

Genetic diversity and population structure revealed by Whole Genome Resequencing of Lushan Cloud Mist Tea (*Camellia sinensis* var. *sinensis*)

Yanli Niu

Dongzhou Xia

Yansong Peng

Zhongxin Duan

Deshui Yu

Xinfeng Jiang

Binsheng Luo

luobins@lsbg.cn

Research Article

Keywords: Lushan Tea, Genetic Diversity, Whole Genome Resequencing, Genetic Structure, Demographic History, SIFT

Posted Date: February 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8748651/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Lushan Cloud Mist Tea is celebrated as one of China's top ten famous teas, distinguished by its unique climate and geographical environment. Our analysis revealed varying levels of genetic differentiation among populations, with the most significant divergence observed between the mid-altitude JDX and low-altitude TYC groups. The low-altitude FSZ group exhibited high nucleotide diversity, suggesting a rich reservoir of genetic resources. In contrast, the low-altitude TYC group displayed comparatively low diversity, making it more susceptible to genetic drift. Additionally, the effective population size of the nine studied populations showed a degree of consistency yet indicated an increasing trend in modern times. For SIFT analysis, sift score showed the most amino acid was under TOLERATED group, indicated that the genome for this tea species was relatively stable. Overall, our findings provide essential insights into the genetic diversity and structure of Lushan Cloud Mist Tea, informing evidence-based strategies for germplasm conservation and utilization.

Introduction

The tea plant (*Camellia sinensis* L.), a pivotal perennial species within the Theaceae family, holds considerable importance as a beverage crop on a global scale. The germplasm resources of the tea plant are crucial for the sustainable development of the tea industry. These resources serve as the cornerstone for comprehensive research, the advancement of resource development, and the strategic utilization of tea plant varieties. Furthermore, they are integral to maintaining the adaptability and long-term survival of tea plant populations. In light of these factors, the examination of genetic diversity and population structure has become increasingly critical, serving as the basis for the development of informed scientific conservation and effective resource utilization approaches.

It remains uncertain about the origin center of *C. sinensis*. Previous research indicated this species was originated in the Jiangnan Tea Region of China, or Central China. Cloud tea tends to grow in the perennial cloud shrouded mountains. For morphology, its leaves are in an elliptical or obovate-lanceolate shape. The edges of the leaves are serrated, which is a typical characteristic. The leaf color is a vibrant green, with a shiny and lustrous surface, indicating the high quality of the tea leaves. The tender leaves may have fine hairs on the back, which gradually shed as the leaves mature.

Molecular marker technologies have been fundamental in understanding tea plants' genetic diversity. For example, Luo et al. (2002) used RAPD markers to elucidate the phylogenetic relationships among 31 distinct tea plant ecotypes in China (Luo et al. 2002). Similarly, Mishra et al. (2009) utilized AFLP markers to investigate the genetic diversity of Darjeeling tea cultivars in India (Mishra et al. 2009). Additionally, Tan et al. (2009) utilized SSR markers to evaluate the genetic variations within 43 tea plant germplasms, uncovering substantial diversity (Tan et al. 2009). Despite the invaluable insights gained from these traditional molecular approaches, their scope in resolving phylogenetic and evolutionary nuances among tea plants remains limited. In contrast, more recent studies have adopted Genotyping-by-Sequencing (GBS) technology, offering a more comprehensive perspective on the genetic diversity of local tea plant germplasm resources and marking a significant progression in the field (Niu et al. 2019; Huang et al.

2022). However, traditional molecular markers lack the accuracy and resolution to accurately find genetic variation at the whole genome level, and it is difficult to find new variants and genes.

Whole-genome resequencing has revolutionized the field of genomics, capturing the complete evolutionary history of species and providing a framework for constructing robust phylogenetic relationships. Its applications extend across plant systematics, population genetics, and evolutionary studies. For example, Qi et al. (2013) leveraged this technology to untangle the evolutionary lineage of cucumber varieties (Qi et al. 2013), and Duan et al. (2017) applied it to shed light on the genetic history of apple cultivars (Duan et al. 2017). The technique has been indispensable in elucidating the genetic makeup of species like watermelon (Guo et al. 2019), grape (Liang et al. 2019), and cannabis (Ren et al. 2021), respectively. Recently, whole-genome resequencing has emerged as a key tool in conserving endangered plant species by offering unprecedented insights into their genetic structures. It is particularly impactful in tea plant genetics research, enhancing our comprehension of the genetic variation and evolutionary dynamics inherent within *Camellia sinensis* species.

The tea plant, a perennial crop of economic significance, displays a wide range of genetic diversity shaped by human cultivation and natural cross-pollination. Despite this, the complexity of high heterozygosity and abundant repetitive sequences has historically impeded research. The seminal publication of the Yun Kang No. 10 tea plant genome heralded a transformative phase in tea genomics (Xia et al. 2017). This was further bolstered by the subsequent genome release of the cultivar Shuchazao (Xia et al. 2020) and the chromosomal-level assembly of reference genomes for cultivars such as Bi Yun (Zhang et al. 2020a), wild DASZ (Zhang et al. 2020b), and Longjing 43 (Wang et al. 2020). Recent advancements in whole-genome resequencing have markedly clarified the genetic diversity and evolutionary trajectory of tea plant varieties. Wang et al. (2020) delineated the evolution of domesticated tea plants (Wang et al. 2020), while An et al. (2020) delved into the genetic diversity and adaptive evolution across different varieties (An et al. 2020). These endeavors have not only enriched our understanding of tea plant evolution but also opened new avenues for enhancing tea breeding programs.

Lushan Cloud Mist Tea is acclaimed as one of China's top ten teas, owing its reputation to the unique climate and geography of its origin. For a long time, Lushan tea has been predominantly propagated via asexual reproduction, yet the genetic consequences of this practice on these distinct geographic tea populations remain unexplored. As genetic diversity forms the cornerstone for species evolution and adaptation, preserving this variability is pivotal for the long-term survival of the species. Hence, this study employs whole-genome resequencing to identify SNP sites, laying the groundwork for analyzing the genetic diversity of Lushan tea resources. Our analysis provides a preliminary glimpse into the extent of genetic variation in Lushan tea, illuminates genetic relationships across varying altitudes and geographical locales in Lushan, and reveals traits of their geographic genetic makeup. These findings deliver theoretical support for the conservation and harnessing of Lushan tea germplasm, facilitating the mining of superior alleles and the selection of prime tea cultivars.

Materials and Methods

Sampling, DNA extraction and sequencing

From March to May 2020, We collected 90 samples from eight sampling points across three different altitudinal gradients in the Lushan Scenic Area, namely DTC, DYS, FSZ, NWS, TH, XRD, JDX and TYC (Fig. 1, Table 1). Upon collection, we promptly froze the leaves in liquid nitrogen. Subsequently, we extracted total genomic DNA using a DNA kit (Tiangen, Beijing, China). Next, we prepared a DNA sequencing library using the TruSeq DNA PCR-free prep kit from Illumina. Finally, we sequenced the library with an insert size of 400 bp on the BGI T7 platform by paired-end (PE) with a read length of 150 base pair (bp) long (Jierui Biotech, Guangzhou).

Reads mapping and SNP calling

Using the software Fastp (version: 0.24.0) (Chen et al. 2018), the raw reads were subject to adapter removal and quality filtering with default parameters. We used the chromosome-level assembled genome of cultivar Longjing 43 (*Camellia sinensis* var. *sinensis* cv. Longjing 43) as the reference (<http://tpia.teaplant.org/download.html>). The genome contains 15 chromosomes and other 1318 contigs. The clean reads were mapped to this reference genome using the BWA-MEM (version: 0.7.19-r1273) program with default parameters (Li & Durbin 2009). The alignment outputs were sorted and indexed with SAMtools (version: 1.20) (Li et al. 2009), and duplicated reads were marked by Picard (version: 3.0) (<http://broadinstitute.github.io/picard/>). Base quality recalibration and local realignment were performed using the Genome Analysis Toolkit (GATK: v4.6.2.0) (McKenna et al. 2010). SNPs were called simultaneously on all samples by GATK Unified Genotyper. After joint calling, we removed InDels (Insertion and Deletion) firstly using Vcftools (version: 0.1.16) (Danecek et al. 2011), and then it was used for SNP filtration with criteria: min depth ≥ 4 , max depth ≤ 100 , min Q value ≥ 30 , and min GQ ≥ 2 . At last we filtered SNPs using `--thin 20` parameter.

Genetic diversity and population structure analysis

Using the filtered SNP datasets, the genetic diversity metrics, including heterozygosity (H_e , H_o), nucleotide diversity estimator of π (Nei & Li 1979), and *private alleles* were implemented using the *populations* module from Stacks (version: 2.68) (Catchen et al. 2013). Principal component analysis was performed using GCTA software (version: 1.94.3) (<http://www.complextaitgenomics.com/software/gcta/>) with all SNPs. Phylogenetic trees were constructed based on the maximum likelihood model using iqtree2 software (version: 2.4.0) (Minh et al. 2020), with the reliability of tree branches verified through bootstrapping (1000 repetitions). We also conduct further SNPs filtering to do genetic structure analysis. Here we removed sites harbored allele frequency less than 0.05 and p value of Hardy-Weinberg that was smaller than 0.0001 using plink software (version: 1.9.0 beta 8) (Purcell et al. 2007). The genetic structure of Lushan tea populations was analyzed using Admixture software (version: 1.3.0) (Alexander et al. 2009), we set K values from 1 to 15. The optimal number of clusters was determined based on the Cross-validation error.

Demographic history inference based on genome SNP data

Using a sequentially Markovian coalescent method implemented in SMC++ (version: v1.15.4) with default parameters (Terhorst et al., 2017), the historical effective population size for all samples was inferred based on whole-genome sequence data. Initially, we obtained VCF files for the population analysis and then converted these files into SMC format. This conversion facilitated the generation for figure plots. Finally, plots were conducted with a generation time of five years and a mutation rate per generation per site of 6.5×10^{-9} (Zhang et al., 2021).

Predicting effects of amino acid substitutions on proteins

To assess whether amino acid substitutions have an impact on protein function and evaluate the harmfulness of genetic variations. We conducted SIFT (Sortig Intolerant From Tolerant) analyses using default parameters based on all SNPs (Sim et al., 2012).

Results

Population genomic data and Genetic Diversity

A total of 287.8367 Gb of high-quality reads were obtained after filtering from the raw bases, amounting to 293.5857 Gb for the 90 collected samples. An average mapping rate of 99.72% (97.67% to 99.88%) was identified. These samples have an average repeat sequence ratio of 5.4599% and an average GC content of 39.46%. The average sequencing depth for these 90 samples were 5.7674x. (Supplementary Table 1).

A total of 2,338,431 SNPs were identified across the eight populations. Approximately above 90% of these SNP sites are located in intergenic regions, with a nearly equal ratio of nonsynonymous to synonymous (Non-syn/Syn) substitutions. An analysis of the SNP site distribution across chromosomes revealed the highest number of SNPs on Chromosome 1 and the highest density on Chromosome 9. Conversely, the lowest number of SNPs was found on Chromosome 15, with the lowest density on Chromosome 6 (**see Fig. 2**). The results of the population genetic diversity analysis are presented in Table 2. The observed heterozygosity of the Lushan tea resources ranges from 0.12142 to 0.13159, averaging 0.12533, while nucleotide diversity (Pi value) ranges from 0.11046 to 0.11995, with an average of 0.11503. Notably, the FSZ population exhibits the highest nucleotide diversity and haplotype diversity, indicating the greatest genetic diversity in this group. In contrast, the TYC population show lower values in both nucleotide and haplotype diversity, suggesting reduced genetic diversity within this group, likely influenced by habitat disturbance as they are situated in core tourist areas. Furthermore, the genetic differentiation coefficient (Fst value) between the TYC and JDX populations is greater than 0.04 (Table 3), indicating a substantial degree of differentiation between them. Conversely, the Fst values between other populations range from 0.022119 to 0.039913, suggesting moderate genetic differentiation among them.

Table 2
Genetic diversity parameters of eight *C. sinensis* populations

Population	Obs Het	Obs Hom	Exp Het	Exp Hom	Pi	Private
DTC	0.12834	0.87166	0.11220	0.88780	0.11726	18224
DYS	0.12315	0.87685	0.10966	0.89034	0.11355	32303
FSZ	0.13159	0.86841	0.11375	0.88625	0.11995	13054
NWS	0.12552	0.87448	0.10988	0.89012	0.11382	41243
TH	0.12172	0.87828	0.10747	0.89253	0.11334	16342
XRD	0.12711	0.87289	0.11290	0.88710	0.11759	28917
JDX	0.12376	0.87624	0.10834	0.89166	0.11425	18773
TYC	0.12142	0.87859	0.09941	0.90059	0.11046	5838

Table 3
Fixation indices(Fst) among eight *C. sinensis* populations

DTC	DYS	FSZ	NWS	TH	XRD	JDX	TYC
DTC	0.026689	0.024839	0.026167	0.031640	0.022119	0.029098	0.037983
DYS		0.025316	0.029749	0.024540	0.029162	0.027122	0.028610
FSZ			0.028406	0.031390	0.024921	0.030048	0.038850
NWS				0.034128	0.026103	0.031481	0.039913
TH					0.033780	0.034208	0.038635
XRD						0.030080	0.039651
JDX							0.042547

Phylogenetic tree construction of Lushan tea populations based on SNPs

Based on the all the SNPs, a maximum likelihood tree consisting of all samples from eight populations was obtained with high bootstrap values (Fig. 3). For the phylogenetic tree, the NWS, and DYC populations were two monogroups, and the remaining six groups were intermingled. The samples collected from DTC population formed 3 separate groups. This dispersion pattern may be related to geographical locations and topography (as Jiandao Gorge forms a curved valley), coupled with the dual natural and cultural attributes of Lushan Mountain. Tourism activities and other human interventions could also influence gene flow among the samples.

Population Principal Component Analysis

Principal component analysis (PCA) was used to further categorize the 90 samples. PC1 accounts for 2.76% of the variance, while PC2 accounts for 1.82%. As depicted in Fig. 4, the NWS and XRD populations were significantly distinguished from TH, XRD, and DYS populations along the PC1 axis, indicating substantial genetic differences between them. PC2, on the other hand, separates the NWS population from the other seven populations. The samples from the FSZ, DYS and XRD populations are closely clustered together on both PC1 and PC2, suggesting a close genetic relationship between these two populations. There is an overlapping area between the samples of the JDX and THG populations, indicating gene flow between these groups. The PCA results also show that the JDX population was highly dispersed, which is consistent with the phylogenetic tree analysis results. This suggests extensive gene flow between the JDX population and other groups.

Population Structure Analysis

The genetic structure of a population is influenced by multiple factors, including mutations, selection, migration, population size, and environmental conditions. Analyzing the genetic structure of populations helps in understanding the evolutionary processes of species. When $K = 2$, the cross-validation error rate (CV error) is minimal (see **Supplementary Fig. 1**). However, integrating results from PCA and phylogenetic tree analysis, a comprehensive analysis of the genetic structure of the six Lushan tea populations was conducted. Thus, genetic structure plots for $K > 1$ were created. According to the Admixture results (Fig. 5), most individuals exhibit a singular genetic background. At $K = 2$, the genetic structure of the DYS, and TH populations show nearly complete homozygosity without gene flow with other groups, indicating its distinct structure, whereas other populations demonstrate genetic admixture, suggesting extensive gene flow between populations. At $K = 3$, except for the MWS population, other population samples exhibit a common ancestral signal. At $K = 4$, JDX and TYC populations share a common ancestor. At $K = 5$ and $K = 6$, many samples show a relatively pure genetic background, with very few being mixed; there is frequent gene flow between populations, with each population displaying both similarities and genetic variations in their genetic backgrounds.

Population demographic history analysis

Our research results indicate that the effective population size has been fluctuating since approximately 20,000 years ago. The greatest decline in population size occurred around 2,000 years ago (see Fig. 6). This pattern holds true for all eight of the populations studied. This trend aligns with the ice age glacial cycle fluctuations, and Lushan is a diverse refuge.

SIFT analyses

For a variant site, a specific score will be given in the end ranging from 0 to 1. When it fall on 0-0.05, the mutation indicates that it has caused a change in the protein structure and is therefore harmful. The closer the score is to 1, the smaller the impact of the mutation on the protein.

The results indicate that the majority of the mutations at these sites are non-synonymous mutations. Each type of mutation was showed in Fig. 7. And among these non-synonymous mutations, the majority of the sites are harmful (Fig. 8).

Discussions

As of January 2021, a total of 388 whole genome sequences of plants have been globally published, providing extensive genomic information for various species. Whole genome sequencing yields a wealth of SNP and variation data, enabling in-depth studies on species characteristics, population evolution, trait gene localization, and more. Tea plants, known for their intricate taxonomy and genetics, pose significant challenges due to their complex genetic backgrounds resulting from cross-breeding among samples, unlike annual or perennial self-incompatible crops.

In biological populations, genetic variation correlates with the rate of evolution. Characterizing genetic diversity unveils a species' population evolutionary history and enables the evaluation of its evolutionary potential. Parameters like H_o , P_i , and Haplotype Diversity offer multifaceted insights into a population's genetic diversity and potential. Higher values within a certain range indicate greater gene richness, genetic diversity, and evolutionary potential. Therefore, understanding the genetic background of germplasm resources is crucial.

In this study, the nearly equal ratio of nonsynonymous mutations to synonymous mutations aligns with An et al.'s (2020) findings (An et al. 2020), where whole genome resequencing revealed a similar pattern across all cultivars. Additionally, Lu et al.'s (2021) sequencing of 120 strains from 8 ancient tea tree populations showed that the variation in ancient tea trees remained unaffected by external natural environmental pressures or artificial breeding, evidenced by a nonsynonymous mutations/synonymous mutations ratio of 1.05.

Whole genome resequencing enhances our understanding of genetic variation and adaptive characteristics (Jing et al., 2023). Lei et al. (2022) also used whole genome sequencing technology to study Yunnan Lincang tea, shedding light on the genetic composition and evolutionary characteristics of the local tea populations (Lei et al. 2022). High-throughput sequencing technology provides a large number of molecular markers for tea plant genetic analysis and serves as a potent tool for studying population genetic patterns. This study, which is conducting a phylogenetic analysis based on whole genome resequencing, elucidates genetic diversity and structure among Lushan tea resources. Notably, the JDX population, situated in a remote location, exhibits the highest nucleotide diversity and richest genetic diversity, while the TYC and TH populations, located in the core tourist areas, display the lowest nucleotide diversity due to potential long-term artificial selection and breeding. These findings align with the results of the phylogenetic tree, principal component analysis, and population structure analysis. The phylogenetic tree demonstrates the wide distribution of DTC population individuals across all three branches, corroborated by scattered DTC population samples on both PC1 and PC2 in PCA analysis. Similarly, the population genetic structure graph indicates the NWS population's complex genetic origin

and richer genetic resources, confirming extensive individual and gene flow among NWS population samples.

The F_{st} genetic differentiation coefficient unveils notable genetic differentiation between the JDX and TYC populations, likely influenced by the core tourist area's location. Conversely, the TH and DYS populations exhibit minimal genetic differentiation, attributed to their relatively less human interference and complex geographical locations. Evidently, changes in the genetic structure of Lushan tea resources are associated with genetic drift and selection.

The Lushan National Scenic Area, a popular tourist destination, experiences a peak tourist season during the National Day holiday coinciding with the prime flowering period of tea plants. The majority of tourist activities are concentrated in the mid-altitude regions (DTC, XRD, and JDX), leading to increased exchanges between tea individuals in these areas. The genetic diversity results indicate that the inbreeding coefficient F_{is} values are lowest in the low-altitude NWS and FSZ populations, signifying minimal inbreeding and robust gene flow with samples from other populations.

Observations of heterozygosity also support these findings, revealing that mid-altitude regions have higher values than high-altitude (DYS) and low-altitude (TYC) populations. Furthermore, cluster analysis, principal component analysis, and population structure analysis demonstrate more distant relationships between samples in the mid-altitude JDX population compared to other populations. Similarly, samples from the high-altitude DYS population and low-altitude THG population exhibit closer relationships within their respective populations than with others. These results align with findings, indicating that increasing hybridization with the popularization of tea cultivation enhances heterozygosity and gene flow between tea populations (Feng et al. 2020).

Tea trees, known for their large genomes, high heterozygosity, and species diversity, hold significant economic importance. In the case of Lushan tea resources, asexual propagation is common, allowing for the preservation of superior genotypes and the swift breeding of new varieties. However, extensive asexual propagation can lead to a loss of genetic diversity and the accumulation of detrimental mutations, potentially impacting tea quality and the tree's ability to adapt to the environment. This can be verified from the results of our analysis of harmful mutations. Among the 31,834 non-synonymous mutation sites, 10,444 of them are harmful mutations, accounting for nearly one-third. This proportion is extremely high, indicating that there are significant challenges regarding its living environment.

Hou et al. (2023) conducted high-depth whole genome resequencing of Tieguanyin mother trees and their asexually propagated progeny, revealing accumulated mutations in adversity-related genes across multiple generations of asexual propagation. Whole genome resequencing proves crucial in identifying genomic variations and predicting hybrid performance for effective breeding strategies (Liu et al., 2021). This study's detection of variations in Lushan tea resources through whole-genome resequencing highlights the importance of expanding the genetic variation of new varieties by leveraging beneficial genes from wild tea, paving the way for innovative germplasm, trait-related gene mining, and improved variety breeding in the future of tea cultivation.

Declarations

Funding

This research was supported by Jiangxi provincial Key Laboratory of Plantation and High Value Utilization of Specialty Fruit Tree and Tea (20241ZDD02045), the Talents Program of Jiangxi Province (PR China) (jxsq2020104003), the Modern Agricultural Industrial Technology System of Jiangxi Province (JXARS-06), National Natural Science Foundation of China (32460785), and the “Xuncheng Talents” Program of Jiujiang City (JJXC2023136).

Author contributions statement

XF. J., and BS. L. conceived and designed the study and served as corresponding authors. YL. N., DZ. X., and YS. P. conducted the field sampling and collected tea plant materials. ZX. D., and DS. Y. performed the sequencing experiments and data acquisition. YL. N., and DZ. X. carried out the bioinformatic analyses and interpreted the sequencing data. YS. P. contributed to data validation and result visualization. BS. L. provided overall project supervision and resources. YL. N., and BS. L. drafted the manuscript. All authors reviewed, revised, and approved the final manuscript.

Data availability statement

The raw data sequenced in this research was deposited in NCBI database with a Project NO. PRJNA1110598 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1110598>) and the SRA NO. was between SRX24527752-SRX24527781, SRX30756424-SRX30756483, respectively.

References

1. Alexander DH, Novembre J, Lange, KJGr (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664
2. An Y, Mi X, Zhao S, Guo R, Xia X, Liu S, Wei C (2020) Revealing Distinctions in Genetic Diversity and Adaptive Evolution Between Two Varieties of *Camellia sinensis* by Whole-Genome Resequencing. *Front Plant Sci* 11:603819
3. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
4. Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890
5. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
6. Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, Wang X, Jiao C, Legall N, Mao LJN (2017) Genome resequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun* 8:249

7. Guo S, Zhao S, Sun H, Wang X, Wu S, Lin T, Ren Y, Gao L, Deng Y, Zhang JJN (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat Genet* 51:1616–1623
8. Huang R, Wang J-Y, Yao M-Z, Ma C-L, Chen LJHR (2022) Quantitative trait loci mapping for free amino acid content using an albino population and SNP markers provides insight into the genetic improvement of tea plants. *Hortic Res* 9:uhab029
9. Jing Z, Cheng K, Shu H, Ma Y, Liu PJBS (2023) Whole genome resequencing approach for conservation biology of endangered plants. *Biodiv Sci* 31:22679
10. Lei Y, Yang L, Duan S, Ning S, Li D, Wang Z, Xiang G, Wang C, Zhang S, Zhang SJFPS (2022) Whole-genome resequencing reveals the origin of tea in Lincang. *Front Plant Sci* 13:984422
11. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) bioinformatics GPDPSJ The sequence alignment/map format and SAMtools. *Bioinformatics*. 25: 2078–2079
13. Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C, Nick P, Du F, Fan PJN (2019) Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat Commun* 10(1):1190
14. Luo J, Shi Z, Shen C, Liu C, Gong Z, Huang YJJTS (2002) Studies on genetic relationships of tea cultivars [*Camellia sinensis* (L.) O. Kuntze] by RAPD analysis. *J Tea Sci* 22:140–146
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly MJG (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
16. Mishra RK, Chaudhury S, Ahmad A, Pradhan M, Siddiqi TOJIJIB (2009) Molecular analysis of tea clones (*Camellia sinensis*) using AFLP markers. *Int J Integr Biology* 5:130–136
17. Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* 76:5269–5273
18. Niu S, Song Q, Koiwa H, Qiao D, Zhao D, Chen Z, Liu X, Wen XJBPB (2019) Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biol* 19:1–12
19. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37(5):1530–1534
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575

21. Qi J, Liu X, Shen D, Miao H, Xie B, Li X, Zeng P, Wang S, Shang Y, Gu XJN (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat Genet* 45:1510–1515
22. Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA, Mumtaz ASJSa (2021) Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Sci Adv* 7(29):eabg2286
23. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40(W1):W452–W457
24. Tan Y, Li J, Liu S, Yan C, Chen JJJTS (2009) Genetic diversity of 43 tea cultivars (*Camellia sinensis* (L.) O. Kuntze) by SSR markers. *Front Plant Sci* 29:271–274
25. Wang X, Feng H, Chang Y, Ma C, Wang L, Hao X, Li A, Cheng H, Wang L, Cui PJN (2020) Population sequencing enhances understanding of tea plant evolution. *Nat Commun* 11(1):4447
26. Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, Zhang Y, Liu Y, Zhu T, Li WJM (2017) The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol Plant* 10:866–877
27. Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q, Liu Y, Yu J, Li F, Li RJM (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol Plant* 13:1013–1026
28. Zhang Q-J, Li W, Li K, Nan H, Shi C, Zhang Y, Dai Z-Y, Lin Y-L, Yang X-L, Tong YJMP (2020a) The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR retrotransposons in driving genome size evolution. *Mol Plant* 13(7):935–938
29. Zhang W, Zhang Y, Qiu H, Guo Y, Wan H, Zhang X, Scossa F, Alseekh S, Zhang Q, Wang P, Xu L, Schmidt MHW, Jia X, Li D, Zhu A, Guo F, Chen W, Ni D, Usadel B, Fernie AR, Wen W (2020b) Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. *Nat Commun* 11(1):3719

Table 1

Table 1 is available in the Supplementary Files section.

Figures

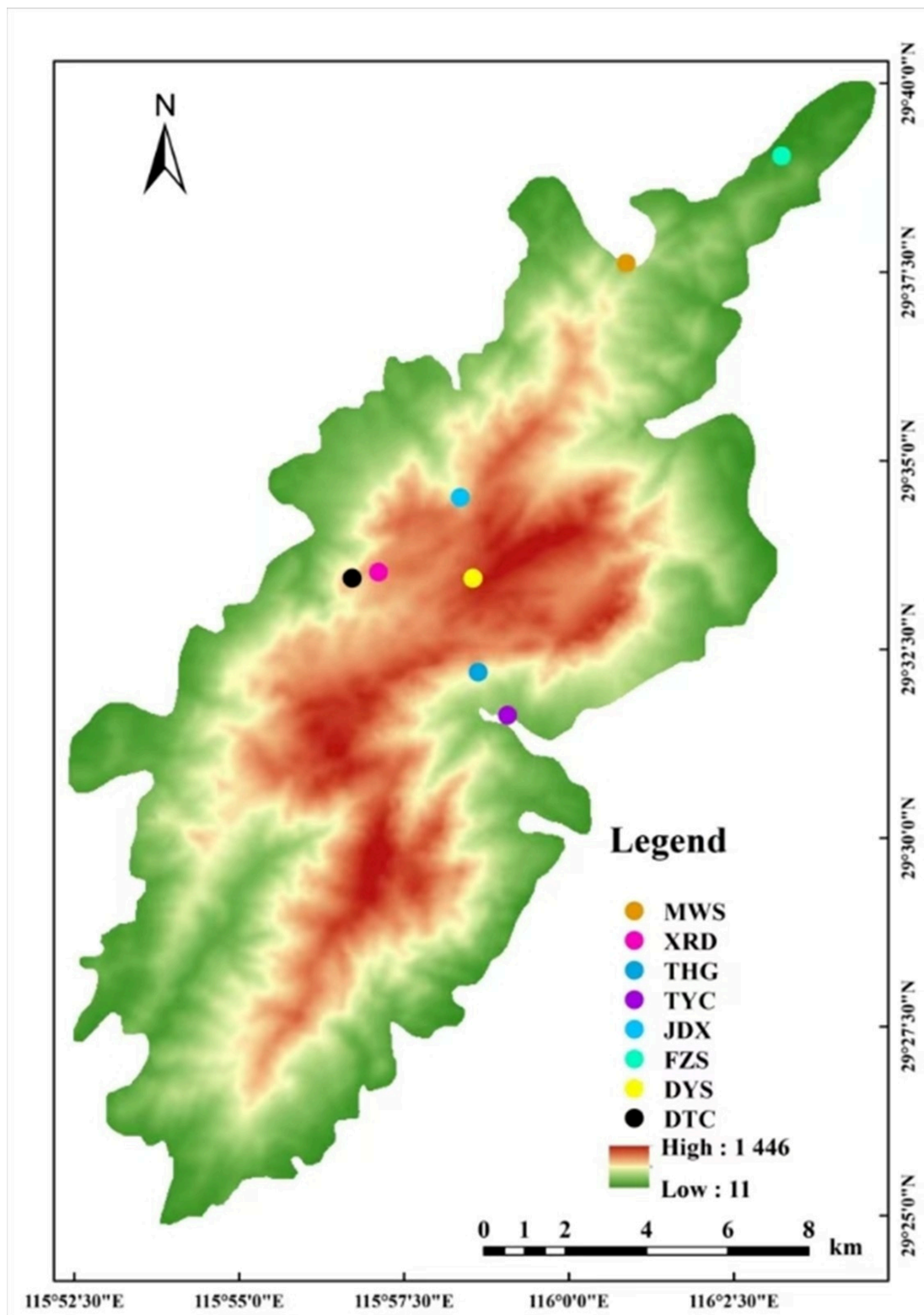


Figure 1

The geographical distribution of eight Lushan tea populations

Figure 3

Phylogenomics tree among eight Lushan tea populations based on genome-wide SNPs. (The circle on the branch represent support rate, the bigger the higher.)

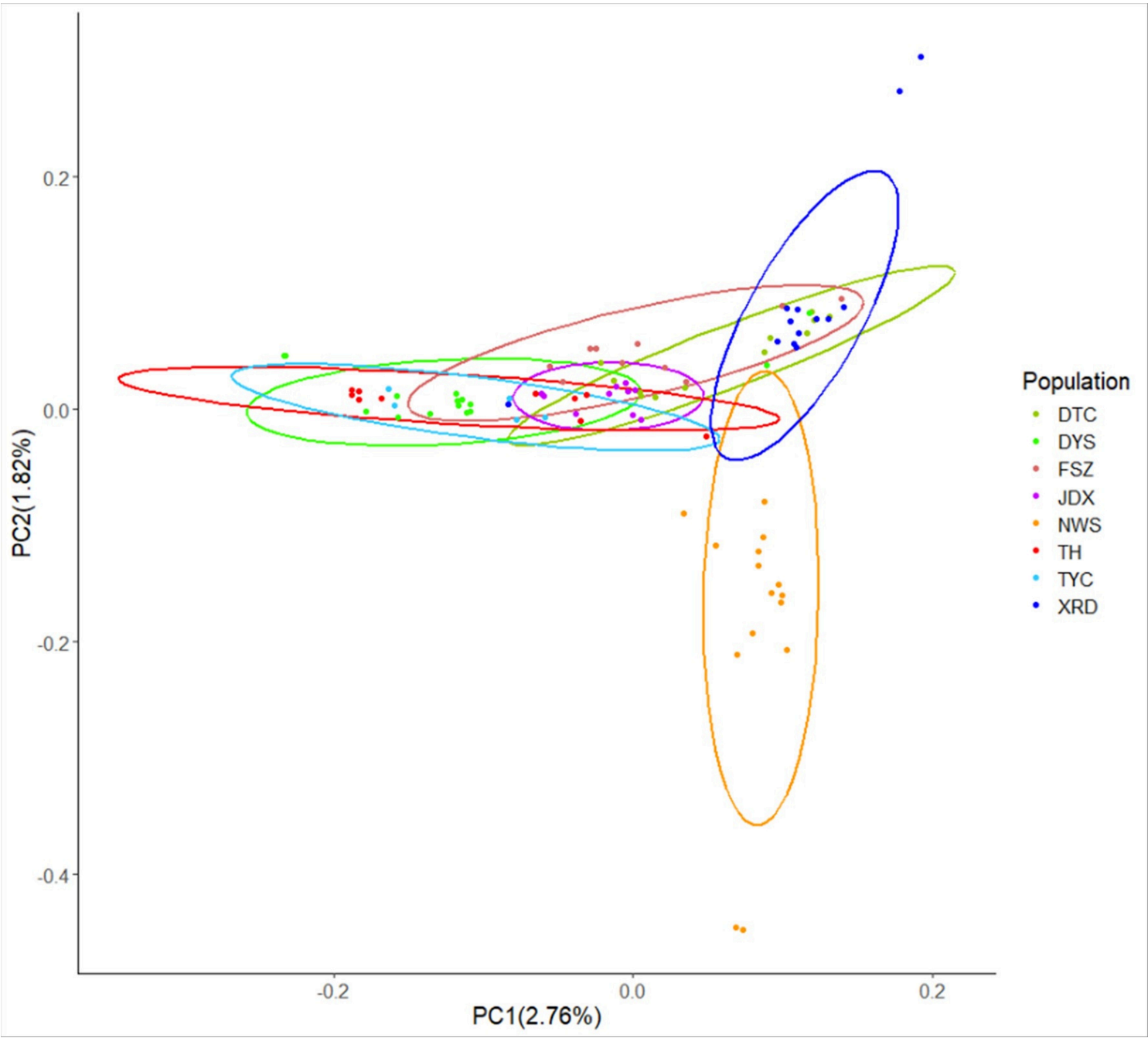


Figure 4

PCA analysis of eight *C. sinensis* populations for whole genome resequencing

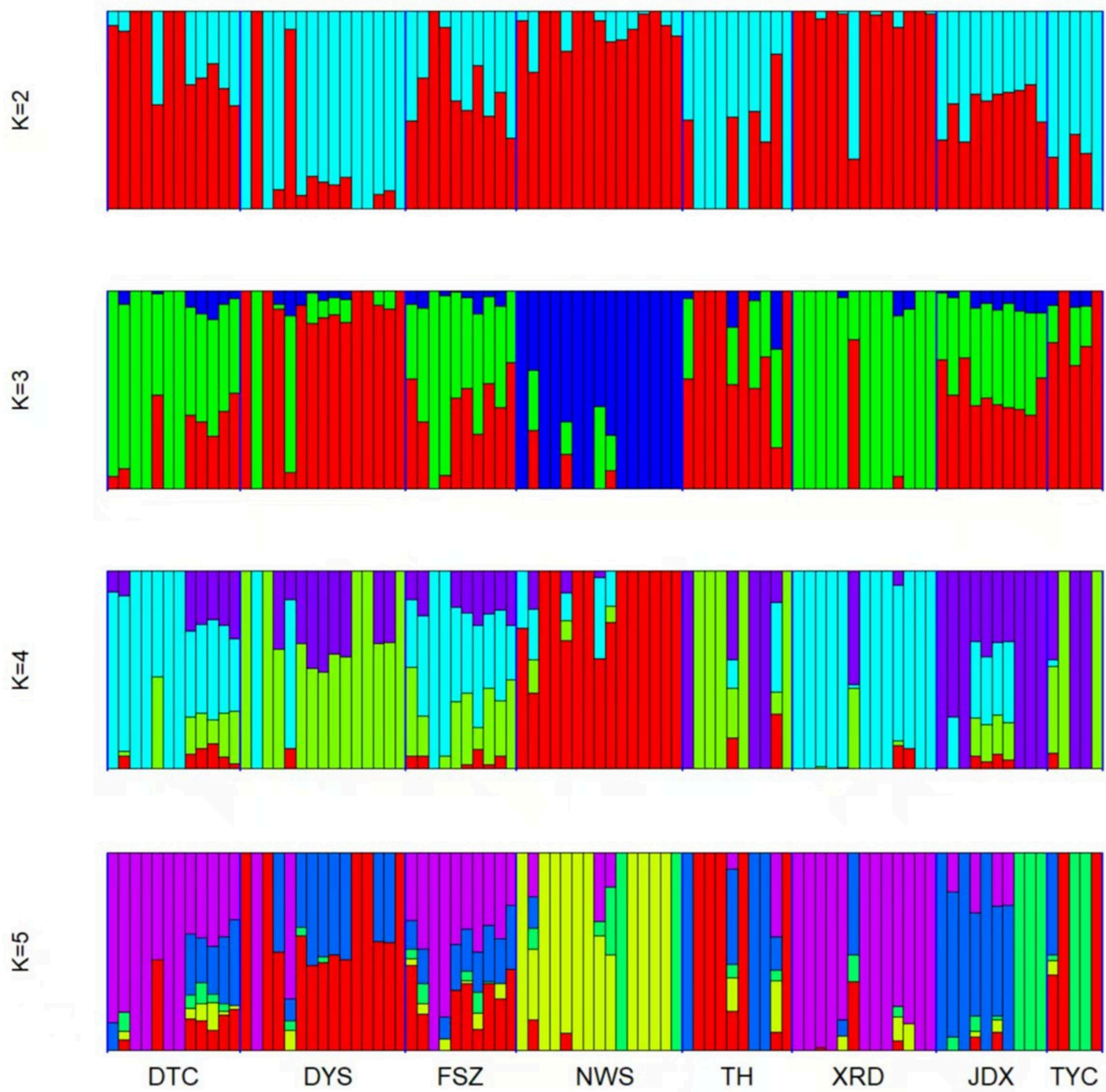


Figure 5

Population Genetic Structure Diagram

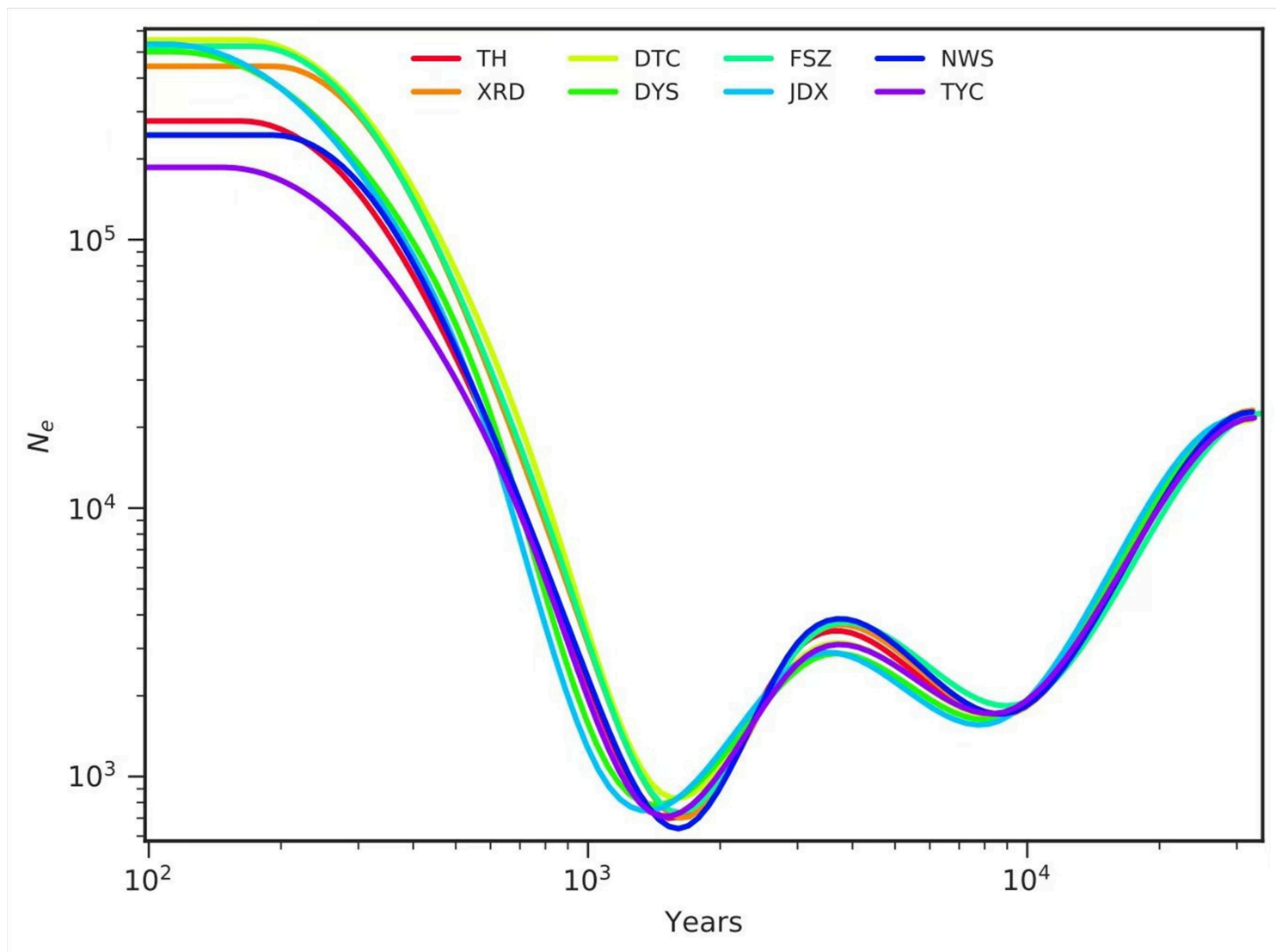


Figure 6

Demographic history of eight *C. sinensis* populations in Lushan

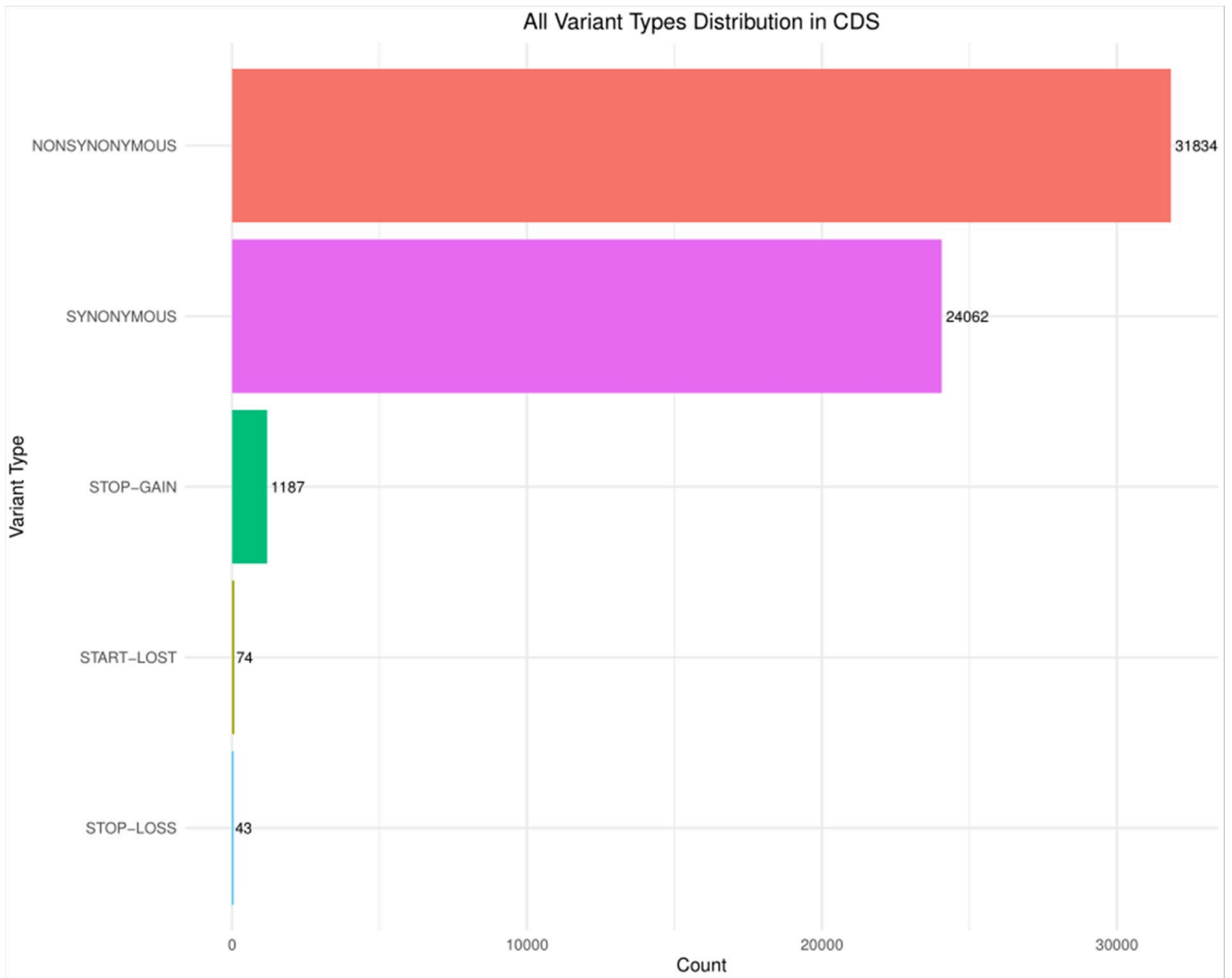


Figure 7

The distribution of all variant types Distribution in CDS

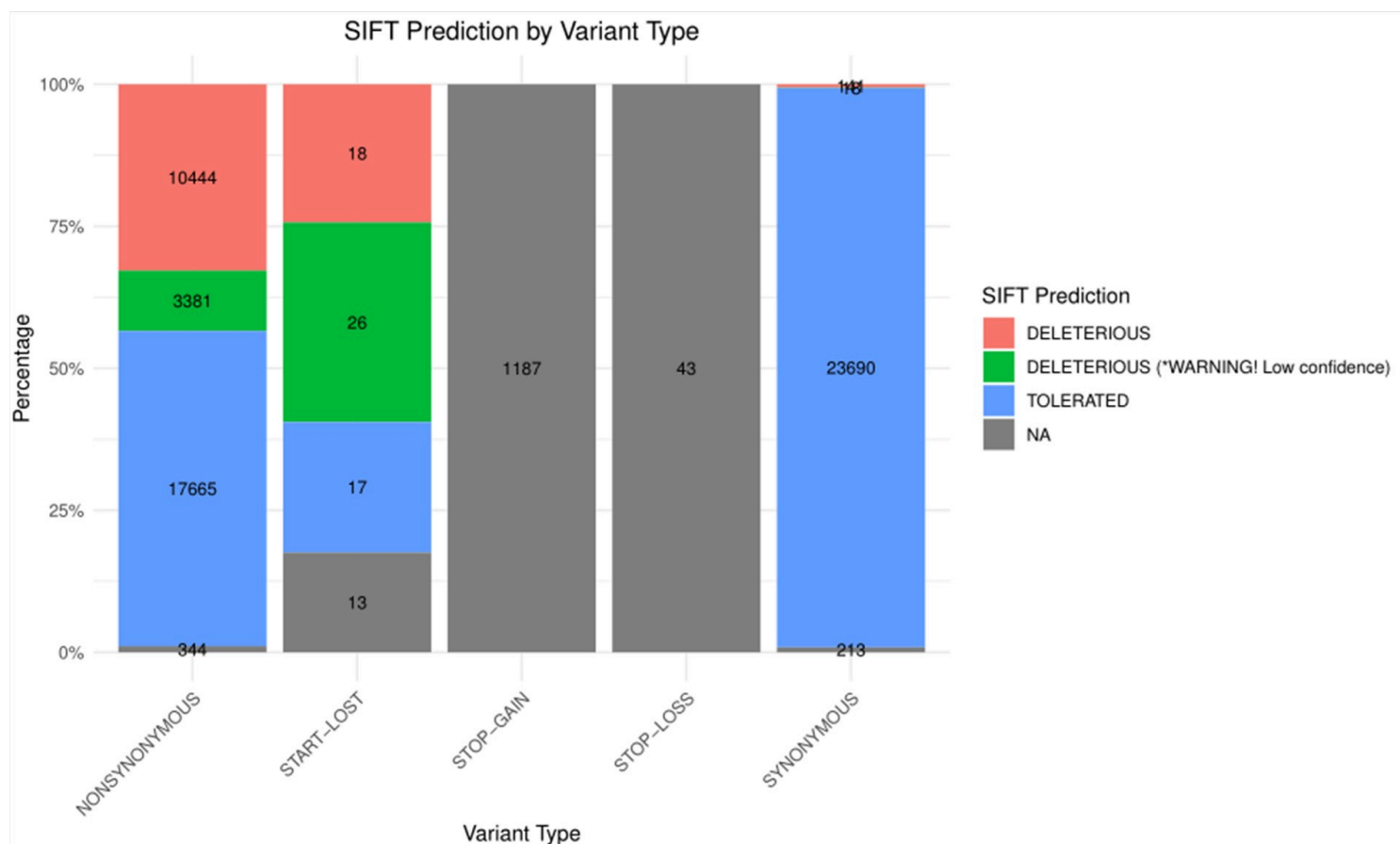


Figure 8

The distribution of SIFT prediction by variant type

Supplementary Files

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

- [Table1.docx](#)
- [FigureS1.png](#)
- [TableS1.csv](#)