# Domestication history reveals multiple genetic improvements of Chinese bayberry cultivars

**Running title: Domestication history of Chinese bayberry**

# Materials and Methods

## *Plant sampling*

A total of 44 domesticated *M. rubra* individuals were collected in orchards in Chongqing, Fujian, Guangdong, Jiangsu, and Zhejiang provinces, including 12 landraces and 32 improved cultivated varieties (**Supplementary Table S1**). Nineteen wild *M. rubra* individuals were also sampled from nine provinces from Yunnan to Zhejiang that cover the entire natural distribution ranges of *M. rubra* in China (**Supplementary Table S2**). Fourteen individuals of six other *Morella* species were also used to resolve the phylogenetic relationship, including 12 from China (one *M. adenophora*, five *M. esculenta*, and six *M. nana*), and two from North America (*M. cerifera* and *M. pensylvanica*). Based on previous studies (Liu *et al.*, 2015; Yang *et al.*, 2021), *Comptonia peregrina* was selected as the outgroup.

## *DNA extraction, library preparation, and RAD-sequencing*

Total genomic DNA was extracted from silica dried leaves using DNA plantzol (Invitrogen Corp, USA) following the manufacturer’s protocol. Single-digest restriction-site associate DNA (RAD-seq) library preparation and sequencing were performed by Beijing Genomics Institute (Shenzhen, China) using the restriction enzyme *EcoR* I and sample-specific barcodes. The digested DNA with *EcoR* I was ligated by P1 adapter. Then, the DNA products were pooled together and sheared using a Covaris S220-DNA Sonicator (Covaris, INC., Woburn, MA, United States). Fragments were retained with a length from 300 bp to 500 bp and selected fragment ends were repaired. An “A” tailing and the P2 adapter were ligated to the target fragments before PCR amplification. Sequencing was conducted on the Illumina HiSeq 2000 platform with a read length of 100bp. For quality checking reasons, the same *C. peregrine* individual was sequenced twice and merged. On average 100 Megabytes raw reads were produced for each sample, which gives a mean read depth of 15 ~ 30× with a genome size of *M. rubra* equals 313 Mb.

## *Reference-guided read alignment, variant calling, and quality control*

Adaptor-free short reads were aligned to the *M. rubra* reference genome (Mru\_ZJU\_2; GenBank assembly accession: GCA\_003952965.2) using BWA-MEM v0.7.17 28. Aligned short reads were sorted by SAMtools v1.7 29 and reads of possible PCA duplicates were marked by Picard v2.21.6. GATK v4.7.1.0 30 HaplotypeCaller and GenotypeGVCFs were applied successively to identify single nucleotide polymorphisms (SNPs). The hard filtering was conducted on raw SNPs by GATK VariantFiltration with options ‘QD<2, MQ<20, FS>60, SOR>3, MQRankSum<-3, ReadPosRankSum<-3, DP<150 and DP>400’. A two-step quality control was further performed on the filtered SNPs as the following: (1) SNPs in the female-specific region (FSR) were excluded 19; (2) sites were filtered out if any of the following three criteria was met: i) more than 30% of individuals were not genotyped, ii) the minor allele frequency (MAF) was lower than 0.01, and iii) more than two alleles were identified. Finally, 595,448 SNPs were retained. For the inferences of phylogenetic relationships, population structure and demographics, only one SNP was kept in every 100 bp sliding window with a step size of 10 bp to remove the effect of linkage disequilibrium using PLINK v1.90 31.

## *Phylogenetic relationship, population structure, and population genetics*

After pruning for linkage disequilibrium, 67,064 SNPs were retrieved for the following analyses. An approximate maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE v2.1.2 32. The best nucleotide substitution model (‘TVMe+R3’) was determined by ‘ModelFinder’ implemented in IQ-TREE. Branch support scores were calculated using a bootstrap of 5,000 replicates.

A SNP-based principal component analysis (PCA) was applied to assess population structure. The first two PCs explained more than 9% of the total variance. Ancestral components and admixtures were then inferred by Admixture v.1.3.0 (Alexander et al., 2009) with the number of genetic clusters K tested from 1 to 7. The best K value was selected when the lowest cross-validation error was found.

Three summary statistics of population genetics, including pairwise nucleotide difference π, pairwise genetic divergence FST, and Tajima’s D, were calculated on the whole data set containing 595,448 SNPs using VCFtools v0.1.17 34.

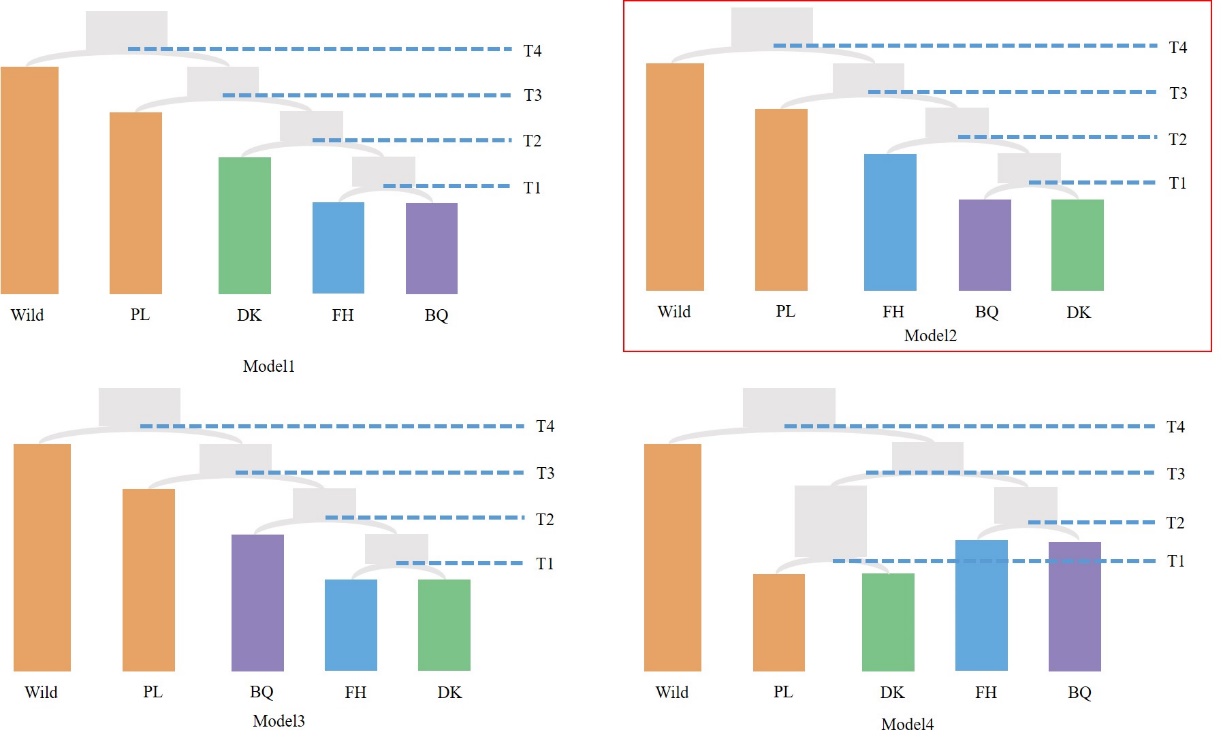
## *Inferences of domestication history*

The demographic and domestication history of *M. rubra* was inferred by Fastsimcoal2 v2.6.0.3 35. Two-dimensional joint site frequency spectrum (2D-SFS) of size 8 x 8 were generated using EST-SFS v2.03 36 for pairs of six population groups inferred from the phylogenetic and clustering analyses, representing the wild (Wild), the landrace (PL), the two most popular cultivars ‘Biqi’ (BQ), ‘Dongkui’ (DK), all other cultivars in a monophyletic group (FH), and a possible admixed / hybrid group (HY). A basic scenario describes an early domestication event generating the landrace, giving rise to two main cultivars, DK and BQ, and from which all other cultivars (FH) were developed. Based on this basic scenario, we first tested four models concerning how these cultivars further diverged from each other (**Fig. S1 a**). Model 1 to 3 depict that DK, BQ, and FH were developed in different orders through two improvement events while model 4, on the other hand, specifies DK developed from a different landrace other than the common ancestor of FH and BQ. 100 replicates were performed for each model with 10,000 coalescent simulations (-n 10,000) and 40 cycles of Expectation Conditional Maximization for likelihood (-L 40). The best-supported model was selected based on Akaike’s weight value 35. Secondly, we further modified the best-supported model to test if the admixed group HY can be a result of hybridization between BQ and FH (model 5), or BQ and PL (model 6), or BQ and wild (model 7) (**Fig. S1 b, Table S3**). For parameter estimation, we calculated the 95% confidence intervals using parametric bootstrap with 100 iterations based on the best point estimates that maximizes the likelihood of the model. The goodness of fit was evaluated by the likelihood ratio of G-statistics. For all models, we assumed a mutation rate of 9.40 × 10-9 per site per year and a generation time of 10 years were used 20. All populations were simulated with constant sizes through time as Tajima’s D values had a minor deviation from zero.

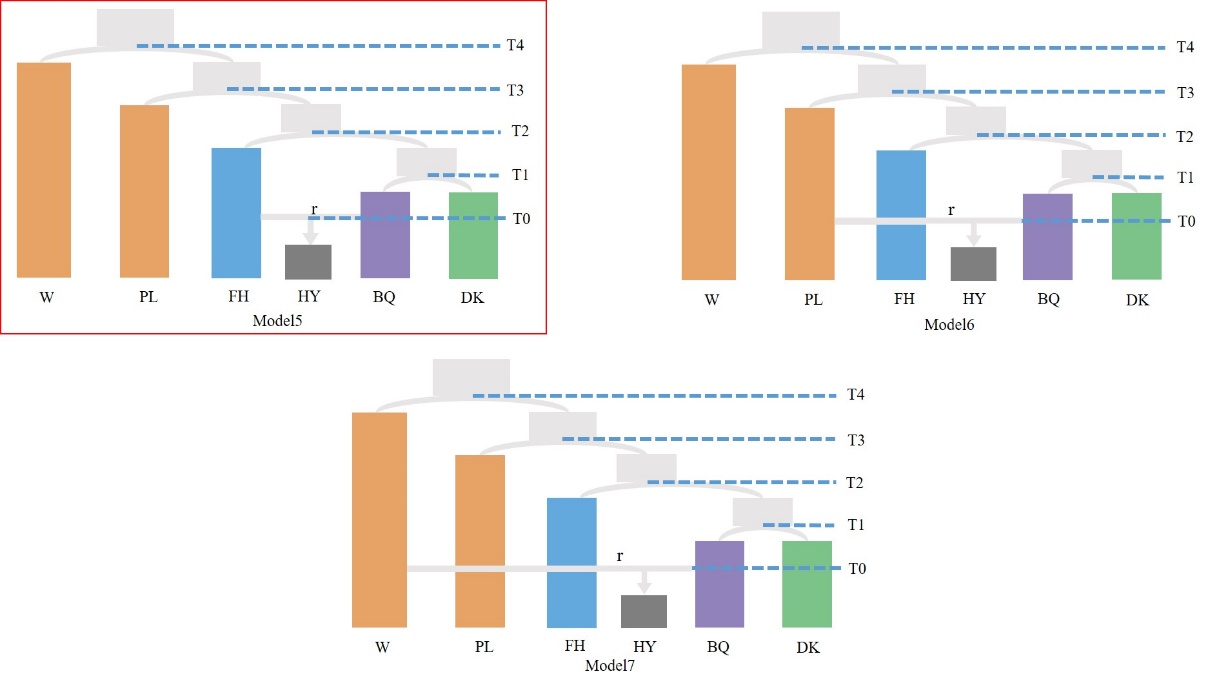
**Supplementary information**

**Figure S1. Seven tested models for the domestication and demographic process of *M. rubra*. a,** Models 1 to 3 depict that DK, BQ, and FH were developed in different orders through two improvement events while model 4, on the other hand, specifies DK developed from a different landrace other than the common ancestor of FH and BQ. **b**, Model 5 to 7 test if the admixed group HY might be a result of hybridization between BQ and FH (model 5), or BQ and PL (model 6), or BQ and wild (model 7) based on the best-supported model.

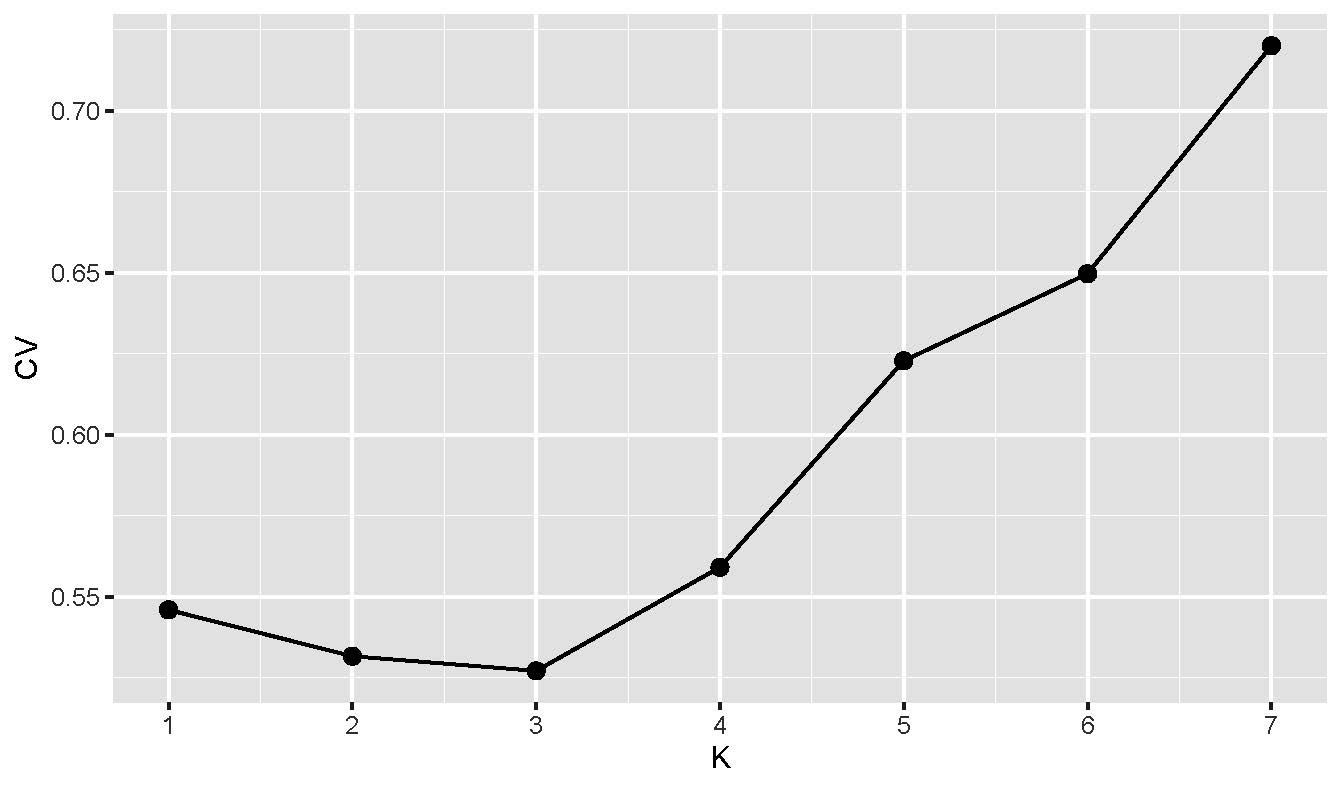
**Figure S1a**.

****

**Figure S1b.**

****

**Figure S2. Cross-validation error regarding the number of clusters for unsupervised population clustering.**

****

**Table S1.** **Details of location and sampling information for cultivated populations of *Morella rubra* investigated in this study.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Collection area | ID | Cultivars | Location | Produced area | Latitude | Longitude | Voucher no. |
| Jiangsu | WXWM | ‘Wumei’ | Mashanzhen, Wuxi, Jiangsu, China | Jiangsu | N31°27’57.14" | E120°07’04.05" | GLH120725001 |
|  | WXBQ | ‘Biqi’1 | Mashanzhen, Wuxi, Jiangsu, China | Zhejiang | N31°27’57.14" | E120°07’04.05" | GLH120725002 |
|  | SZLD | ‘Langdangzi’ | Xishan, Suzhou, Jiangsu, China | Jiangsu | N31°06’58.83" | E120°17’10.35" | GLH120725003 |
|  | SZWZ | ‘Wangzi’anhai’ | Dongshan, Suzhou, Jiangsu, China | Fujian | N31°05’25.25" | E120°26’04.57" | GLH120725005 |
|  | SZDK | ‘Dongkui’1 | Xishan, Suzhou, Jiangsu, China | Zhejiang | N31°06’58.83" | E120°17’10.35" | GLH120725006 |
|  | SZDY | ‘Dayexidi’ | Dongshan, Suzhou, Jiangsu, China | Jiangsu | N31°05’25.25" | E120°26’04.57" | GLH120725007 |
|  | SZXY | ‘Xiaoyexidi’ | Dongshan, Suzhou, Jiangsu, China | Jiangsu | N31°05’25.25" | E120°26’04.57" | GLH120725008 |
| Zhejiang | HZZS | ‘Zaose’ | Suoqianzhen, Hangzhou, Zhejiang, China | Zhejiang | N30°04’55.25" | E120°18’35.84" | GLH120620001 |
|  | HZDT | ‘Datanmei’ | Yuhang district, Hangzhou, Zhejiang, China | Zhejiang | N30°25’02.58" | E120°17’11.60" | GLH120804002 |
|  | HZCS | ‘Chise’ | Suoqianzhen, Hangzhou, Zhejiang, China | Zhejiang | N30°04’55.25" | E120°18’35.84" | GLH120620002 |
|  | SYSH | ‘Shenhong’ | Yitingzhen, Shaoxing, Zhejiang, China | Zhejiang | N29°59’45.10" | E120°58’01.90" | GLH120626001 |
|  | SYSJ | ‘Shuijing’1 | Yitingzhen, Shaoxing, Zhejiang, China | Zhejiang | N29°59’45.10" | E120°58’01.90" | GLH120626002 |
|  | YYSJ | ‘Shuijing’2 | Zhangtingzhen, Ningbo, Zhejiang, China | Zhejiang | N30°02’36.22" | E121°15’16.04" | GLH120801003 |
|  | YYFH | ‘Fenghong’ | Zhangtingzhen, Ningbo, Zhejiang, China | Zhejiang | N30°02’36.22" | E121°15’16.04" | GLH120801001 |
|  | YYWQ | ‘Wanqi’ | Zhangtingzhen, Ningbo, Zhejiang, China | Zhejiang | N30°02’36.22" | E121°15’16.04" | GLH120801002 |
|  | YYLZ | ‘Lizhi’ | Zhangtingzhen, Ningbo, Zhejiang, China | Zhejiang | N30°02’36.22" | E121°15’16.04" | GLH120801004 |
|  | CXBQ | ‘Biqi’2 | Henghezhen, Ningbo, Zhejiang, China | Zhejiang | N30°06’02.77" | E121°16’00.25" | GLH120804006 |
|  | CXZQ | ‘Zaoqi’ | Henghezhen, Ningbo, Zhejiang, China | Zhejiang | N30°06’02.77" | E121°16’00.25" | GLH120726002 |
|  | CXZD | ‘Zaodao’ | Henghezhen, Ningbo, Zhejiang, China | Zhejiang | N30°06’02.77" | E121°16’00.25" | GLH120726003 |
|  | JBCQ | ‘Ciqi’ | Cichengzhen, Ningbo, Zhejiang, China | Zhejiang | N29°54’40.97" | E121°35’15.45" | GLH120802001 |
|  | FHQY | ‘Qingying’ | Chunhuzhen, Ningbo, Zhejiang, China | Zhejiang | N29°37’14.01" | E121°35’34.05" | GLH120802002 |
|  | FHHY | ‘Huangying’ | Chunhuzhen, Ningbo, Zhejiang, China | Zhejiang | N29°37’14.01" | E121°35’34.05" | GLH120802003 |
|  | ZSWD | ‘Wandao’ | Zhoushan, Hangzhou, Zhejiang, China | Zhejiang | N29°59’51.15" | E122°12’43.68" | GLH120629001 |
|  | NBWZ | ‘Wuzi’ | Henghezhen, Ningbo, Zhejiang, China | Zhejiang | N30°06’02.77" | E121°16’00.25" | GLH120804001 |
|  | NHZM | ‘Zhenmei’ | Ninghai, Ningbo, Zhejiang, China | Zhejiang | N29°17’04.25" | E121°26’31.34" | GLH120804003 |
|  | LXZT | ‘Zaotanmei’ | Majianzhen, Jinhua, Zhejiang, China | Zhejiang | N29°18’37.31" | E119°36’22.51" | GLH120616001 |
|  | XJDK | ‘Dongkui’2 | Xianju, Taizhou, Zhejiang, China | Zhejiang | N28°51’41.58" | E120°45’17.62" | GLH120623001 |
|  | TZTZ | ‘Tongzi’ | Tingpangzhen, Taizhou, Zhejiang, China | Zhejiang | N29°02’28.46" | E121°21’21.59" | GLH120803001 |
|  | LHZD | ‘Zaoda’ | Linhai, Taizhou, Zhejiang, China | Zhejiang | N28°52’48.97" | E121°09’10.98" | GLH120804005 |
|  | WZDA | ‘Ding’ao’ | Chashanzhen, Wenzhou, Zhejiang, China | Zhejiang | N27°54’28.74" | E120°42’18.71" | GLH120615001 |
| Fujian | FJDY | ‘Daye’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601008 |
|  | FJDK | ‘Dongkui’3 | Fujian academy of agricultural sciences | Zhejiang | N26°07’44.76" | E119°19’57.83" | SZD120601002 |
|  | FJYS | ‘Yingsi’anhai’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601003 |
|  | FJRS | ‘Ruansi’anhai’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601004 |
|  | FJFG | ‘Fugongyihao’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601005 |
|  | FJTZ | ‘Tezaomei’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601006 |
|  | FJB | ‘Baiyangmei’ | Fujian academy of agricultural sciences | Fujian | N26°07’44.76" | E119°19’57.83" | SZD120601007 |
|  | FJLZ | ‘Luozi’ | Jinfengzhen, Changle, Fujian, China | Fujian | N25°58’38.86" | E119°35’04.71" | SZD120602002 |
|  | FJRB | ‘Ranbingsanhao’ | Ningde, Fujian, China | Fujian | N26°39’28.30" | E119°33’31.11" | SZD120601010 |
|  | FJBX | ‘Baxiandao’ | Jinfengzhen, Changle, Fujian, China | Fujian | N25°58’38.86" | E119°35’04.71" | SZD120602001 |
|  | FJBY | ‘Baiyangsanhao’ | Putian, Fujian, China | Fujian | N25°26’46.07" | E118°58’57.22" | SZD120601009 |
| Guangzhou | GZDK | ‘Dongkui’4 | Guangdong academy of agricultural sciences | Zhejiang | N23°09’16.80" | E113°21’26.82" | GLH120911001 |
|  | GZWS | ‘Wusu’ | Guangdong academy of agricultural sciences | Guangdong | N23°09’16.80" | E113°21’26.82" | GLH120911003 |
| Chongqing | CQBQ | ‘Biqi’3 | Xinglongzhen, Chongqing, China | Zhejiang | N29°55’00.30" | E106°42’32.52" | GLH120907003 |

**Table S2: Details of location and sampling information for wild *Morella* and outgroup species investigated in this study.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Taxon | ID | Location | Latitude | Longitude | Altitude(m) | Voucher no. |
| *Morella rubra* | ZJGT | Gutianshan, Zhejiang, China | N29°8’41.77" | E118°9’11.18" | 455 | CN201307041 |
|  | ZJDP | Dapanshan, Zhejiang, China | N28°59’27.16" | E120°26’26.54" | 683 | CN201307051 |
|  | ZJWY | Wuyanling, Zhejiang, China | N27°42’20.40" | E119°40’30.83" | 882 | CN20130707 |
|  | FJNP | Mangdangshan, Fujian, China | N26°41’39.52" | E118°07’32.67" | 253 | CN130324 |
|  | FJZS | Zhangzhou, Fujian, China | N24°35’30.44" | E117°56’49.70" | 210 | CN130321 |
|  | FJXC | Xiacangcun, Fujian, China | N24°35’02.18" | E117°39’30.86" | 36 | GLH20130123001 |
|  | JXFH | Lushan, Jiangxi, China | N29°33’40.03" | E115°53’04.20" | 725 | GLH12021129001 |
|  | JXNC | Nanchang, Jiangxi, China | N28°42’30.05" | E115°42’13.98" | 324 | GLH20130120001 |
|  | JXGZ | Chongyi, Jiangxi, China | N25°40’23.07" | E114°20’29.19" | 549 | GLH20130120002 |
|  | GZYS | Guangzhou, Guangdong, China | N23°08’58.35" | E113°22’17.03" | 50 | GLH120911004 |
|  | HNHT | Luoshancun, Hunan, China | N26°49’28.16" | E109°37’04.46" | 299 | JXJ20130808 |
|  | HNJS | Jishou, Hunan, China | N28°17’13.20" | E109°43’18.85" | 279 | JXJ20130809 |
|  | GXXA | Xingan, Guangxi, China | N25°46’21.78" | E110°29’08.54" | 292 | GLH121102001 |
|  | GXYS | Yanshan, Guangxi, China | N25°04’47.22" | E110°17’55.27" | 206 | GLH121103001 |
|  | GXNN | Nanning, Guangxi, China | N23°05’51.84" | E108°18’37.28" | 124 | GLH121106001 |
|  | GZLS | Leigongshan, Guizhou, China | N26°30’1.04" | E106°44’2.82" | 1135 | GLH121201001 |
|  | CQNC | Nanchuan, Chongqing, China | N29°20’47.10" | E107°14’32.77" | 766 | GLH120907001 |
|  | YNDW | Daweishan, Yunnan, China | N22°55’29.21" | E103°40’46.38" | 2257 | CN130601 |
|  | YNML | Malipo, Yunnan, China | N23°8’16.67" | E104°41’23.71" | 1171 | CN130603 |
| *M. adenophora* | GXHX | Hengxian, Guangxi, China | N22°31’52.97" | E109°11’17.15" | 74 | L131214 |
| *M. nana* | YNFM | Fuminxian, Yunnan, China | N25°18’21.31" | E102°41’24.88" | 1958 | L13121101 |
|  | YNXW | Xuanweishi, Yunnan, China | N27°2’55.94" | E101°68’04.88" | 2341 | L131208 |
|  | YNWS | Wenshanxian, Yunnan, China | N23°19’27" | E103°13’34" | 2583 | MYR2013429005 |
|  | YNPL | Panlong, Kunming, Yunnan, China | N25°10’12.94" | E102°47’2.33" | 2050 | L131207 |
|  | YNHK | Haikouzhen, Kunming, Yunnan, China | N24°48’33.77" | E102°37’25.29" | 1965 | L131210 |
|  | YNZJ | Zhaojiaocun, Kunming, Yunnan, China | N25°05’13.64" | E102°34’10.37" | 1888 | L131301 |
| *M. esculenta* | YNAL | Ailaoshan, Yunnan, China | N24°32’25" | E101°01’36" | 2504 | MYR2013520001 |
|  | YNHN | Shilinxian, Yunnan, China | N24°50’34.08" | E103°18’15.30" | 1966 | CN120919 |
|  | YNMWS | Wenshanxian, Yunnan, China | N23°19’27" | E103°13’34" | 2583 | MYR2013429003 |
|  | SCYB | Yanbianxian, Sichuan, China | N27°06’08.62" | E101°46’01.19" | 2074 | JXJ20130806 |
|  | SCMY | Miyixian, Sichuan, China | N27°02’55.94" | E101°58’04.88" | 1952 | JXJ20130805 |
| *Morella pensylvanica* | USBH | Harvard campus, Cambridge, MA, USA | N42°22’37.21" | W71°06’59.98" | 8 | F12758 |
| *Morella cerifera* | USR | NCSU campus, Raleigh, NC, USA | N35°47’04.79" | W78°40’55.54" | 127 | F1007038 |
| *Comptonia peregrina* | USR | NCSU campus, Raleigh NC, USA | N35°47’04.79" | W78°40’55.54" | 127 | F12756 |

**Table S3. Comparison of alternative demographic models tested using fastsimcoal2.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Log10L** | **AIC** | **ΔAIC** | **ωi** |
| Model1 | -764282.07 | 3519675.00 | 7740.81 | 0.00 |
| **Model2** | **-762601.18** | **3511934.19** | **0.00** | 1.00 |
| Model3 | -764362.85 | 3520047.03 | 8112.84 | 0.00 |
| Model4 | -764206.58 | 3519327.36 | 7393.17 | 0.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Log10L** | **AIC** | **ΔAIC** | **ωi** |
| **Model5** | **-1137584.39** | **5238805.69** | **0.00** | **1.00** |
| Model6 | -1137652.48 | 5239119.27 | 313.58 | 0.00 |
| Model7 | -1137726.74 | 5239416.27 | 610.58 | 0.00 |

For each model, the table shows the maximum likelihood estimate of the model (log10L), Akaike's information criterion score (AIC), the difference in AIC value of each model from that of the strongest model (ΔAIC) and AIC weight (ωi). The best-supported model is indicated in bold.