

Application of genotyping-by-sequencing data on inferring the phylogeny of *Curcuma* (Zingiberaceae) from China

Heng Liang

Sichuan Agricultural University

Yan Zhang

Sichuan Agricultural University

Jiabin Deng

guizhou education university

Gang Gao

Yibin University

Chunbang Ding

Sichuan Agricultural University

Li Zhang

Sichuan Agricultural University

Xiaofang Yu

Sichuan Agricultural University

Yonghong Zhou

Sichuan Agricultural University

Ruiwu Yang (✉ yrwu@sicau.edu.cn)

Sichuan Agricultural University

Research article

Keywords: Genotyping-by-sequencing, ipyrad, phylogeny, *Curcuma*

Posted Date: September 27th, 2019

DOI: <https://doi.org/10.21203/rs.2.15210/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Genotyping-by-sequencing (GBS), as one of the next generation sequences, has been applied to large scale genotyping in plants, which is poor in morphological differentiation and low in genetic divergence among different species. *Curcuma* is a significantly medicinal and edible genus. Improvement efforts of phylogenetic relationships and disentangling species are still a challenge due to poor morphology and lack in a reference genome.

Result: A high-throughput genomic sequence data which was obtained through GBS protocols was used to investigate the relationships among 8 species with 60 total samples of *Curcuma*. Through the use of the ipyrad software, 437,061 loci and 997,988 filtered SNPs without reliance upon a reference genome were produced. After quality control (QC) of the filtered SNPs, 1,295 high-quality SNPs were used to clarify the phylogenetic relationships among *Curcuma* species. Based on these data, a supermatrix approach was used to speculate the phylogeny, and the phylogenetic trees and the relationships were inferred .

Conclusions: Varying degrees of support can be explained, as well as the diversification events for Chinese *Curcuma*. The diversification events showed that the third intense uplift of Qinghai–Tibet Plateau (QTP) and formation of the Hengduan Mountains may speed up *Curcuma* interspecific divergence in China. The PCA suggested the same topology of the phylogenetic tree. The genetic structure analysis revealed that extensive hybridization may exist in Chinese *Curcuma*. Additionally, the GBS will be a promising approach for the phylogenetic and systematic study in the future.

Background

Curcuma L. (Zingiberaceae) is a genus of 40 species of aromatic herbs which occur in subtropical and tropical Asia. About 10 *Curcuma* species are distributed in China, of which 6 species are treated as vital Chinese folk herbal medicine in Traditional Chinese Medicine (TCM), and the extraction of rhizomes exhibits activity of anti-inflammatory, anticancer and HIV–1 protease inhibitory [1].

In recent years, the phylogeny of *Curcuma* has been given much attention, and the strong recovery support among the major lineages within the genus were topics of research [2–4]. But its phylogeny and taxonomy also have different controversy and rifts, because of the poor diagnostic characters (aerial part, underground part, floral morphology, and so on) [5]. The morphological differences in all species of the genus are neither unique nor universal. The cytological study of *Curcuma* had proved that extensive polyploidy exists in this genus. The basic chromosome numbers in different populations have different features [6]. So far, on account of the influence of polyploidy on the complexity of species, the relationships in *Curcuma* have been unknown. With the development of molecular biotechnology in the past several decades, the phylogenetic studies of *Curcuma* have resulted in much new progress [7–9]. Unfortunately, DNA regions still have some limitations in species identification, particularly sister or

closely related species, and they show a low genetic divergence between species [10]. However, the plants with a complex genome also show uncertain relationships in species-level and species boundaries [11]

Next generation sequencing (NGS) has made great progress in recent years, and high-throughput sequencing stands in the breach [12–14]. It comprised a number of procedures such as restriction-site associated DNA sequencing (RAD-Seq) and genotyping-by-sequencing (GBS), producing large amounts of genomic sequence data quickly and effectively [15–16]. Recently, many systematic biologists have begun to pay attention to how to use the reduced-representation of genomic approaches to solve some problems which are hard to work out in traditional ways. These methods are very useful to phylogenetic studies, which not only produce many genome wide phylogenetic information locis in lack of a reference genome or non-model species, but also simplify the laboratory protocols, saving cost and efficiency, particularly in the species lineages less than 60 million years old [17–19]. Although large numbers of single-nucleotide polymorphism (SNP) markers that analyzed population structure and genetic diversity were provided by a NGS-based method, the methods also have some drawbacks and uncertainties including short sequence reads (50–100 bp). There might appear some mistakes in sampling error, mutation and complex bioinformatics analysis, making it hard to apply to deeper timescale studies. But nonetheless, NGS has still successfully helped many researchers to understand the phylogeny for some representative family members of plants [20–23].

Genotyping-by-sequencing (GBS) approach obtained numerous sequence tags from individual species and the tags could be used for polymorphism detection of multispecies coalescent-based tree inference methods and polymorphism detection without a reference genome sequence. Thus, it is almost fit for any organism [24]. The data based on GBS, using a multispecies coalescent-based approach have been shown to be highly useful for some plant studies taking gene history variation into account [25–26].

In this study, the main objectives of this study were (1) to test GBS performance for inferring the phylogeny of Chinese *Curcuma*; (2) to evaluate phylogenetic data analysis the lineages clade; (3) to obtain a resolved phylogeny with traditional maximum likelihood (ML), maximum parsimony (MP) and Bayesian methods; (4) to evaluate the divergence time and population structure of several Chinese *Curcuma*. The work will become a plowing ahead in this genus, and determine if the data base on GBS can help the phylogeny of *Curcuma*.

Results

Sequences discovery and characterization

From the project, the sequences of 8 species a total of 61 samples (Table 1) returned 53.15 Gb, with an average of 892.2 Mb per sample. In order to obtain reliable data ($Q_{20} \geq 93.74\%$, $Q_{30} \geq 85.10\%$), some low quality data were left out. The NCBI BioProject ID: PRJNA557061 can have access to Raw Reads.

Using the denovo clustering method in ipyRAD, 437,061 unique GBS locis across all 60 samples were revealed. After quality control (QC) of the filtered SNPs by Plink, filtering SNPs reduced the total to 1,295

base pairs.

Maximum parsimony inference

The maximum parsimony method was used to analyze the datasets. Fig. 1 was bootstrap 50% majority-rule consensus tree. The bootstrap values were marked on the top of branch in each species group.

The same species from different areas were clustered together, and species group had the highest bootstrap support (8 out of 13 clades with 100%, 6 out of 13 clades with 80–99%, and none < 50%). The group *C. longa* formed a monophyletic branch (BS = 100); the group *C. wenyujin* (BS = 100) was sister to *C. aromatica* (BS = 100); the group *C. amarissima* (BS = 100) was sister to *C. sichuanensis* (BS = 100); the group *C. kwangsiensis* first clustered with *C. yunnanensis* (BS = 97) and *C. phaeocaulis* (BS = 100). Compared to the previous study, these analyses got different clades [3].

Maximum likelihood inference

The maximum likelihood method got the same clades for MP tree. The ML analyses had a higher bootstrap support (8 out of 13 clades with 100% and none < 90%) (Fig. 1). The bootstrap values were marked on the top of branch in each species group. The group *C. longa* formed a monophyletic branch with 100% support. The group *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis* clustered a big clade with 100% support. The group *C. wenyujin* and *C. aromatica* clustered with other clades with 99% support. The relationship between *C. longa* and the clade consisting of *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis* was close, which was 100%. The genus relationship of the sister group in the MP analysis was the same in the ML analysis.

Bayesian inference

The Bayesian tree (Fig. 2) showed the same cluster and clades to MP tree and ML tree. The Bayesian analysis had a higher support than MP tree (12 out of 15 clades with 100%, 3 out of 15 clades with 85–99%, and none < 85%). The sister clades which differentiated to the species group *C. wenyujin* and *C. aromatica* had a principle of maximum differentiation support in MP (BS = 99), ML (BS = 75) and Bayesian (PP = 85) tree.

PCA inference

The PCA based on high-quality SNPs could reflect the clustering and relationship among different individuals in a visual graphic from 60 *Curcuma* sample. Obviously, the same species had a better cluster. The genetic background could be reflected by the distance of species. In Fig. 3, different species had their own individual locations which were not overlapped. The *C. longa* was far from other species. The *C.*

wenyujin was close to *C. aromatica*. The *C. sichuannensis* and *C. amarissima* formed a closely related cluster. The group of *C. phaeocaulis*, *C. yunnanensis* and *C. kwangsiensis* had a close cluster. The PCA suggested the same topology of the phylogenetic tree (MP, ML and Bayesian).

Population structure inference

On the strength of the number of different assumed ancestral kinships (K), the composition of ancestor species was shown in a visual graphic by the Bayesian algorithm. The structure was run K from 3–12 for 20 times (Fig. S1). With a list as a sample, the proportion of different color in samples expressed different ancestors. The population structure analysis was similar to that of the phylogenetic tree and PCA. All species had the same individual colors (Fig. 4). When $K = 3$, the results suggested that the Chinese *Curcuma* samples branches could be classified into three groups. The *C. longa* had an individual gene pool (blue). Most gene pool (green) comprised the *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis*. The *C. wenyujin* and *C. aromatica* were composed of most gene pool (red). With the number of ancestor increasing, more species groups were distinguished clearly. When $K = 3–12$, the *C. longa* could be distinguished by constitution of gene pool, and the results could be the origin of ancestors, which were different from other groups. The *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis* had a close relationship, and they could be distinguished in $K = 5$. The *C. kwangsiensis* and *C. aromatica* consisted of two or more gene pools, and most likely hybrid origin.

Divergence time inference

Based on the GBS dataset in the chronophylogenetic analyses (Fig. 2), the major divergence times and events of Chinese *Curcuma* species occurred in the Miocene (~7.45 Mya). The *C. longa* was an independent branch, splitting at ~6.43 Mya, and intraspecific diversification was diverged at ~2.45 Mya during the late Pliocene and Quaternary. Another branch consisted of *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis*. The clusters A (*C. amarissima* and *C. sichuanensis*) and B (*C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis*) were estimated to be ~4.83 Mya. The divergence of the A lineages started ~3.57 Mya in the Pliocene, and then intraspecific diversification was estimated around ~1.89 Mya for *C. sichuanensis* and ~0.67 Mya for *C. amarissima*. The B lineages were diverged at ~3.74 Mya, the *C. phaeocaulis* split at ~1.45 Mya and the *C. yunnanensis* split at ~0.81 Mya. The branch which consisted of *C. wenyujin* and *C. aromatica* was diverged at ~6.08 Mya during the Miocene and Pliocene; subsequent diversification events ~1.17 Mya diverged the species *C. wenyujin* and ~0.56 Mya diverged the species *C. aromatica*.

Discussion

Phylogeny and biological implication

In this study, the phylogeny and evolution of Chinese *Curcuma* were explained by using GBS data firstly, which was never applied to Zingiberaceae. The phylogenetic relationships of eight main Chinese *Curcuma* species are reasonable. Unfortunately, our samples are only a subset of the species in *Curcuma*. It is always impossible to compare with previous results directly, because of the limited and not matched taxon samples. However, our study supports several hypotheses about the origin of *Curcuma* [4]. Based on GBS data, we inferred the earliest appearance of *Curcuma* in China during the Miocene, a period when large-scale orogenesis and other geological events had an influence on the development and speciation of the flora in this area [27–28]. The interspecific divergence events of Chinese *Curcuma* occurred in the late Miocene. As the third intense uplift of Qinghai–Tibet Plateau (QTP) and formation of the Hengduan Mountains began in the Pliocene [29–30], it changed the geographical environment and climate in Asia [31–32]. These events may speed up *Curcuma* interspecific divergence, when it arrived at China.

Hybridization and polyploidy are ubiquitous in plants and play an important role in speciation and diversification [33–34]. The cytotoxicity and chromosome doubling can promote the production of a large number of new phenotypes in a short time, and polyploidy tend to be more adaptive than parents [35–37]. Hybrids and polyploidization are found in the regions where there are species overlap and changes in the living environment. In the study, Chinese *Curcuma* may have a hybrid origin. According to previous studies about molecular data, *Curcuma* was likely to hybridize among distantly related species [4], and our results seem to support this opinion. Because our samples are only a subset of the species in *Curcuma*, it is difficult to discuss their origin further. In addition, there was a complex history of *Curcuma*.

4.2. Phylogeny and biological implication of the phylogeny of *Curcuma*

Many researches have shown that the “next-generation” as a new sequence approach and inference method have started to conquer some disputed problems in the relationship of closely phylogenetic lineages and estimate the recent divergence times and events [38,20]. This study was the first to use GBS to analyze the phylogeny of *Curcuma*, which produced level of genomic sequence data as massive as possible. Based on the sequence data, it aims to generate high-quality SNPs to infer the evolution and phylogenetic relationships of the radiation of Chinese *Curcuma*. Over 400,000 loci and 997,988 filtered SNPs were produced from 8 species 60 samples and 1 outgroup. The high-quality SNPs “coverage” among all samples were obtained with the method of quality control (QC) of the filtered SNPs. Compared to many previous studies of this group, the quality of our dataset had a significant increase. The concatenated supermatrix data from QC of the filtered SNPs returned a strong support for the 8 species relationships of Chinese *Curcuma*. From fabricated different phylogenetic tree, the trees showed fairly consistent topology and recovered a similar clades with high support. A previous study demonstrated that several mistakes will appear in inference by using a single phylogenetic tree method [39]. Especially, in the multispecies coalescent supermatrix data and model, there might exist inappropriate conclusions [40]. Some studies noted that the approach of obtaining numerous sequence data is relatively ideal for establishing and reconstructing the phylogeny [41–42]. Researches in molecular evolutionary biology

often lack in locis and variable sites for inference. Therefore, more locis and variable sites may offer an opportunity for studies of complex groups mechanisms and history indeed. Due to the short age of locis about 100 base, a bias needs to overcome [43–44].

With the development of sequencing technology, the deficiency was apt to be solved and recovered by more accurate and longer data locis from more population samples. The new sequencing technology will prove more accurate and better phylogeny [13]. For example, the Paired-end Illumina sequencing technology becomes more available and reliable, and it obtains longer sequences and locis from GBS methods [45]. It is also significant to improve our comprehension about the context of variation when the sample genetically differentiated in slight and it had a large geographical distribution. Furthermore, the improved model studies have increasingly shown that a small number of samples in continuous geographic ranges can be misleading evidence that a population does not exist[46]. The species phylogenetic tree inference also suggested that a better and optimal group size can be estimated from more samples [47]. In this study, 7 out of 8 of the species had multiple accessions, and greater sample in species overcame some problems by sequence errors.

Conclusions

According to the study, the next step of the job with Chinese *Curcuma* is to collect more species samples from different geographical distribution, especially *Curcuma kwangsiensis*, and from subgenus *Curcuma* and other subgenus of *Curcuma* in the world [2]. Besides, the use of next-generation sequencing methods such as ddGBS-seq and ddRad-seq can produce longer reads. The transcriptome sequencing methods have succeeded in recovering the phylogeny in recently radiated genus and mining for phylogenetic study [48].

While the step to use new sequencing techniques is expected to improve the reliability of phylogenetic tree analysis and inference methods, these methods need to combine with simulated and authentic data to determine the best method. The *Curcuma* has a complex evolutionary history. In order to make sure the reliability of its relationship and origin, besides depending on advances in sequencing technology, more taxa will play a crucial role.

Methods

Material and sequencing

The total 60 specimens which belong to 8 *Curcuma* species were used in this study, and *Hedychium coronarium* was used as outgroup (Table 1). All species collecting information were obtained in Table 1 and identified by Prof. Ruiwu Yang of Sichuan Agricultural University. All materials used in this study were stored at Sichuan Agricultural University. The genomic DNA isolation was carried out on fresh leaves by CTAB method. The GBS libraries for each sample were prepared using the protocol outline as previously described [24, 49]. The PstI (CTGCAG) as a restriction enzyme was used to digest all DNA samples. The

library produced the DNA fragments that were ligated by a barcode adaptor and the sticky ends were corrected by a common adaptor. After then, a Qiagen MinElute 96-well PCR purification kit was used in the clean-up step to clean up the products. After PCR, the PicoGreen and a qPCR machine were used to examine the quality about the PCR products. All samples were pooled into a single GBS library. The genome research and bioinformatics about the GBS library were carried out on Illumina Hiseq PE150 sequencer on Novogene Bioinformatics Technology

Clustering

The raw reads were performed using the ipyRAD 0.7.29 pipeline [50] from the Illumina FASTQ files. Following seven sequential steps, the ipyRAD pipeline can obtain species or higher variation across clades in clustering and alignment method based on specific parameters in the ipyRAD documentation (<https://github.com/dereneaton/ipyrad>). This is unlike other pipelines [51]. The RAD parameter settings were as follows: Sequences were clustered within samples by 90% similarity via the uclust function in USEARCH [52]. Clusters of less than 10 sequences were discarded and the minimum number of individuals per cluster was set to 5. Any locus that was heterozygous among more than two samples was discarded. The rest of clusters were processed as loci. The phylogenetic matrix was assembled with the loci.

Concatenation-based species tree inference

The dataset used to analyze the phylogeny tree was the high-quality SNPs via quality control of the filtered SNPs \times QC \times by PLINK. The maximum likelihood (ML) and maximum parsimony (MP) method with bootstrap in the software PAUP* 4.0a134 [53–54] were used to generate the phylogenetic tree. The dataset were implemented in RAxML8.2.8 [55], substitution models GTR for datasets used the Akaike Information Criterion (AIC) in MrModeltest [56] after calculation in PAUP* 4.0a134 [54]. The MrBayes v.3.1.2 was used to analyze the dataset based on Bayes theorem [47].

Phylogenetic structure and divergence time inference

Principal component analysis (PCA) is a purely mathematical method that reflects the clustering between groups and is based on the degree of SNPs in difference individual by EIGENSOFT [57]. The 1,295 high-quality SNPs from 60 *Curcuma* sample were obtained by the computer software ADMIXTURE[58]. In this study, 60 *Curcuma* samples were analyzed by STRUCTURE. The kinships (K) value was from 2–12 by ADMIXTURE. BEAST1.7.5 (Bayesian evolutionary analysis by sampling trees) [47] is an efficient and flexible computer software package based on the Bayesian method for molecular sequence mutation evolution analysis. It was the first software package using the relaxed molecular clock to estimate divergence time in the phylogenetic tree [59]. On the basis of the high-quality SNPs, the BEAST was used to estimate Chinese *Curcuma* species divergence time, and the Bayesian tree was dated by setting

divergence time between *Hedychium coronarium* and *Curcuma* as 42.4 Mya (37.4–47.4 Mya)[60]. Substitution model and site heterogeneity model used the optimal model based on the Bayesian analysis to select the model “relaxed clock” and Markov chain Monte Carlo (MCMC) runs 200 million generations, and every 1,000 steps in the samples was to ensure effective sample size (ESS) in each parameter greater than 100. The output file assessed convergence in Tracer1.5 [61]. The phylogenetic tree used TreeAnnotator1.5.3 to discard 25% as burnin. Finally, the divergence time were analyzed and obtained in FigTree1.1.2.

Abbreviations

GBS: genotyping-by-sequencing; QC: quality control; QTP: qinghai–tibet plateau; TCM: traditional chinese medicine; NGS: next generation sequencing; RAD-Seq: restriction-site associated DNA sequencing; SNP: single-nucleotide polymorphism; ML: maximum likelihood; MP: maximum parsimony; PCA: principal component analysis; K: Kinships

Declarations

Acknowledgements

We would like to thank Yang Yang (Novogene Bioinformatics Technology, Tianjing) for assistance in GBS library preparation. We also thank Dr. Zhimeng Wang (College of Life Science, Peking University, Beijing) for helping us run pyRAD and kindly providing additional help in assembly programs. My sincere gratitude also goes to Dr. Keyan Zhang (Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan) for valuable comments for this study.

Authors' contributions

HL and R-WY designed the experiments and analyzed the data. HL wrote the original manuscript. YZ, GG and J-BD collected the plant materials. H-YZ and R-WY identified the materials. X-FY, LZ and C-BD assisted with manuscript preparation. All authors read and approved the final manuscript.

Funding

This work was supported by the R&D Infrastructure and Facility Development Program of Sichuan (NO. 2018TJPT0013).

Availability of data and materials

These plant materials are required for the collection of plant samples. The plant materials are maintained in accordance with the institutional guidelines of the College of Life Sciences, Sichuan Agricultural

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1 College of Life Science, Sichuan Agricultural University, Yaan, 625014, China.2 School of Geography and Tourism, Guizhou Education University, Guiyang, 550018, China.3 College of life sciences and food engineering, Yibin University, Yibin, 644000, China.4 College of Science, Sichuan Agricultural University, Yaan, 625014, China.5 College of Landscape Architecture, Sichuan Agricultural University, Chengdu, 611130, China.6 Triticeae Research institute, Sichuan Agricultural University, Chengdu, 611130, China.

References

- 1.Deng JB, Ding CB., Zhang L, Zhou YH, Yang RW. Relationships among six herbal species (*Curcuma*) assessed by four isozymes. *Phyton*. 2011;80, 181–88.
- 2.Záveská E, Fér T, Šída O, Krak K, Marhold K, Leong-Škorničková J. Phylogeny of *Curcuma* (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus *Ecomata*. *Taxon*. 2012;61, 747–63.
- 3.Chen J, Zhao J, Erickson DL, Xia N, Kress WJ. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. *Mol. Ecol. Resour*. 2015;15, 337–48.
- 4.Záveská E, Fér T, Šída O, Marhold K, Leong-Škorničková J. Hybridization among distantly related species: examples from the polyploid genus *Curcuma* (Zingiberaceae). *Mol. Phylogenet. Evol*. 2016;100, 303–21.
- 5.Leong-Škorničková J, Šída O, Záveská E, Marhold K. History of infrageneric classification, typification of supraspecific names and outstanding transfers in *Curcuma* (Zingiberaceae). *Taxon*. 2015;64, 362–73.

6. Leong-Škorničková J, Šída O, Jarolímová V, Sabu M, Fér T, Trávníček P, Suda J. Chromosome numbers and genome size variation in Indian species of *Curcuma* L. (Zingiberaceae). *Ann. Bot. (Oxford)* 2007;100: 505–26.
7. Kress WJ, Prince LM, Williams KJ. The phylogeny and a new classification of the gingers (Zingiberaceae), Evidence from molecular data. *Am. J. Bot.* 2002;89, 1682–96.
8. Leong-Škorničková J, N.-S, Lý. *Curcuma pambrosima* sp.nov. (Zingiberaceae) from central Vietnam. *Nordic J. Bot.* 2010;28:652–55.
9. Chen J, Xia NH, Pollen morphology of Chinese *Curcuma* L. and *Boesenbergia* Kuntz (Zingiberaceae): taxonomic implications. *Flora* 2011;206,458–67.
10. Newmaster SG, Ragupathy S. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Mol. Ecol. Resour.* 2009;9, 172–80.
11. Blanco-Pastor JL, Vargas P, Pfeil BE. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *Plos One* 2012;7:e39089.
12. Lemmon EM, Lemmon AR, High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Systemat.* 2013; 44:99–121.
13. Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Syst. Biol.* 2016;65:910–24.
14. Godden GT, Jordan-Thaden IE, Chamala S, Crowl AA, Garcia N., Germain-Aubrey, et al. Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. *Plant Ecol. Div.* 2013;5, 427–50.
15. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 2008;3, e3376.
16. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One* 2012;7:e32253.
17. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT, Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 2013;66:526–38.
18. Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. Resolving postglacial phylogeography using high-throughput sequencing. *Proc. Natl. Acad. Sci. U.S. A.* 2010;107, 16196–200.
19. Eaton DAR., Spriggs EL, Park B, Donoghue MJ. Misconceptions of missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Syst. Biol.* 2017;66, 399–412.

20. Fernández-Mazuecos M, Mellers G, Vigalondo B, Sáez L, Vargas P, Glover BJ. Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Syst. Boil.* 2017;67, 250–68.
21. Hauser DA, Keuter A, McVay JD, Hipp AL, Manos PS. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). *Am. J. Bot.* 2017;104, 1581–95.
22. Stansell Z, Hyma K, Fresnedo-Ramírez J, Sun Q, Mitchell S, Björkman T, et al. Genotyping-by-sequencing of *Brassica oleracea* vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Hortic. Res.* 2018;5.38–48.
23. Nemati Z, Blattner FR, Kerndorff H, Erol O, Harpke, D. Phylogeny of the saffron-crocus species group, *Crocus* series *Crocus* (Iridaceae). *Mol. Phylogenet. Evol.* 2018; doi: 10.1016/j.ympev.2018.06.036.
24. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 2011; 6, e19379.
25. Kumar S, Filipowski AJ, Battistuzzi FU, Pond SLK, Tamura K. Statistics and truth in phylogenomics. *Mol. Biol. Evol.* 2012;29, 457–72.
26. Ward JA, Bhangoo J, Fernández-Fernández F, Moore P, Swanson JD, Viola R, et al. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation, *BMC Genomics* 2013;14, 2–2.
27. Zhe KZ, Jian HW. The impact of major geological events on Chinese flora. *Biodiversity science.* 2017;25, 123–125. doi: 10.17520/biods.2016120
28. Zhe KZ, Huang J, Ding WN. The impact of major geological events on Chinese flora. *Biodiversity science.* 2017;25, 123–125. doi: 10.17520/biods.2016120
29. Li J, Fang X. Uplift of the Tibetan Plateau and environmental changes. *Chin. Sci. Bull.* 1998;43, 1569–1574.
30. Shi YF, Li JJ, Li BY, Yao TD, Wang SM, Li SJ, et al. Uplift and environmental changes of Qinghai-Xizang (Tibetan) Plateau in the late Cenozoic. *Acta Geographica Sinica* 1998;10–20. doi: [10.11821/xb199901002](https://doi.org/10.11821/xb199901002).
31. Li JJ. The Qinghai - Tibet plateau uplifting and environmental evolution in Asia: Article collection of academician. Science Press. Beijing, China. 2006;
32. Liu XD, Dong BW. Influence of the Tibetan Plateau uplift on the Asian monsoon-arid environment evolution. *Chin. Sci. Bull.* 2013;58, 2906–2919.
33. Mallet J. 2007. Hybrid speciation. *Nature*, 446, 279–83.

34. Abbott R, Albach D, Ansell S, Arntzen J, Baird S, Bierne N, et al. Hybridization and speciation. *J. Evol. Biol.* 2013;26, 229–46.
35. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, et al. Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 2008;42, 443–61.
36. Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, et al. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* 2013;341, 658–59.
37. Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shores N, Sorenson AL, et al. Polyploidy can drive rapid adaptation in yeast. *Nature*, 2015;519, 349–52
38. Escudero M, Eaton DAR, Hahn M, Hipp AL. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex*(Cyperaceae). *Mol. Phylogenet. Evol.* 2014;79, 359–67.
39. Lee JY, Joseph L, Edwards SV, A species tree for the Australo-Papuan fairy-wrens and allies (Aves: Maluridae). *Syst. Biol.* 2011;61, 253–71.
40. Sukumaran J, Knowles LL, Multispecies coalescent delimits structure, not species. *Proc. Natl. Acad. Sci. U.S. A.* 2017;114, 1607–12.
41. Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O., Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 2017;8, 14363.
42. Martin CH, Cutler JS, Friel JP, Denning TC, Coop G, Wainwright PC. Complex histories of repeated gene flow in cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution* 2015;69, 1406–22.
43. Eaton DAR, Ree RH, Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Syst. Biol.* 2013;62, 689–706.
44. Hipp AL, Manos PS, Gonzalez-Rodriguez A, Hahn M, Kaproth M, McVay JD, et al. Sympatric parallel diversification of major oak clades in the Americas and the origin of the Mexican species diversity. *New Phytol.* 2018;217, 439–452.
45. Lemmon AR, Lemmon EM. High throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Syst. Biol.* 2012;61, 745–61.
46. Gompert Z, Buerkle CA. What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evol. Appl.* 2016;9, 909–23.
47. Heled J, Drummond AJ, Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 2010;27, 570–80.

- 48.Valderrama E, Richardson JE, Kidner CA, Madriñán S, Stone GN. Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L. F., Zingiberaceae). *Mol. Phylogenet. Evol.* 2018;119: 13–24.
- 49.Grabowski PP, Morris GP, Casler MD, Borevitz JO. Population genomic variation reveals roles of history, adaptation and ploidy in switchgrass. *Mol. Ecol.* 2014;23:4059–073.
- 50.Eaton DA, Overcast I. ipyrad: interactive assembly and analysis of RADseq data sets. 2016.Available at: <http://ipyrad.readthedocs.io/>.
- 51.Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks:building and genotyping loci de novo from short-read sequences. *G3* 2011;1, 171–182.
- 52.Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26, 2460–461.
- 53.Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39, 783–91.
- 54.Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version b4. Sinauer Associates, Sunderland, Massachusetts. 2002.
- 55.Stamatakis. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2014;22, 2688–690.
- 56.Nylander JAA. MrModelTest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. 2004.
- 57.Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 2006;38, 904–09.
- 58.Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 2011;28: 2239–252.
- 59.Pybus OG. Relaxed phylogenetics and dating with confidence. *Public Library of Science*, 2006; 4e:88.
- 60.Zheng ML. A phylogenetic study on the tribe Zingibereae (Zingiberaceae). Ph.D. dissertation. Xishuangbanna Tropical Botanical Garden Chinese Academy of Sciences. Yunan, CN. 2010.
- 61.Drummond A, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 2007;7, 214.

Tables

Table 1

Sample of species names, with source details and total loci after processing with ipyRAD, of the 61 samples used in phylogenetic analyses.

Sample	Species	Source details	Total of loci after ipyrad
SC1	<i>Curcuma longa</i>	Muchuan, Sichuan, cultivated	26,977
SC3	<i>Curcuma longa</i>	Fulu, Sichuan, wild	17,898
SC4	<i>Curcuma longa</i>	Cuiping, Sichuan, wild	39,772
SC19	<i>Curcuma longa</i>	Yibin, Sichuan, wild	30,974
SC18	<i>Curcuma longa</i>	Qianwe, Sichuan, wild	26,311
GZ3	<i>Curcuma longa</i>	Xingyi, Guizhou, wild	28,912
SC20	<i>Curcuma longa</i>	Shuangliu, Sichuan, cultivated	32,886
YN11	<i>Curcuma longa</i>	Menga, Yunnan, wild	20,606
SC5	<i>Curcuma longa</i>	Dayi, Sichuan, cultivated	34,053
SC21	<i>Curcuma sichuanensis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	21,536
SC22	<i>Curcuma sichuanensis</i>	Chongzhou, Sichuan, cultivated	33,411
YN29	<i>Curcuma sichuanensis</i>	Yiwu, Yunnan, cultivated	19,503
SC2	<i>Curcuma sichuanensis</i>	Sanjiang, Sichuan, cultivated	26,973
SC23	<i>Curcuma sichuanensis</i>	Cuiping, Sichuan, wild	24,230
SC6	<i>Curcuma sichuanensis</i>	Ziyang, Sichuan, cultivated	27,677
YN10	<i>Curcuma sichuanensis</i>	Menga, Yunnan, cultivated	38,515
SC24	<i>Curcuma sichuanensis</i>	Chongzhou, Sichuan, cultivated	31,072
SC25	<i>Curcuma sichuanensis</i>	Qianwei, Sichuan, cultivated	20,902
SC26	<i>Curcuma sichuanensis</i>	Muchuan, Sichuan, cultivated	21,588
SC11	<i>Curcuma sichuanensis</i>	Weiyuan, Sichuan, cultivated	28,768
SC8	<i>Curcuma sichuanensis</i>	GAP Base, Sichuan, cultivated	30,753
SC9	<i>Curcuma sichuanensis</i>	Sanjiang, Sichuan, wild	28,621
SC10	<i>Curcuma sichuanensis</i>	Chongzhou, Sichuan, cultivated	23,043
SC7	<i>Curcuma sichuanensis</i>	Yibing, Sichuan, wild	28,068
YN15	<i>Curcuma sichuanensis</i>	Menglun, Yunnan, cultivated	13,089
GX9	<i>Curcuma sichuanensis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	31,039
YN32	<i>Curcuma amarissima</i>	Mengkang, Yunnan, wild	16,865
YN7	<i>Curcuma amarissima</i>	Menga, Yunnan, wild	17,222
YN8	<i>Curcuma amarissima</i>	Menga, Yunnan, wild	17,923
GD4	<i>Curcuma yunnanensis</i>	Huaxian, Guangdong, wild	32,976
GD5	<i>Curcuma yunnanensis</i>	Huaxian, Guangdong, wild	24,819
GX10	<i>Curcuma phaeocalis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	23,145
YN27	<i>Curcuma phaeocalis</i>	Yiwu, Yunnan, wild	22,636
SC14	<i>Curcuma phaeocalis</i>	Shuangliu, Sichuan, cultivated	34,212
GX11	<i>Curcuma phaeocalis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	29,557
SC13	<i>Curcuma phaeocalis</i>	Jianwei, Sichuan, cultivated	25,208
GX4	<i>Curcuma phaeocalis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	35,825
GZ2	<i>Curcuma phaeocalis</i>	Anlong, Guizhou, wild	31,112
YN13	<i>Curcuma phaeocalis</i>	Menga, Yunnan, wild	14,425
YN6	<i>Curcuma phaeocalis</i>	Mengxing, Yunnan, wild	21,440
YN31	<i>Curcuma phaeocalis</i>	Mengkang, Yunnan, wild	21,642
SC15	<i>Curcuma phaeocalis</i>	Shuangliu, Sichuan, cultivated	13,089
YN21	<i>Curcuma phaeocalis</i>	Xishuangbanna, Yunnan, wild	16,744
SC12	<i>Curcuma phaeocalis</i>	Chongzhou, Sichuan, cultivated	18,289
YN19	<i>Curcuma phaeocalis</i>	Menglun, Yunnan, wild	13,489
YN16	<i>Curcuma phaeocalis</i>	Dalu, Yunnan, wild	26,946
YN17	<i>Curcuma phaeocalis</i>	Xishuangbanna, Yunnan, wild	23,180
YN28	<i>Curcuma phaeocalis</i>	Yiwu, Yunnan, wild	28,065
YN4	<i>Curcuma kwangsiensis</i>	Medicinal, Botanical Garden, Guangxi, cultivated	17,001
GX14	<i>Curcuma aromatica</i>	Medicinal, Botanical Garden, Guangxi, cultivated	26,739
YN5	<i>Curcuma aromatica</i>	Mengkang, Yunnan, wild	24,435
SC16	<i>Curcuma aromatica</i>	Jianyang, Sichuan, wild	26,817
YN12	<i>Curcuma aromatica</i>	Yiwu, Yunnan, wild	18,289
ZJ1	<i>Curcuma wenyujin</i>	Taoshan, Zhejiang, cultivated	23,259
ZJ2	<i>Curcuma wenyujin</i>	Meiyu, Zhejiang, wild	27,430
GX12	<i>Curcuma wenyujin</i>	Medicinal, Botanical Garden, Guangxi, cultivated	18,273
GD2	<i>Curcuma wenyujin</i>	Sanshui, Guangdong, cultivated	17,080
GX1	<i>Curcuma wenyujin</i>	Medicinal, Botanical Garden, Guangxi, cultivated	22,593
GX3	<i>Curcuma wenyujin</i>	Medicinal, Botanical Garden,	28,736

Figures

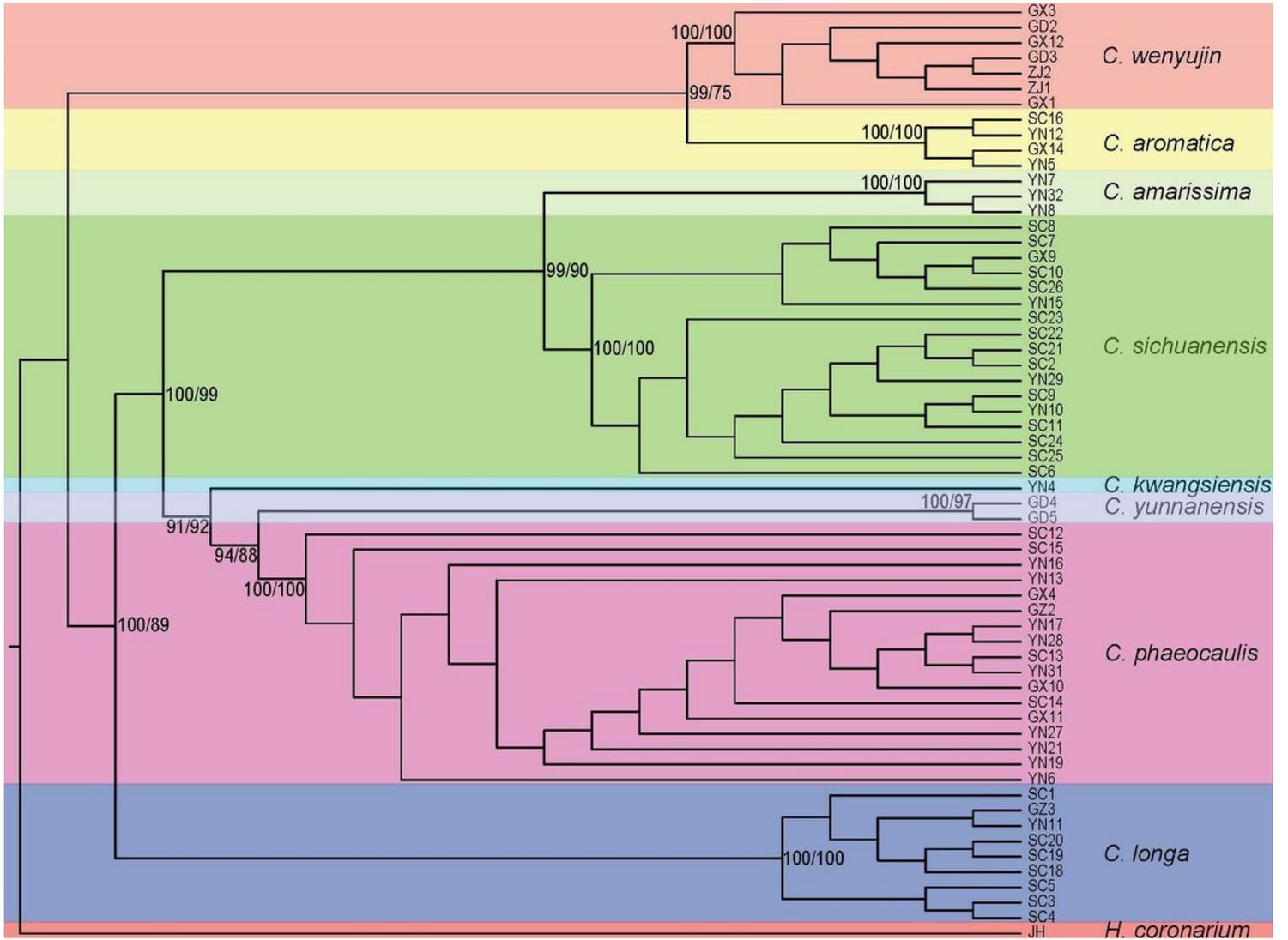


Figure 1

Phylogeny tree of Chinese *Curcuma* resulting from maximum likelihood and maximum parsimony (ML bootstrap values/MP bootstrap values). The phylogeny tree shows all 61 samples in this study, different species with different colors. The bootstrap values (maximum likelihood bootstrap values/maximum parsimony bootstrap values) in every species clade.

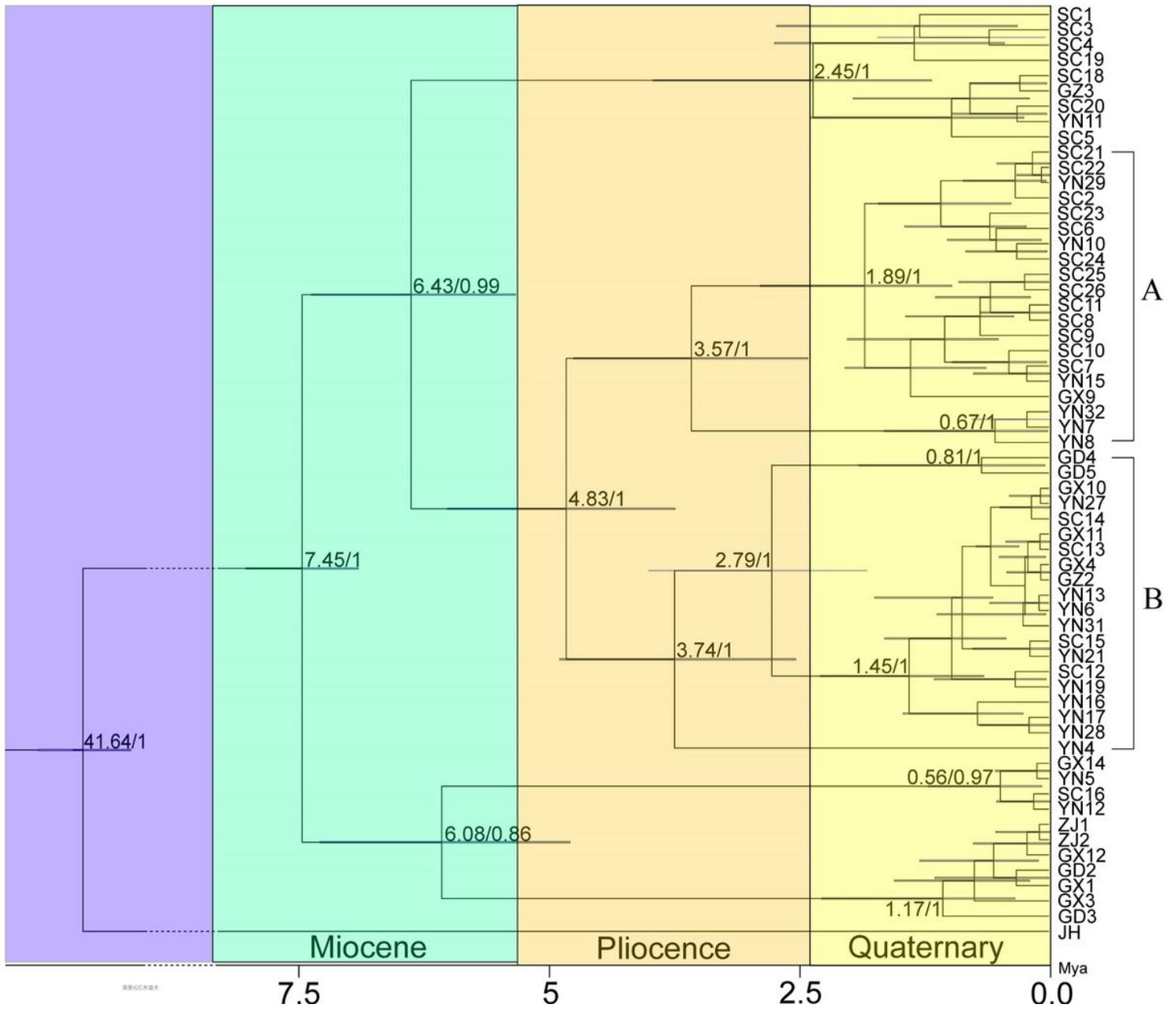


Figure 2

Phylogeny tree resulting from the Bayesian inference and divergency time of species clade in BEAST (Divergence time/Bayesian support). The A lineages was consisted of *C. amarissima*, *C. sichuanensis*, *C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis*. And B was made up (*C. kwangsiensis*, *C. yunnanensis* and *C. phaeocaulis*. Different geologic age with different color.

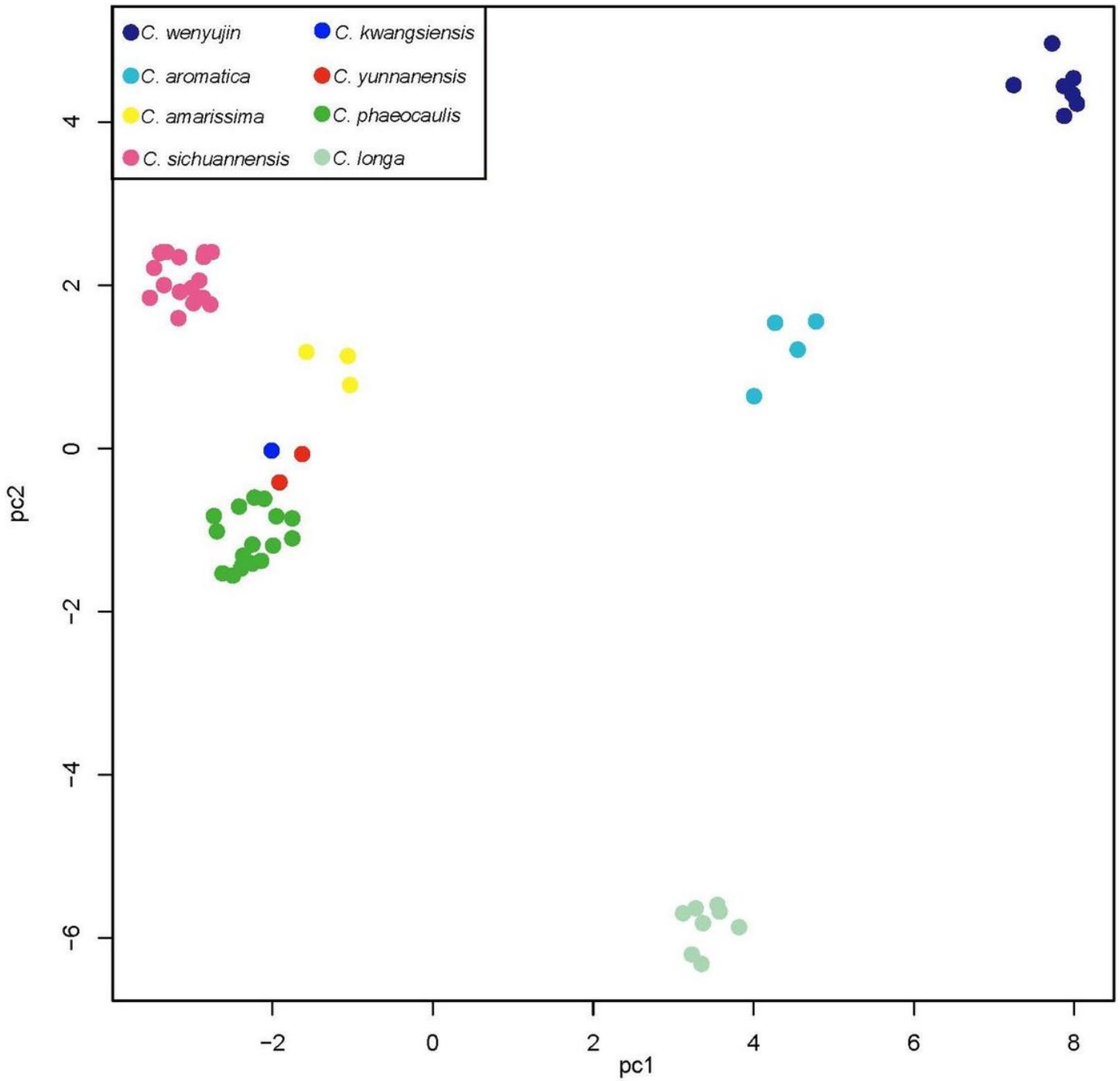


Figure 3

The principle components analysis of Chinese Curcuma. Principal component analysis (PCA) of the 60 samples.

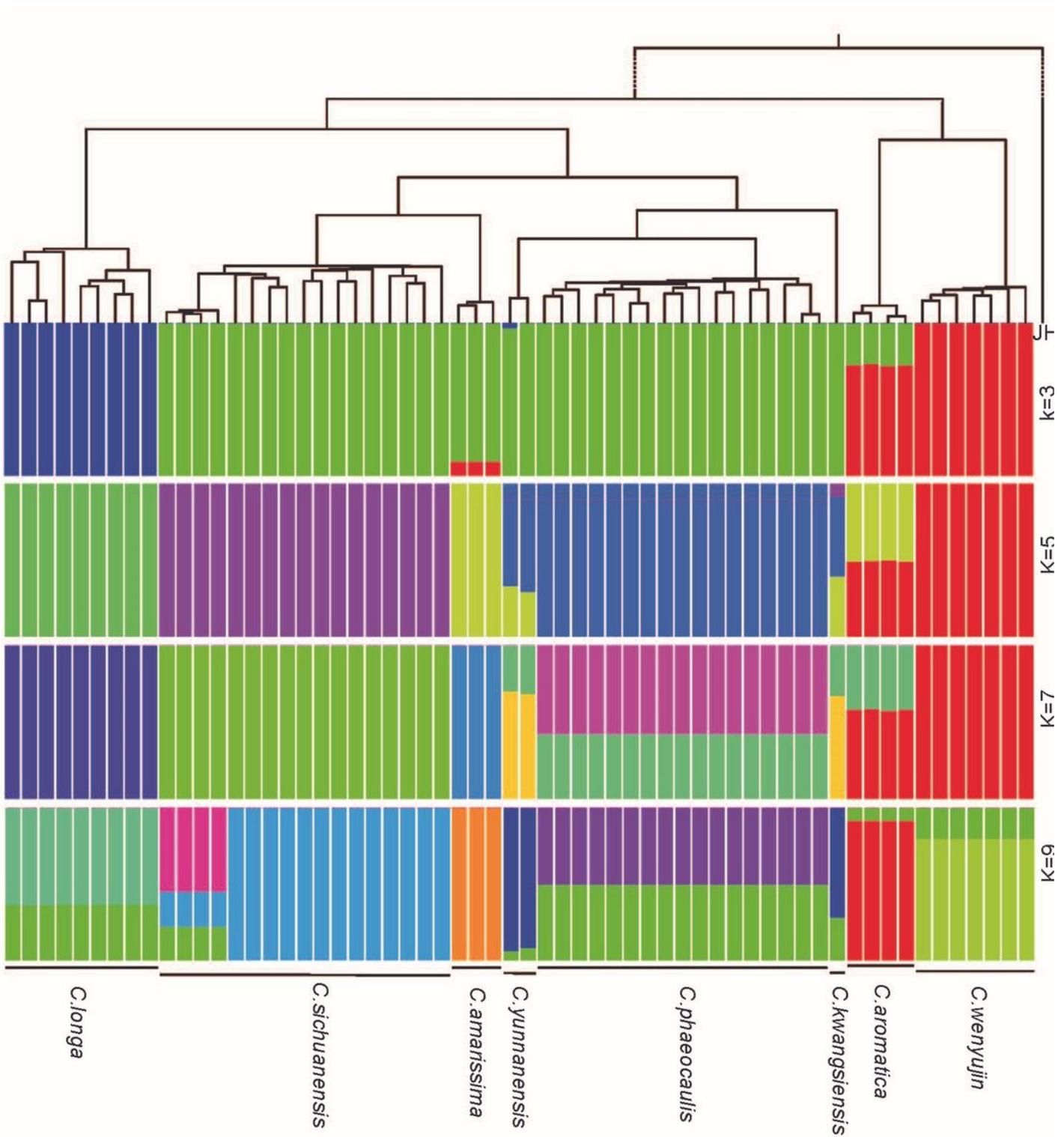


Figure 4

Population genetic structure of Chinese Curcuma. When the kinship (K) = 3, 5, 7 and 9, each vertical bar represents a Curcuma sample and different color represents different ancestral background in putative.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.pdf](#)