a Markov chain Monte Carlo (MCMC) value set at 130000. These values were large enough to stabilize log (alpha), Ln likelihood and the MCMC chain converging to consistent end results [47]. The method of Evanno *et al.* [50] was used for finding the most likely value of  $k$  by plotting the log probability of  $L(k)$  and the  $\Delta k$  of the data over multiple runs using the program STRUCTURE HARVESTER [51]. In order to compare runs with the same value of  $k$ , we calculated the symmetric similarity coefficients (SSC) using the Greedy algorithm as implemented in the CLUMPP software [52]. The groups of runs with  $SSC \geq 0.8$  were combined and bar plots were prepared using the software program DISTRUCT [53]. The GenAlex program [54] was used for calculating pairwise genetic distances between individuals [55] and principal coordinate analysis (PCA).

**Haplotype network analyses.** The evolutionary network reconstruction and ancestral haplotype inference based on cpDNA data were carried out using the software program NETWORK V4.5.1.6 (www.fluxus-engineering.com) following the median-joining method [56]. The ambiguous loops were resolved following the method of Crandall & Templeton [57] assuming that rare haplotypes and widespread haplotypes are more likely located at the tip and interior respectively and singletons are more likely derived from the same population. We chose *Agava* sp., *Yucca gloriosa* and *Asparagus setacea* as outgroups for the evolutionary network analysis. Phylogenetically *Asparagus* is basal to *Agava* and *Yucca*, which are closely related to *Dracaena* [58]. Thus we used *Asparagus* as the basal outgroup to root the evolutionary network.

**Molecular dating and demographics of cpDNA.** We used the 'Isolation with migration' (IM) coalescent model as imple-

mented in the program IMA2 [59,60] to estimate the divergence times of four geographically distinct groups of populations. We calculated the divergence time,  $t_i$  (i as the node of diverging groups) and the time since most recent common ancestor (TMRCA) of the species and each group ( $t_{TMRCA}$ ). The resulting values were converted to absolute time scale,  $T$  in years using the formula  $T = t/\mu k$ , where  $\mu$  is the number of substitution per site per year (s/s/y) and  $k$  is the length of the sequence [61]. We estimated  $\mu$  as  $0.146476 \pm 0.063280 \times 10^{-9}$  s/s/y based on uncorrelated exponential clock model (Table S1 and File S1), and the value of  $k$  is 1814 bp. The newick format of a tree for IMA2 based on pairwise  $F_{ST}$  values among four groups (Table S2) was constructed using NTSYS software package. The haplotype in Southern Indochina (S-I), which was found to be ancestral based upon haplotype network analysis was used for rooting the haplotype network.

The departure from population demographic equilibrium was assessed using Tajima's  $D$  [62] and Fu's  $F_s$  [63] statistics using the ARLEQUIN software. The past population dynamics and the estimates of most recent common ancestor or  $T_{TMRCA}$  were further analyzed through Bayesian skyline plot [64] as implemented in the program BEAST V1.5.4 [65]. The Bayesian skyline reconstruction and lineages through time analysis were conducted using TRACER V1.5 [66] with burn-in of 10% of chains. The details of IMA2 and BEAST analysis procedure are given in File S1. The evidence for population demographic growth was investigated through mismatch distribution analysis using the software program ARLEQUIN. The pair-wise mismatch distribution is expected to be multimodal in samples drawn from populations at demographic equilibrium and unimodal in populations that have undergone a recent demographic expansion

[67,68] or a range expansion with high levels of migration between neighboring populations [69]. The goodness-of-fit of mismatch distributions to the expected distribution under a sudden expansion model [68] was tested using the sum of squared deviations (SSD) and the raggedness index [HRag; 70].

**Population bottleneck analyses of nSSR.** The excess in heterozygosity ( $H_c$ ) as compared to the expected heterozygosity under mutation drift equilibrium ( $H_{eq}$ ) for a given allelic diversity is an indication of recent reduction in the effective population size [71,72]. We used the software program BOTTLENECK [73] to assess any significant excess in heterozygosity ( $H_c > H_{eq}$ ) through Wilcoxon sign-rank test and sign test with 5000 replications. We performed the analysis under Infinite allele model (IAM) and two-phase model of microsatellite evolution with the proportion of single step stepwise mutation (SMM) set to 70% and the variance to 30%. The population bottleneck is also expected to change the allele frequency distribution [74]. The allele frequency mode-shift indicator test was performed to test for significant departure in allele frequency distribution as compared to equilibrium expectations.

The population divergence under genetic drift alone or the balance between gene flow and genetic drift was assessed using the software program 2MOD V0.2 [http://www.rubic.rdg.ac.uk/~mab/software.html]. The genetic drift model assumes that fragmented populations are diverging by drift alone, and the gene flow–drift model assumes population allele frequencies are determined by the balance between genetic drift and gene flow. Two independent runs of 500000 iterations were performed and the initial 10% of the iterations were used as the burn-in. The Bayesian factors were calculated using TRACER to test the relative contribution of drift and gene flow to the population divergence.

## Results

### Genetic Diversity

The length of aligned consensus sequences of cpDNA fragments, *atpB-rbcL* and *trnD-trnT* of 140 individuals representing 15 *D. cambodiana* populations were 858 bp and 990 bp respectively. The *atpB-rbcL* region included two 1bp indels and the *trnD-trnT* region included one 8 bp (only in one individual of the population TL), one 17 bp, one 5 bp and two 1bp indels. After excluding gaps, the length of combined cpDNA consensus sequence length was 1814 bp, which included 15 polymorphic sites and 18 haplotypes (Table 1; Figure 1).

Overall, only 4 (H1, H5 H8, H16) of the 18 haplotypes were shared among populations (Table 1; Figure 1). Only one haplotype (H5) was shared among three populations (HF, LA, SY) in different groups (NE-I, N-I, HN). The other three haplotypes were shared between two or three populations within groups (H1: NX, ML, JG; H8: JC, HF; H16: TL, BA). The remaining 14 haplotypes were fixed within each population. The haplotype diversity and nucleotide diversity of cpDNA were 0.911 and  $1.81 \times 10^{-3}$  respectively (Table 2). The total genetic diversity was 0.968. The genetic diversity within population was 0.198. Observed heterozygosity ( $H_o$ ) estimated from nSSR was from 0.474 to 0.706 (average was 0.637). Expected heterozygosity ( $H_e$ ) estimated from nSSR was from 0.747 to 0.919 (average was 0.948).

### Population Differentiation and Phylogeographic Structure

The hierarchical AMOVA based upon both cpDNA and microsatellites data (Table 2 and Table 3) revealed significant species level genetic structure ( $F_{ST-cp} = 0.850$ ,  $P < 0.05$ ;

$F_{ST-SSR} = 0.329$ ,  $P < 0.01$ ) and within-group population divergence ( $F_{SC-cp} = 0.833$ ,  $P < 0.01$ ;  $F_{SC-SSR} = 0.322$ ,  $P < 0.01$ ). Although cpDNA data revealed a significant genetic differentiation among four groups ( $F_{CT-cp} = 0.104$ ,  $P < 0.01$ ), microsatellite data showed no significant genetic differentiation among groups ( $F_{CT-SSR} = 0.001$ ,  $P > 0.05$ ). The gene flow ( $Nm$ ) among populations based on cpDNA and microsatellite data was 0.0882 and 0.5099 respectively. The range-wide isolation by distance (IBD) in *D. cambodiana* was statistically non-significant ( $r_{cp} = 0.027$   $P = 0.437$ ;  $r_{SSR} = 0.199$   $P = 0.084$ ). A significant phylogeographic structure based on both cpDNA and microsatellites data was detected ( $G_{ST-cp} = 0.796$ ,  $N_{ST-cp} = 0.876$ ,  $N_{ST-cp} > G_{ST-cp}$ ,  $P < 0.05$ ;  $F_{ST-SSR} = 0.329$ ,  $R_{ST-SSR} = 0.449$ ,  $R_{ST-SSR} > F_{ST-SSR}$ ,  $P < 0.05$ ; Table 2).  $G_{ST}$  of cpDNA was approached to 1 and  $G_{ST}$  of SSR was 0.921.

### Genetic Clusters

In the STRUCTURE analysis, the posterior probability (LnP(D)) gradually increased reaching the highest value corresponding to  $K = 9$  (Figure 2A). Similarly, the  $\Delta K$  value also showed the highest peak at  $K = 9$  (Figure 2A). Although few peaks of LnP(D) at  $K$  values of 12, 15 and 18 were present, we chose  $K = 9$  as the optimal number of genetic clusters. The value of  $K = 9$  corresponded to the highest peak of LnP(D) at the lowest  $K$  value and the highest  $\Delta K$  value, the criteria for choosing the optimal  $K$  value as recommended in the documentation of STRUCTURE software (http://pritch.bsd.uchicago.edu/software). The occurrence of nine genetic clusters ( $K = 9$ ), a value higher than the geographically defined four sampling regions suggests a high genetic differentiation among populations than among regions. The bar-plot depicting ancestry of samples (Figure 2B) further indicated that the populations from Northeastern Indochina were more or less genetically similar to Northern Indochina and represents genetically diverse group. The populations in the Hainan Island and Southern Indochina showed complex genetic admixture with membership in more than one cluster. The population in the Hainan Island represents a genetically distinct group as compared to other populations analyzed in the present study.

### Haplotype Network Analysis

The rooted phylogenetic network analysis of 18 cpDNA haplotypes (Figure 1b and Figure S2) showed that only haplotypes from the S-I group connected to the outgroups. The haplotype H14 showed most connections with other haplotypes suggesting haplotypes in S-I are derived from haplotype H14. This suggests that the S-I group is ancestral [75]. The haplotype networks of other three groups are complex and formed several loops. The connections among H3, H5, and H8 were resolved following the methods of [57]. With the exception of S-I group, the phylogenetic clustering of other three groups was not congruent with their geographic locations. The relationships among haplotypes H1, H4, H9 and H10 were clearly interpretable based on their geographic location after resolving the loops following the minimum evolution theory. Most haplotypes in the Northeastern group (NE-I) have derived from H11 and one haplotype, (H5) has derived from H14. Haplotypes in the Northern Indochina group (N-I) have derived from three ancestral haplotypes: H1 from H5, H2 from H11 and H4 from H14.

### Molecular Dating and Historical Demography

The coalescent times of the most recent common ancestor ( $T_{TMRCA}$ ) based on the results of IMA2 and BEAST program were 11.48 (5.46–22.39) million years ago (Ma) and 9.77 (3.69–18.64)

**Table 2.** Genetic diversity and phylogeographic structure estimated from cpDNA and nSSR data of *Dracaena cambodiana*.

Group	cpDNA						nSSR				
	$H_S$	$H_T$	$h$	$\pi \times 10^{-3}$	$G_{ST}$	$N_{ST}$	$H_O$	$H_E$	$F_{ST-SSR}$	$R_{ST}$	$F_{IS}$
S-I	0.297	0.958	0.772	0.86	0.690	0.807	0.580	0.856	0.343	0.321	0.080
NE-I	0.289	0.963	0.784	1.35	0.700	0.732	0.706	0.890	0.298	0.438	0.011
N-I	0.081	0.739	0.643	0.72	0.891	0.817	0.678	0.919	0.324	0.406	0.047
HN	–	–	0.573	0.63	–	–	0.474	0.747	0.395	0.701	0.170
Overall	0.198	0.968	0.911	1.81	0.796	0.876*	0.637	0.948	0.329	0.449*	0.050

Genetic diversity for each loci and each population was summarize in Table S3 and Table S4.

$H_S$ , genetic diversity within populations;  $H_O$ , observed heterozygosity;  $H_E$ , expected heterozygosity;  $H_T$ , total genetic diversity;  $h$ , haplotype diversity;  $\pi$ , nucleotide diversity;  $G_{ST}$ ,  $N_{ST}$ ,  $F_{ST-SSR}$  and  $R_{ST}$ , genetic differentiation for and phylogeographic signal test; \*,  $N_{ST} > G_{ST}$ ,  $F_{ST-SSR} > R_{ST}$ ;  $P < 0.05$ , means significant phylogeographic structure;  $F_{IS}$ , inbreeding coefficient; –, Populations that are smaller than two cannot be analyzed in the PERMUTCPSSR program.

doi:10.1371/journal.pone.0060102.t002

Ma respectively (Figure 3). The  $T_{TMRC}$  of the 18 cpDNA haplotypes was 15.80 (5.29–29.88) Ma. The posterior probability distributions of all parameters showed a high degree of convergence (Figure S1). These results suggest that the origin of *D. cambodiana* dates back to Tertiary period from Middle Oligocene to Miocene. Although coalescent times of four groups dates back to Tertiary, the terminal lineages show more recent divergence dating back to the Quaternary period (Figure 4c, Figure S3, and Figure S4).

The mismatch distribution was multimodal (Figure 4a), indicating a demographic equilibrium. The SSD value (0.011,  $P < 0.05$ ; Table 4) and HRag value (0.053,  $P < 0.05$ ) also supports the demographic equilibrium. Moreover, both Tajima's  $D$  value (0.549,  $P > 0.05$ ) and Fu's  $F_S$  (–2.597,  $P > 0.05$ ) value also in agreement with the demographic equilibrium.

### Population Bottleneck Analysis

The population bottleneck analysis showed a mode-shift in allele frequency distribution in three populations (ML, MM and LA) from Northern Indochina (Table S5). These three populations (ML, MM and LA) along with two populations from Northeastern Indochina (PX and HF) and one population from Hainan Island (SY) showed significant deviation (Wilcoxon sign rank test,  $P < 0.05$ ; sign test,  $P < 0.05$ ) from mutation drift equilibrium (Table S5). Among these populations, two populations (ML and MM) showed strong evidence for recent population bottleneck with statistical significance in all three tests. The population divergence analysis using 2MOD software program highlighted genetic drift as a major evolutionary force in *D. cambodiana* ( $P_{\text{geneflow}} = 0.00014$ ; Bayes Factor<sub>geneticdrift vs. geneflow</sub> = 101.464).

## Discussion

### High Genetic Diversity and Limited Gene Flow

The genetic diversity estimates of *D. cambodiana* were higher than most high-altitude tropical, subtropical and temperate tree or shrub species in Southeast Asia [mean cpDNA  $H_T = 0.79$  in 76]. The diversity parameters ( $H_E$ ) and genetic differentiation ( $F_{ST}$ ) estimated using nuclear microsatellites were also high as compared to many tree species [77]. Plant species can survive during glacial periods in heterogeneous habitats [16] suggesting that the topologically heterogeneous mountains may have served as reservoirs of genetic diversity as evident in *Lithocarpus* sp. in tropical Asia [78]. A high genetic diversity observed in *D. cambodiana* is in agreement with the postulation that high genetic diversity could be attributable to long-term evolution of plant populations isolation in refugia under fluctuating climatic conditions [14,79,80].

The limited seed dispersal and low levels of gene flow may act as a driving force of divergence among populations [81]. The rareness of shared haplotypes and prevalence of private alleles within populations of *D. cambodiana* indicate that the gene flow through seeds between populations is limited despite possessing berry type fruits adapted for bird-mediated long-distance dispersal. The flower of *D. cambodiana* is entomophilous and therefore pollen flow among distant populations is unlikely. The significant population genetic differentiation and pronounced larger  $N_{ST}$  than  $G_{ST}$  is an indication of limited gene flow among populations. High  $G'_{ST}$  of cpDNA and nSSR support the strong population structure with limited gene flow among populations.

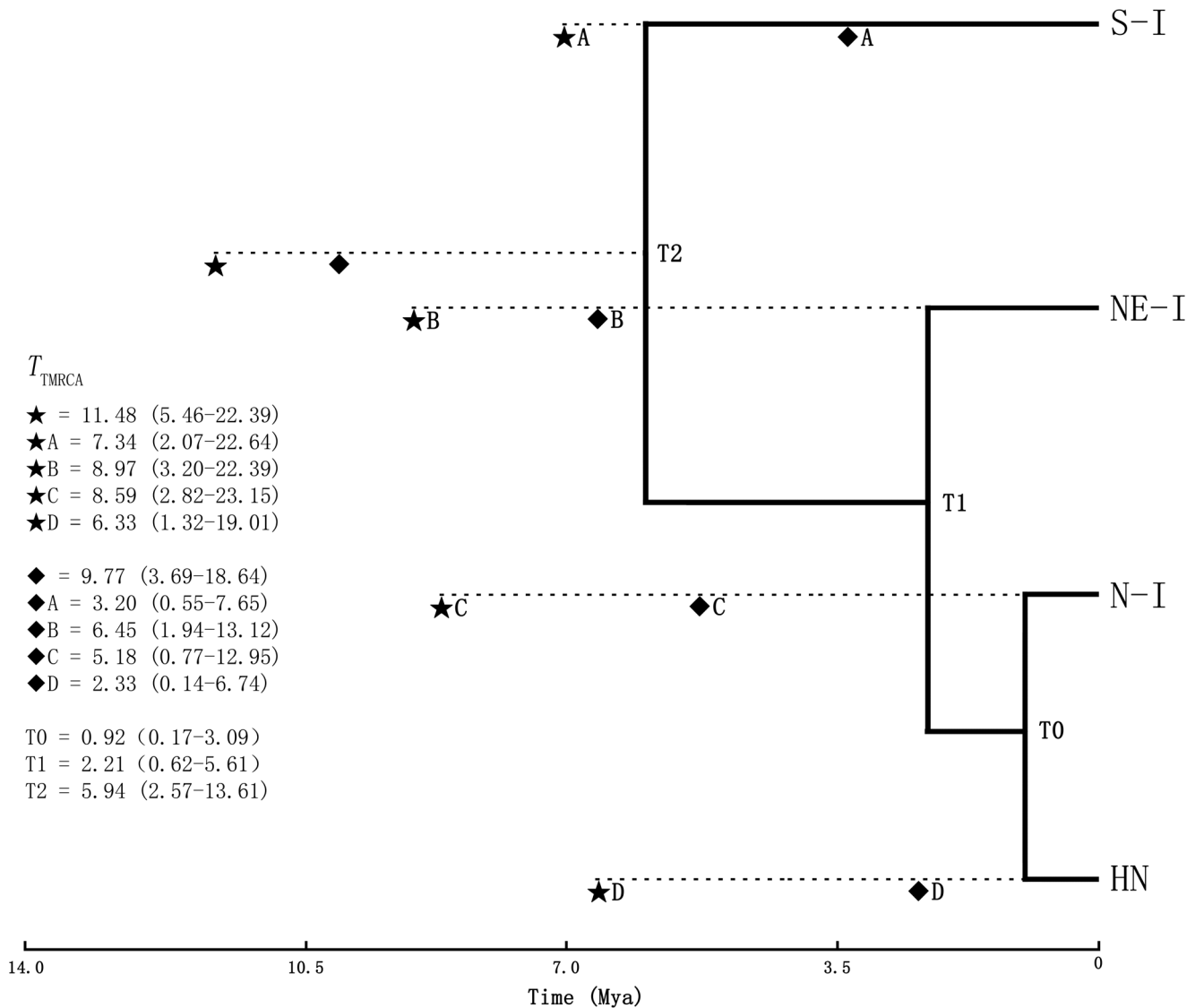
**Table 3.** Results of Analysis of Molecular Variance (AMOVA, cpDNA/nSSR) of *Dracaena cambodiana*.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Indices
Among groups	3/3	15.94/125.78	0.05/0.002	10.39/0.10	$F_{CT} = 0.104^*/F_{CT} = 0.001$
Among populations within groups	11/9	38.10/392.27	0.37/0.75	74.62/32.22	$F_{SC} = 0.833^*/F_{SC} = 0.322^*$
Within populations	125/350	9.26/549.79	0.07/0.01	15.00/0.22	$F_{ST-cp} = 0.850^*$ , $N_m = 0.0882/F_{ST-SSR} = 0.329^*$ , $N_m = 0.5099^{\#}$
Within individuals	–/363	–/566.50	–/1.56	–/67.46	

\* $P < 0.05$ ;

$\#$  for cpDNA,  $N_m = 0.5 (1 - F_{ST})/F_{ST}$ ; for microsatellite,  $N_m = 0.25 (1 - F_{ST})/F_{ST}$  (Wright 1931; Wright 1965).

doi:10.1371/journal.pone.0060102.t003



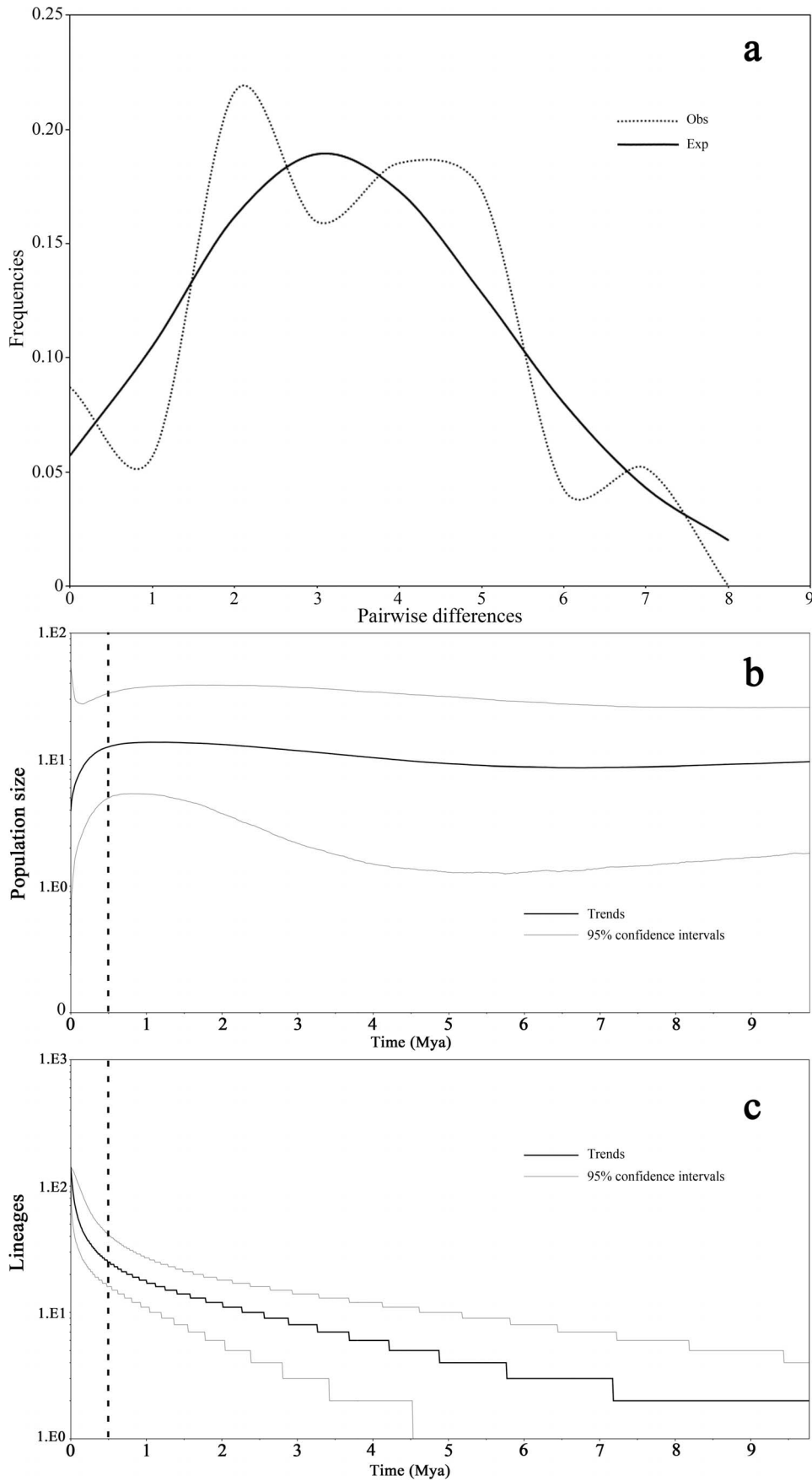
**Figure 2. Results of Bayesian model based clustering of individuals using the STRUCTURE program. a)** Distribution of likelihood value  $\ln P(D)$  and the distribution of model parameter  $\Delta K$ . All estimated values are based on 20 replicates and values of  $\ln P(D) (\pm SD)$  and  $\Delta K$  are plotted against its corresponding  $K$ . **b)** Barplots showing the results of the Bayesian cluster analysis. The colour in each barplot represents the probability of each individual to belong to an admixture group. Populations are ordered based on their population IDs as given in Table 1. doi:10.1371/journal.pone.0060102.g002

Although our cpDNA haplotype data showed a high genetic divergence between populations in Hainan Island and Indochina Peninsular, the microsatellite data provided insight into contemporary gene flow. The significant differences between corresponding  $F_{ST-SSR}$  and  $R_{ST-SSR}$  values indicated that population subdivision between Hainan Island and Indochina Peninsular reflect a phylogeographic structuring. Similarly,  $F$ -statistics and Bayesian model based clustering also revealed strong differentiation between populations in Hainan Island and Indochina Peninsular (South, North and Northeast) corroborating the cpDNA-based results and providing evidence for limited gene flow between populations. The low  $M_m$  value ( $<1$ ) also suggests that the limited migration among populations may not offset the divergence due to genetic drift [82]. The bottleneck analysis supported the genetic drift biased genetic differentiation confirming the low levels of among-population gene flow. The significant positive  $F_{IS}$  for some of the microsatellite loci is an indication that inbreeding occurs in

several populations [83]. Thus, strong genetic differentiation among *D. cambodiana* populations appears to be the result of the combined effect of distinctly limited gene flow and significant inbreeding within populations.

### Tertiary Origin and Rapid Divergence

Past geological events and associated climate change including the uplift of Himalayan-Tibetan Plateau and Quaternary glaciations [10,11,13] may have played an important role in the biological diversification, species extinction, speciation and distribution in the Asian tropics. The present topology of Asian mountains and the biogeographical boundaries in the Indochina Peninsula may have formed during the rapid uplift of Himalayan-Tibetan Plateau [11,84,85,86], erecting barriers for plant migration [9,12,18]. The present geographical distribution of *D. cambodiana* is closely linked to the topographical features of the Indochina Peninsula (Figure 1a), indicating genetic divergence of



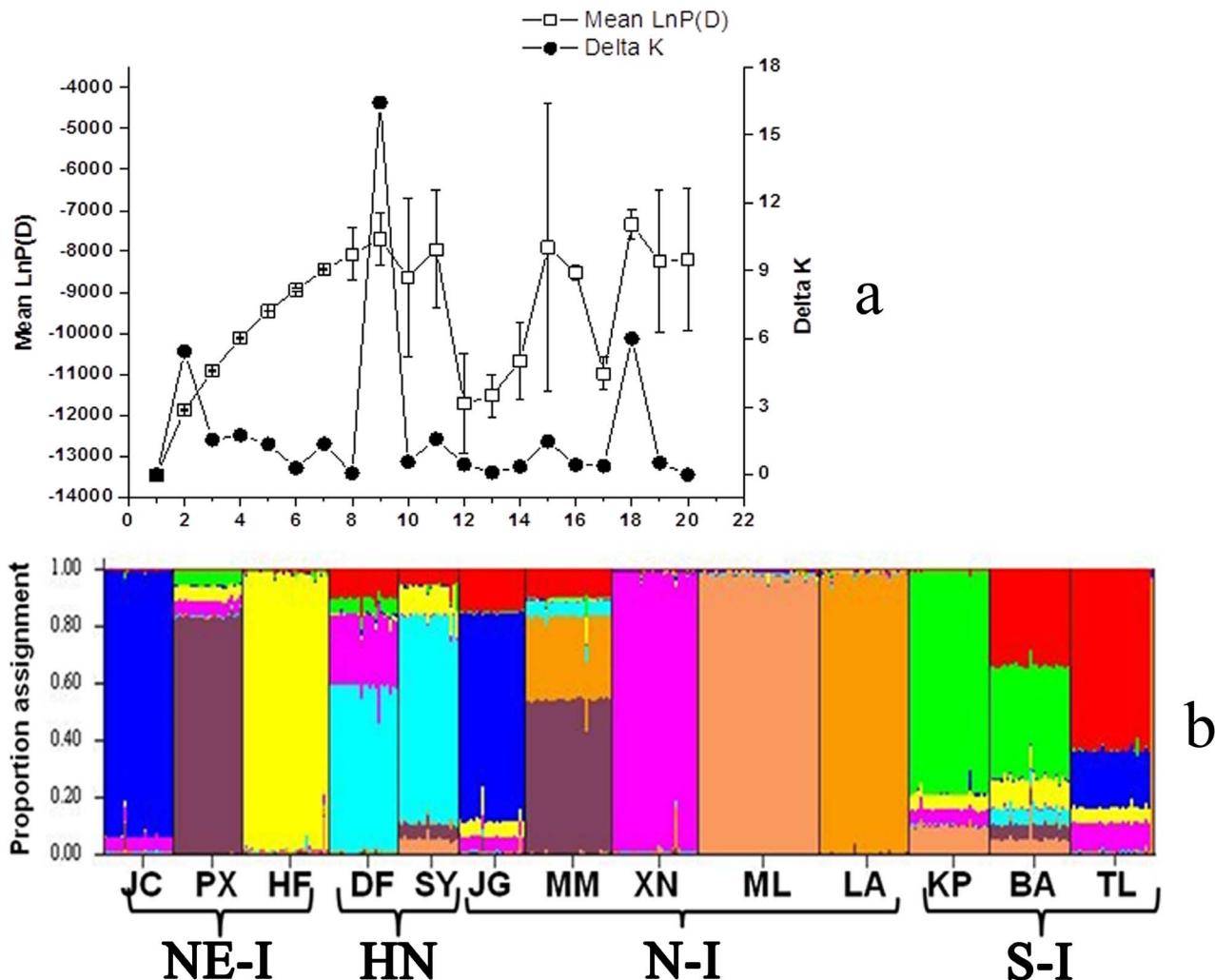


**Figure 3. The Phylogenetic tree representing evolutionary relationships of four geographically based groups of *Dracaena cambodiana* populations.** The coalescent times of the most recent common ancestor ( $T_{\text{TMRCA}}$ ) for overall populations and each group (A, B, C, D) based on IMA2 (★) and BEAST (◆) analyses. The divergence times among the four geographic groups are T0 T1 and T2. Dashed lines corresponds to  $T_{\text{TMRCA}}$ . The 95% highest posterior density intervals (95%HPD) are given in parentheses. The unit time is in million years ago (Ma). doi:10.1371/journal.pone.0060102.g003

*D. cambodiana* was significantly impacted by Tertiary extrusions of mountain. The mammalian fossils data of Tertiary origin also suggest that the middle Miocene environmental change and geological events related to the collision of the Deccan plate with Eurasia may have contributed to species extinction and speciation [87]. Although BEAST and IMA differ in the assumptions of genealogy samplers for coalescence process [88], these two coalescence-based methods yielded similar results of TMRCA (9.77–15.80 Ma), confirming the robustness of time estimates based on molecular dating. Such TMRCA dating suggests that regional genetic diversification of *D. cambodiana* are of Tertiary origin, correlated with the rapid uplift of the Himalayan-Tibetan Plateau at 7–10 Ma [11], suggesting the influence of the uplift of Himalayan-Tibetan Plateau on genetic divergence. The historical events along with relationship between latitude and altitude

indicate that *D. cambodiana* may have migrated to the north from the south as a megathermal rainforest element before the geological uplift of the north (Figure 1d) tracking the temperature increase during the early Tertiary [6]. This suggest that the uplift of mountains may have contributed to the increase in genetic diversity in north Indochina Peninsula. These evidences suggest that the genetic stock resulting from initial differentiation of *D. cambodiana* during the Tertiary may have persisted and served as the source material for further diversification.

The Quaternary period (~2.4 Ma to the present) is considered as one of the most important periods for genetic diversification and speciation [79,80,89] in a variety of ecosystems due to frequent changes in the climate and associated series of glaciations or ice ages [90,91,92]. The  $T_{\text{TMRCA}}$  of *Dysosma versipellis* haplotypes corresponds to Pleistocene era suggesting the effect of glacial and



**Figure 4. Population demographics of *Dracaena cambodiana*.** a: mismatch distribution. dashed lines are observed and black lines are expected values. b (change in population size over time) and c (divergence of lineages): x-axis is the median time in millions years ago (Ma); black lines represent median estimate and gray lines represent 95% confidence intervals. doi:10.1371/journal.pone.0060102.g004

**Table 4.** The neutrality tests and mismatch distribution analysis within four groups and overall populations of *Dracaena cambodiana*.

Group	SSD	HRag	Tajima's <i>D</i>	Fu's <i>F<sub>s</sub></i>
S-I	0.079	<b>0.263</b>	0.679	-1.169
NE-I	<b>0.074</b>	<b>0.264</b>	0.751	0.420
N-I	0.026	0.101	-0.003	-0.941
HN	0.132	0.483	0.875	1.473
Overall	<b>0.011</b>	<b>0.053</b>	0.549	-2.597

Tajima's *D* and Fu's *F<sub>s</sub>* and their significance values are reported. The sum of squared deviation (SSD) and raggedness index (HRag) are for testing the sudden expansion model. The bolds indicate  $P < 0.05$ .

doi:10.1371/journal.pone.0060102.t004

interglacial cycle on genetic divergence of plants [93]. Relatively rapid diversification of *D. cambodiana* terminal lineages with short branches corresponding to the Pleistocene period (Figure 4c, Figure S3 and Figure S4) indicates a rapid recent divergence of populations caused by Quaternary climate change [5,17,94]. However, skyline plot of cpDNA and bottleneck analyses of microsatellites suggested that *D. cambodiana* experienced a recent reduction in population, suggesting a Quaternary climate change also induced population size change. Thus, we inferred that evolutionary processes such as extinction-recolonization due to climate change during the Pleistocene may have contributed to the fast diversification in *D. cambodiana*.

## Supporting Information

**Figure S1 The posterior probability convergence of  $t_0$ ,  $t_1$ ,  $t_2$  and  $t_{\text{TMRCa}}$ .** A: convergence of time parameters for each node in Figure 3. B: convergence of time parameter for the most recent common ancestor.  
(PDF)

**Figure S2 The phylogenetic network of haplotypes with outgroups.** O-1 = *Agave sp.*, O-2 = *Yucca gloriosa*, O-3 = *Asparagus plumosus*. Red dots are missing haplotypes or missing samples.  
(TIF)

**Figure S3 The bayesian tree for evolutionary rate estimation based on uncorrelated lognormal relaxed clock.** Haps are the cpDNA haplotypes. The numbers above branches are the median rates of substitutions per site per million years. The numbers below the branches denotes the Bayesian posteriors. The height-median divergence time (boldfaces) and the 95%HPD time ranges (blue bars) marked at the nodes for the posteriors are more than 0.6. The coalescent times of the most recent common ancestor ( $T_{\text{TMRCa}}$ ) of 18 cpDNA haplotypes is 15.80 (5.29–29.88) Ma.  
(TIF)

## References

- Hooghiemstra H, van der Hammen T (1998) Neogene and Quaternary development of the neotropical rain forest: the forest refugia hypothesis, and a literature overview. *Earth-Sci Rev* 44: 147–183.
- Colinvaux PA, De Oliveira PE, Bush MB (2000) Amazonian and neotropical plant communities on glacial time-scales: the failure of the aridity and refuge hypotheses. *Quaternary Sci Rev* 19: 141–169.
- Ejlsdå J, Lovett JC (2004) Geographical patterns of old and young species in African forest biota: the significance of specific mountain areas as evolutionary centres. *Biodivers Conserv* 6: 325–346.
- Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends Ecol Evol* 19: 639–644.
- Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, et al. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.
- Morley RJ (2007) Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In: Bush MB, Flenley JR, editors. *Tropical Rainforest Response to climate change* New York: Springer Berlin Heidelberg, pp. 1–31.
- van der Hammen T, Hooghiemstra H (2000) Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Sci Rev* 19: 725–742.

**Figure S4 The bayesian tree of *D. cambodiana* individuals based on Bayesian Skyline model.** The height-median divergence time (boldfaces) and the 95%HPD time ranges (blue bars) marked at the nodes. The coalescent times of the most recent common ancestor ( $T_{\text{TMRCa}}$ ) of all individuals is 9.77 (3.69–18.64) Ma.  
(TIF)

**Table S1 log10 Bayes factors of three different relaxed clock models.**  
(DOCX)

**Table S2 Pairwise genetic differentiation ( $F_{\text{ST}}$ ) among four groups.** The permutations for significance tests were 1023. *P*-value for all  $F_{\text{ST}}$  is less than 0.001.  
(DOCX)

**Table S3 The diversity and genetic structure parameters (*F*-Statistics, global *R*-Statistics and heterozygosity values) corresponding to each of the six microsatellite loci.**  
(DOCX)

**Table S4 The genetic parameters of populations of *D. cambodiana* in Hainan Island and Indochina Peninsular:** expected and observed heterozygosity ( $H_E$  and  $H_O$ ), and mean observed number of alleles per polymorphic locus ( $A$ ), mean allelic richness per polymorphic locus ( $A_R$ ), number of private alleles ( $A_P$ ),  $F_{\text{IS}}$  and  $F_{\text{ST-SSR}}$  per population.  $H_E$  in \* indicate significant deviations from HWE (at 0.05 significance level). The negative value of  $F_{\text{IS}}$  indicates heterozygosity excess.  
(DOCX)

**Table S5 Analysis of population bottleneck based on IAM and TPM model of microsatellite evolution and Mode shift test for allele frequency distribution.** Significance of gene diversity excess ( $H_e > H_{eq}$ ) was tested using Sign test and Wilcoxon signed ranks test (Luikart & Cornuet, 1998) based on 5000 replications. \* $P < 0.05$ . NL = normal L-shaped distribution and MS = mode-shift in allele frequency distribution.  
(DOCX)

**File S1 Supplementary materials and methods.**  
(PDF)

## Acknowledgments

We thank Tran Xuan Cuong for sample collection in Pumat National Park, Vietnam, Ashley Thomson for comments on the manuscript. We also thank Giovanni G Vendramin, Ivan SCOTTI and an anonymous reviewer for their valuable comments on the submitting process.

## Author Contributions

Conceived and designed the experiments: JLZ LZ QML. Performed the experiments: JLZ LZ QML. Analyzed the data: JLZ SD SN. Contributed reagents/materials/analysis tools: JLZ LZ DML SD SN. Wrote the paper: JLZ LZ QML SD SN.

8. Colinvaux PA, De Oliveira PE, Moreno JE, Miller MC, Bush MB (1996) A long pollen record from lowland Amazonia: forest and cooling in glacial times. *Science* 274: 85–88.
9. Tapponnier P, Lacassin R, Leloup PH, Schärer U, Zhong Dalai, et al. (1990) The Ailao Shan/Red River metamorphic belt: Tertiary left-lateral shear between Indochina and South China. *Nature* 343: 431–437.
10. Molnar P, England P, Martinod J (1993) Mantle dynamics, uplift of the Tibetan Plateau, and the Indian Monsoon. *Rev Geophys* 31: 357–396.
11. An Z, Kutzbach J, Prell W, Porter S (2001) Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan Plateau since Late Miocene time. *Science* 411: 62–66.
12. Yin A, Harrison TM (2000) Geologic evolution of the Himalayan-Tibetan orogen. *Annu Rev Earth Planet Sci* 28: 211–280.
13. Clift PD, Hodges KV, Heslop D, Hannigan R, Van Long H, et al. (2008) Correlation of Himalayan exhumation rates and Asian monsoon intensity. *Nat Geosci* 1: 875–880.
14. Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends Plant Sci* 3: 432–438.
15. McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *0012–9658* 86: 2088–2098.
16. Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proc Natl Acad Sci U S A* 103: 12447–12450.
17. Riddle B (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. *Trends Ecol Evol* 11: 207–211.
18. Hall R (1998) The plate tectonics of Cenozoic SE Asia and the distribution land and sea. In: Hall R, Holloway JD, editors. *Biogeography and Geological Evolution of SE Asia*. Leiden, The Netherlands: Backhuys Publishers. pp. 99–131.
19. Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, et al. (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science* 289: 1343–1346.
20. Pearson R (2006) Climate change and the migration capacity of species. *Trends Ecol Evol* 21: 111–113.
21. Petit RJ, Kremer A, Wagner DB (1993) Finite island model for organelle and nuclear genes in plants. *Heredity* 71: 630–641.
22. Marrero A, Almeida RS, González-Martín M (1998) A new species of the wild dragon tree, *Dracaena* (Dracaenaceae) from Gran Canaria and its taxonomic and biogeographic implications. *Botanical Journal of the Linnean Society* 128: 291–314.
23. Chen XQ, Turland NJ (2000) *Dracaena Vandellii* ex Linnaeus. In: Wu CY, Raven PH, editors. *Flora of China*: Beijing: Science Press; St Louis: Missouri Botanical Garden Press. pp. 215–217.
24. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11–15.
25. Chiang T, Schaal BA, Peng C (1998) Universal primers for amplification and sequencing a noncoding spacer between the *atpB* and *rbL* genes of chloroplast DNA. *Bot Bull Acad Sinica* 39: 245–250.
26. Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol Ecol* 4: 129–131.
27. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-w-improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
28. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
29. Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematic. *Ann Mo Bot Gard* 87: 482–498.
30. Ingvarsson PK, Ribstein S, Taylor DR (2003) Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. *Mol Biol Evol* 20: 1737–1740.
31. Zhang L, Li QM (2010) Isolation and characterization of microsatellite markers in an endangered species *Dracaena cambodiana* (Liliaceae). *Am J Bot* 97: e91–e93.
32. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
33. Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237–1245.
34. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* 1: 47–50.
35. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4: 535–538.
36. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* 5: 453–455.
37. Rowley DB, Pierrehumbert RT, Currie BS (2001) A new approach to stable isotope-based paleoaltimetry: implications for paleoaltimetry and paleohypsometry of the High Himalaya since the Late Miocene. *Earth Planet Sci Lett* 5836: 1–17.
38. Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
39. Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59: 1633–1638.
40. Weir BS, Cockerham CC (1984) Estimating *F*-Statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
41. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.
42. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2: 618–620.
43. Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145: 1219–1228.
44. Rohlf FJ (2000) NTSYS-pc V2.10c: numerical taxonomy and multivariate analysis system. Exeter Software, New York.
45. Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* 163: 1467–1482.
46. Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources* 10: 556–557.
47. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
48. Ostrowski MF, David J, Santoni S, McKhann H, Reboud X, et al. (2006) Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Mol Ecol* 15: 1507–1517.
49. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
50. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
51. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4: 359–361.
52. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
53. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4: 137–138.
54. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.
55. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
56. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
57. Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959–969.
58. Bogler DJ, Simpson BB (1996) Phylogeny of Agavaceae based on ITS rDNA sequence variation. *Am J Bot* 83: 1225–1235.
59. Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A* 104: 2785–2790.
60. Hey J (2010) Isolation with migration models for more than two populations. *Mol Biol Evol* 27: 905–920.
61. Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol* 3: e193.
62. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
63. Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
64. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185–1192.
65. Drummond AJ, Rambaut A (2007) BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
66. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from <http://beast.bio.ed.ac.uk/Tracer>.
67. Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555–562.
68. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9: 552–569.
69. Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol* 20: 76–86.
70. Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 64: 591–600.
71. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.
72. Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol* 12: 228–237.
73. Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* 90: 502–503.

74. Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Mol Ecol* 7: 963–974.
75. Avise JC (2000) *Phylogeography: the history and formation of species*. Cambridge, Massachusetts, London, England: Harvard University Press.
76. Qiu Y-X, Fu C-X, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol* 59: 225–244.
77. Shepherd LD, Perrie LR (2011) Microsatellite DNA analyses of a highly disjunct New Zealand tree reveal strong differentiation and imply a formerly more continuous distribution. *Mol Ecol* 20: 1389–1400.
78. Cannon CH, Manos PS (2003) Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). *J Biogeogr* 30: 211–226.
79. Willis KJ, Niklas KJ (2004) The role of Quaternary environmental change in plant macroevolution: the exception or the rule? *Philos Trans R Soc B Biol Sci* 359: 159–172.
80. Hewitt G (2000) The genetic legacy of Quaternary ice ages. *Nature* 405: 907–913.
81. Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, S B, et al. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.
82. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16: 97–159.
83. Wright S (1965) The interpretation of population structure by *F*-Statistics with special regard to systems of mating. *Evolution* 19: 395–420.
84. Lacassin R, Leloup PH, Tapponnier P (1993) Bounds on strain in large tertiary shear zones of SE Asia from boundinage restoration. *J Struct Geol* 15: 677–692.
85. Huchon P, Le Pichon X, Rangin C (1994) Indochina Peninsula and the collision of India and Eurasia. *Geology* 22: 27–30.
86. Searle MP (2006) Role of the Red River Shear zone, Yunnan and Vietnam, in the continental extrusion of SE Asia. *J Geol Soc London* 163: 1025–1036.
87. Ducrocq S, Chaimanee Y, Sutethorn V, Jaeger J (1995) Mammalian faunas and the ages of the continental Tertiary fossiliferous localities from Thailand. *J Southeast Asian Earth Sci* 12: 65–78.
88. Kuhner M (2009) Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol* 24: 86–93.
89. Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc B Biol Sci* 359: 183–195.
90. Ruddiman WF, Raymo ME, Martinson DG, Clement BM, Backman J (1989) Pleistocene evolution: northern hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography* 4: 353–412.
91. Webb TI, Bartlein PJ (1992) Global changes during the last 3 million years: climatic controls and biotic responses. *Annu Rev Ecol Syst* 23: 141–173.
92. Lisiecki LE, Raymo ME (2005) A Plio-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* 20: 1–17.
93. Qiu Y-X, Guan B-C, Fu C-X, Comes HP (2009) Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dysoxma versipellis*. *Mol Phylogenet Evol* 51: 281–293.
94. Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293: 2242–2245.