

Quick Identification of Genetic Provenances and Best Practice Woodland Management of Chinese Pine (*Pinus tabuliformis*)

Hanhan Xia (✉ xiahanhan@zhku.edu.cn)

Zhongkai University of Agriculture and Engineering

Minyi Yang

Zhongkai University of Agriculture and Engineering

Feicui Yang

Zhongkai University of Agriculture and Engineering

Haoran Xie

Zhongkai University of Agriculture and Engineering

Juan Li

Zhongkai University of Agriculture and Engineering

Research

Keywords: *Pinus tabuliformis*, population structure, provenance identification, mitochondrial sequence system, breeding program

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-639788/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Chinese pine (*Pinus tabuliformis*) is one of the most widespread, ecologically and economically important tree species in North China. In this study, we analyzed and compared the genetic diversity and population structure of 158 individuals from 17 populations of *P. tabuliformis*, by group III, a new mitochondrial marker system (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) with two other groups of marker systems. In contrast to the conservation in the evolution of the mitochondrial sequence, the new mitochondrial marker system of *P. tabuliformis* shows as extremely high polymorphism as 25, whose haplotypes are more than four times of the group I marker system (*nad1-2*, *nad4-3*, and *nad5-1*) as 8 haplotypes. Although the group II, *nad7-1* (19 haplotypes), showed high resolution in the provenance identification of *P. tabuliformis*, the new mitochondrial marker system is more accurate for detection of specific groups like HL, WT and NS and powerful to differentiate populations between GD and SS. The results suggested that the new mitochondrial marker system is as high as the resolution of GBS (genotype by sequencing). It is much more available and will be of great help to provenance identification and molecular assisted breeding of *P. tabuliformis*. This study will make theoretical foundation for the following studies on the evaluation, cultivation and germplasm management of *P. tabuliformis* populations and aid the breeding, biodiversity and conservation programs of forest species.

Introduction

Chinese pine (*Pinus tabuliformis* Carr.), as one of the dominant coniferous and most important reforestation tree species in northern China, is ecological and economic important (Wu,1995;Ying,et al. 2004; Mao, Wang,2011)So far, genetic improvement researches mainly focused on provenance test, seedling breeding, reforestation program and tree physiology, have carried on to keep stable development of the woodland and germplasm management (Chen et al.2019;He et al.2020;Zhang et al.2020). Large breeding and tree improvement programs were initiated recently with the primary aim to increase forest coverage in harsh climate zones (Zhang, et al.1998; Zhang, et al.1999). Long-term provenance trials, according to the phenology, phenotypic variations, growth, germination and other indicators, found that *P. tabuliformis* across the whole geographic distribution could be divided into seven climatic types (Xu,1992;Chen et al.2019;He et al.2020). Previous studies show evidence of strong local adaptation as transplantation across climate zones results in high mortality and inferior growth of the transplanted populations (Xu,1992;Gray et al.2016;Zhao et al. 2014). So the mining and use of natural local population resources or populations with close genetic relationship will increase survival rate and be more beneficial (Chen et al.2019;He et al.2020).

However, long-term field experiments and observations are time-consuming and labor intensive(Xu,1992;Rony et al. 2009;Pan et al. 2020). A complementary method is genomic analysis, including organelle genomic data, to reveal differences at the DNA level between different geographical populations without unstable environmental interference, and are effective for evaluating the genetic diversity of germplasm, population dynamics in conservation and breeding programs (Meng et al.2007; Neale et al.2014; Luo et al.2020;Zurn et al. 2020). Considering the cost and difficulty of genomic analysis Holliday et al.2017, Xia et al. 2018) especially for non-model forest plants, organelle sequences analysis is a convenient choose for identification of particular population and the determination of their purity (Du,et al.2009; Zhou et al.2010) There have been many reports and studies on the diversity of *P. tabuliformis* germplasm resources by molecular marker(Chen et al.2008; Wang et al.2011;Hao et al. 2018).However, most of the studies focused on the pedigree genetic structure and demographic history of *P. tabuliformis*. The reported mitochondrial markers *nad1-2*, *nad4-3* and *nad5-1*, chloroplast gene fragments *rpl16* and *trnS-trnG* reflect the low level of genetic variation of *P. tabuliformis* (Chen et al.2008; Wang et al.2013;Hao et al.2018)especially compared with the genetic variation provided by the nuclear genome sequence(Xia et al. 2018) . The resolution of the classical mitochondrial and chloroplast marker is low and difficult to locate the origin of ancient *P. tabuliformis* by the limited information. Although the resolution of nuclear genome data by GBS (Genotyping by Sequencing) is good, the cost of sequencing and difficulty in data analysis is still a problem for technology extension and universally utilization. Although, high level of polymorphism was observed in *nad7-1* of *P. tabuliformis* (Xia et al. 2020), there are some population like GD (Guandi Mountain) and SS (Songshan Mountain) could not be differentiated. In this study, we combined *nad1-2*, *nad4-3*, *nad5-1* and *nad7-1* together to form a new marker system. We compared and analyzed the genetic diversity of different geographic natural populations of *P. tabuliformis* by 3 groups (*nad1-2*, *nad4-3* and *nad5-1*; *nad7-1*; *nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) of mt marker systems. It is expected that the new marker system is easiest, with highest resolution to identify the provenance of *P. tabuliformis*, which is easy operated and nice for technology extension. Identifying the provenance of nature population, is one of the powerful tools for genetic improvement of forest species. The result will provide direct evidence for provenance identification and the origin of *P. tabuliformis*. This study will support the theoretical basis for the tree conservation and utilization of germplasm resources, which is of great significance for protecting high-quality germplasm of forests species.

Materials And Methods

Data collection

The data in this study was collected from Wang et al. in 2011 (Wang et al. 2011) and Xia et al. in 2020 (Xia et al. 2020). The data in Wang et al. in 2011(Wang et al. 2011) included the sequences of *nad1-2*, *nad4-3* and *nad5-1* of 184 individuals from 17 populations of *P. tabuliformis*. And the data in Xia et al. in 2020 (Xia et al. 2020) included the sequences of *nad7-1* of 158 individuals from 17 populations. After removing the individuals with missing data, 158 individuals from 17 populations, all of which containing the the four mitochondrial segments (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*), were extracted for the following analysis.

Data analysis

The alignment of the sequences have been made in Clustal x 1.81(Thompson et al.1997). Then different sequences were obtained by manual proof. Both point mutation and insertion deletion were processed in the same proportion, and each insertion deletion was independent of other insertion deletions. A haplotype can be understood as a combination of all nucleotides on a DNA fragment different from other haplotypes. VcfTools v1.0.12 (Danecek et al. 2011).was used to calculate each population. Genetic diversity parameters have been made: individual number (N) and haplotype diversity (nh). At last, the

comparison of three groups of marker systems, group 1 (*nad1-2*, *nad4-3* and *nad5-1*), group 2 (*nad7-1*) and group 3 (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) have been made.

Results

Mitochondrial genetic variation patterns of nad1-2, nad4-3, nad5-1 and nad7-1

Haplotype of mitochondrial fragments in *P. tabuliformis* showed a clear regional specific distribution pattern. The mitochondrial molecular marker system (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) obtained 25 specific haplotypes in 158 individuals (Table 1). Three geographically western groups are dominated by M10-H6, while RG is dominated by M10-H12 independently. The largest geographic zone, containing 10 groups, are dominated by M4-H1 or M5-H1. There are some independent populations, such as NC dominated by M3-H5 and WT dominated by M1-H3, separate from the northern component (Figure 1, Table 1). The genetic content of the populations ZW1, ZW2, NS and GY in the southern zone are complex and different from each other, composed of M8-H19, M3-H5, M9-H16 and M8-H14, respectively.

The results were consistent with the population structure obtained by marker group I (*nad1-2*, *nad4-3*, *nad5-1*) (Table S1) and marker group II (*nad7-1*) (Table S2) that the mitochondrial genetic structure of *P. tabuliformis* can be divided into three large groups, indicating the stability of the new marker system. However, the marker group III, the new marker system can further distinguish some isolated and independent groups which could not be separated on marker group I and group II (Fig. 1). For example, so far GD (Guandi Mountain) and SS (Songshan Mountain) can only be distinguished at the level of nuclear genomic sequence (Xia et al.2018), now can be simply identified by mitochondrial markers. And compared with marker group I (*nad1-2*, *nad4-3*, *nad5-1*) (Table S1) and marker group II (*nad7-1*), more haplotypes were identified in HL (Helan Mountain), WT (Wutai Mountain) and NS (Ningshan Mountain) by the new marker system, which shows much higher resolution.

Discussion

Polymorphism distribution of mitochondria

Generally, the variation pattern of mitochondrial genome is similar to that of nuclear gene that it is divided into three major distribution regions as the south, the north and the west. From the perspective of population genetic diversity, the genetic variation of population in the central and western regions of *P. tabuliformis* are higher than the eastern regions, and the southwest population has the most abundant genetic resources. Among the southwest populations, the genetic variation of NS (Ningshan) is the highest. The results are in consistency with observation got by previous long-term provenance experiments and field studies (Chen et al.2019; He et al.2020; Xu et al.1992; Zhao et al.2014). It is indicated that NS might be one of the origins of ancient *P. tabuliformis*. The Southwest China, as part of the "Sino-Japanese Floristic Region", has a complex topography and plant diversity. There are superior natural conditions, complex habitats and Climate conditions are very suitable for the growth of local *P. tabuliformis*. In recent years, due to intentional protection, human beings have less damage to the original habitat, which has a positive effect on the maintenance and accumulation of genetic variation.

A new mitochondrial marker system for efficient provenances identification of P. tabuliformis

So far, the mitochondrial fragments widely used in *P. tabuliformis* is *nad1-2*, *nad4-3* and *nad5-1*. They have relatively low levels of variation among populations, as only two to four haplotypes when each fragment used alone and eight haplotypes when the three fragments combined (Wang et al. 2011). We found a high level of variation in the *nad7-1* region of *P. tabuliformis* in a recent study, which can identify the maternal sources of different provenances, twice the sum of *nad1-2*, *nad4-3* and *nad5-1*(Xia et al.2020). In this study, we combined the four markers (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) to form a new system and found the genetic structure of *P. tabuliformis* with obvious geographical pattern. By comparing the three groups of mitochondrial marker system, it is found that the number of haplotypes in identified by group III system (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) is 25, significantly more than that of group I (8 haplotypes) and *nad7-1*.

Although *nad7-1* could achieved a relatively high resolution (19 haplotypes), the new marker system (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) is more accurate in the detection of specific populations like HL, WT and NS, and can even identified the populations like GD and SS, which could not be distinguished by *nad7-1*. We have developed GBS (Genotyping by sequencing) to study the population structure and genetic variation of *P. tabuliformis* (Xia et al., 2018). The resolution of the new marker system (*nad1-2*, *nad4-3*, *nad5-1* and *nad7-1*) with rich length variation is comparable to that of the GBS. However, the new mitochondrial system does not need do the sequencing in the company, while as simple as the PCR run gel in the lab and easy to complete in every lab. The results of this study will provide the most efficient method for identifying the maternal origin of *P. tabuliformis*, and provide technical support for molecular assisted breeding of improved varieties.

By years of investigation and collection of germplasm resources, we conducted management and protection of *P. tabuliformis* for a long time. In this study, the new mitochondrial system composed by the developed *nad7-1* and three classical markers (*nad1-2*, *nad4-3* and *nad5-1*) was reported for the first time, which could identify the genetic differences and population genetic structure characteristics among *P. tabuliformis* populations much more effective and easier. The molecular marker system obtained in this study is good to identify genetic variation, which provide primer support and technology extension for the following study of *P. tabuliformis* plantation. The study of origin will support the introduction and growth of local *P. tabuliformis* populations. The new marker system will provide a powerful tool and analysis platform for promoting the identification and utilization of *P. tabuliformis* germplasm resources, selecting breeding program. Provenance selection is very important for the survival rate and quality of plantation. The adaptability and growth performance of various geographical provenances is significant different in varied introduction areas. Therefore, this study provides a marker system and platform for the analysis of different provenances of *P. tabuliformis* plantations, and lays foundation for further directional breeding and planting promotion.

Prospect

Since the *P. tabuliformis* is distributed in the area with advanced cultural development, most natural forests have been damaged due to various economic and social disturbance. The remaining natural forests are precious for research and production, as a repository for various genetic resources in breeding program. We should protect the decreasing wild resources of *P. tabuliformis* and expand high-quality plantation. In production and afforestation, more attention should be paid to provenance selection. For example, in the process of seed testing and seeding treatment, different measures and standards should be taken according to provenances, to save seeds and improve the yield and quality of seedlings, so that greater economic benefits can be achieved.

Declarations

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgments

The authors thank National natural science foundation of China (No. 32001244) for the funding support.

References

1. Mao, J.F.; X.R.Wang 2011. Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan Plateau. *American Naturalist*, 177, 424–439.
2. Wu, C.Y. 1995. Vegetation of China, 2nd ed.; Beijing, China: Science Press. (in Chinese)
3. Ying, T.S.; M.L.Chen; H.C.Chang 2004. Atlas of the gymnosperms of China. Beijing, China: China Science and Technology Press.
4. Chen, X.Y.; J.X.Meng; X.Q.Zhou; H.W.Yuan; S.H.Niu; Y.Li 2019. Genetic variation analysis of needle morphological anatomy and physiological indexes of *Pinus tabuliformis* geographical populations. *Journal of Beijing Forestry University*, 41: 19-30. (in Chinese)
5. He, K.K.; W.B.Wu; Z.J. Zhang; X.G.Hu; F.X.Han; S.H.Niu; Y.Li 2020. Genetic structure variation of *Pinus tabuliformis* plantation in Beijing and population difference with Shanxi mountains. *Journal of Beijing Forestry University*, 42: 33-42. (in Chinese)
6. Zhang, X.W.; Q.J.Yu; W.Q. Zhang; L.Q.Zhao; L.Y.Zhang; G.S.Luo; Q.Jia; Z.Q.Jia 2020. Research on wood properties and growth process of *Pinus tabuliformis* building timber forest with different tree ages. *Journal of Central South University of Forestry and Technology*, 40: 122-131.
7. Zhang. C.X.; Y.Li.; X.M.Chen 1999. Study on isozyme site selection of *Pinus tabuliformis*. *Journal of Beijing Forestry University*, 22: 7-12. (in Chinese)
8. Zhang, C.X.; Y.Li; X.H.Shen; X.M.Chen 1998. Genetic diversity of two isozyme systems in 10 natural and artificial populations of *Pinus tabuliformis*. Proceedings of the third National Symposium on biodiversity conservation and sustainable utilization (in Chinese)
9. Xu, H.C. 1992. Geographic variation and provenance selection of *Pinus tabuliformis* Carr. Beijing. (in Chinese): China Forestry Press
10. Gray, L.K.; A.Hamann; S.John; D.Rweyongeza; L.Barnhardt; B.R.Thomas 2016. Climate change risk management in tree improvement programs: selection and movement of genotypes. *Tree Genetics & Genomes*, 12, 23.
11. Zhao, W.; J.X.Meng; B. Wang; L.S. Zhang; Y.L.Xu; Q.Y.Zeng; X.R.Wang, 2014. Weak crossability barrier but strong juvenile selection supports ecological speciation of the hybrid pine *Pinus densata* on the Tibetan Plateau. *Evolution*, 68, 3120-3133.
12. Pan, Y.Y.; L.P.Jiang; G.Y. Xu; J.Q.Li; B.Y.Wang; Y.X.Li; X.Y.Zhao 2020. Evaluation and selection analyses of 60 *Larix kaempferi* clones in four provenances based on growth traits and wood properties. *Tree Genetics & Genomes*, 16, 27.
13. Rony, C.; R.Baalbaki; P.Kalaitzis; S.N.Talhouk 2009. Molecular characterization of Lebanese olive germplasm. *Tree Genetics & Genomes*, 5, 109–115.
14. Luo, F.X.; J.L.Norelli; N.P. Howard; M.Wisniewski; H.Flachowsky; M.V.Hanke; C.Peace 2020. Introgressing blue mold resistance into elite apple germplasm by rapid cycle breeding and foreground and background DNA-informed selection. *Tree Genetics & Genomes*, 16, 28.
15. Meng, L.H.; R.Yang; R.J.Abbott; G.Miehe; T.H. Hu; J.Q. Liu 2007. Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Molecular Ecology*, 16, 4128– 4137.
16. Neale, D.B.; J.L.Wegrzyn; K.A. Stevens; A.V.Zimin; D. Puiu; M.W. Crepeau; C.Cardeno; M.Koriabine; A.E.Holtz-Morris; J.D.Liechty.; et al. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, 15, R59.
17. Zurn, J.D.; A.Nyberg; S.Montanari J.Postman; D.Neale; N.Bassil 2020. A new SSR fingerprinting set and its comparison to existing SSR- and SNP-based genotyping platforms to manage *Pyrus* germplasm resources. *Tree Genetics & Genomes*, 16, 72.
18. Holliday, J.A.; S.N.Aitken; J.E.K.Cooke; B.Fady; S.C.González-Martínez; M.Heuertz; J-P.Jaramillo-Correa; C. Lexer; M. Staton; W.R.Whetten; et al. 2017. Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Molecular Ecology*, 26: 706-717.
19. Xia, H.H.; B.S.Wang; W. Zhao; J.Pan; J.F.Mao; X.R. Wang 2018. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environment, and geography on population structure in *Pinus tabuliformis*. *Evolutionary Applications*, 11: 1931-1945.
20. Du, F.K.; R.J.Petit; J.Q. Liu 2009. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Molecular Ecology*, 18: 1396-1407.
21. Zhou, Y.F.; R.J.Abbott; Z.Y.Jiang; F.K.Du; R.I.Milne; J.Q.Liu 2010. Gene Flow And Species Delimitation: A Case Study Of Two Pine Species With Overlapping Distributions In Southeast China. *Evolution*, 64: 2342-2352.

22. Chen, K.M.; R.J.Abbott; R.I.Milne; X.M.Tian; J.Q.Liu 2008.Phylogeography of *Pinus tabuliformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. *Molecular Ecology*, 17: 4276-4288.
23. Hao Q.; De, L.G.; D.S.Guo; H.Y.Gu; F.S.Hu; Y.Han; Z.L.Song; H.Y.Liu 2018.The critical role of local refugia in postglacial colonization of Chinese pine: joint inferences from DNA analyses, pollen records, and species distribution modeling. *Ecography*, 41: 592-606.
24. Wang, B.S.; J.F.Mao; J.Gao; W.Zhao; X.R.Wang 2011.Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Molecular Ecology*, 20: 3796-3811.
25. Wang, B.S.; J.F. Mao; W.Zhao; X.R. Wang 2013.Impact of Geography and Climate on the Genetic Differentiation of the Subtropical Pine *Pinus yunnanensis*. *Plos One*, 8(6).
26. Xia, H.; Zhao, W.; Y.Shi; X.R.Wang; B. Wang 2020 Nov 3; Microhomologies Are Associated with Tandem Duplications and Structural Variation in Plant Mitochondrial Genomes. *Genome Biology and Evolution*, 12:1965-1974.
27. Thompson, J.D.; T.J.Gibson; F.Plewniak; F.Jeanmougin; D.G. Higgins 1997.The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research*, 25: 4876-4882.
28. Danecek, P.; A.Auton; G.Abecasis; C.A.Albers; E.Banks; M.A. DePristo; R.E. Handsaker; G.Lunter; G.T. Marth; S.T.Sherry; et al. 2011.The variant call format and VCFtools. *Bioinformatics*, 27: 2156-2158.

Tables

Table 1. Geographical locations, sample size (N), number of haplotypes (nh) of the *P. tabuliformis* populations identified by mark system (*nad1-2+nad4-3+nad5-1+nad7-1*)

Group III marker system (<i>nad1-2+nad4-3+nad5-1+nad7-1</i>)																					
Pop	N	nh	M1-H3	M2-H5	M3-H4	M3-H5	M4-H1	M4-H3	M4-H10	M5-H1	M5-H2	M5-H8	M5-H10	M5-H13	M8-H9	M8-H11	M8-H12	M8-H14	M8-H15	M8-H19	M H'
DS	8	2	0	0	0	0	6	0	0	0	0	2	0	0	0	0	0	0	0	0	0
GD	8	1	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GY	13	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	2	0	0
HL	11	2	0	0	0	0	0	0	1	0	0	0	10	0	0	0	0	0	0	0	0
HZ	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JL	7	3	0	0	0	0	5	0	0	1	1	0	0	0	0	0	0	0	0	0	0
KY	15	2	0	0	0	0	11	0	0	0	0	0	0	0	4	0	0	0	0	0	0
LK	6	2	0	0	0	0	2	0	0	0	0	0	0	4	0	0	0	0	0	0	0
NC	7	2	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NS	14	6	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	1
QL	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SS	8	2	0	0	0	0	7	0	0	1	0	0	0	0	0	0	0	0	0	0	0
RG	11	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
TB	8	1	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WT	16	2	13	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ZW1	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
ZW2	8	3	0	1	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	158	27	13	2	1	12	51	1	1	3	1	2	10	4	4	1	1	11	2	8	1

Figures

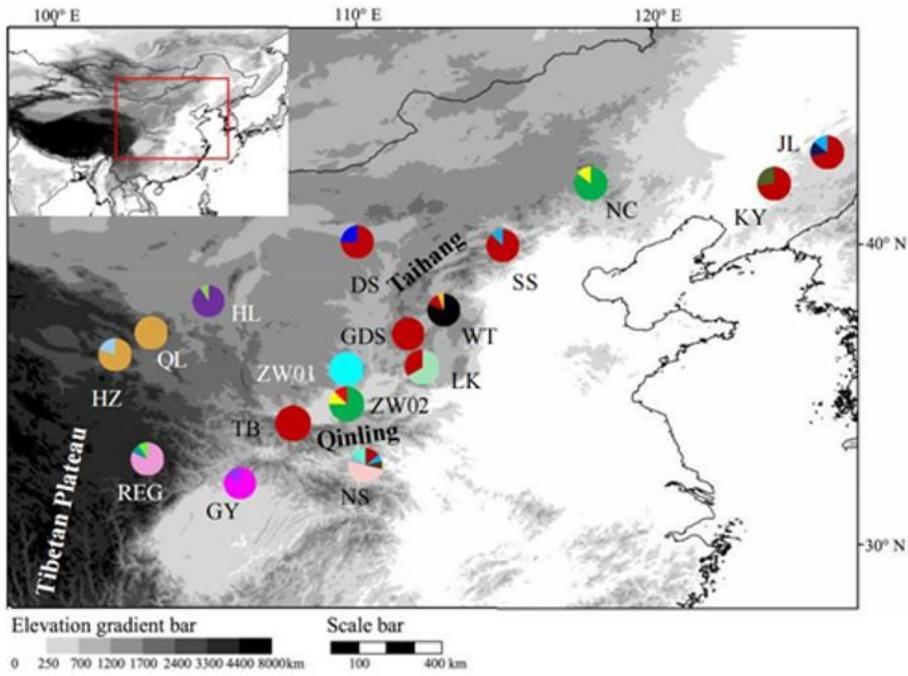


Figure 1

Pie charts show the proportions of mitotypes by nad1-2, nad4-3, nad5-1 and nad7-1 system Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.