

Comparative transcriptomes and genetic diversity assessment of two Sichuan peppers (*Zanthoxylum bungeanum* and *Zanthoxylum armatum*)

Yangchuan Deng (✉ 409207145@qq.com)

Sichuan Agricultural University

Li Xiang

Wei Gong

Tingyu Ma

Li Yang

Xiangxiao Meng

Shilin Chen

Meng Ye

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences

Research article

Keywords: *Zanthoxylum*, transcriptome, unigene, EST-SSR, genetic diversity

Posted Date: June 19th, 2019

DOI: <https://doi.org/10.21203/rs.2.10186/v1>

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

Abstract

Background: Species of *Zanthoxylum* (Sichuan pepper) are diverse and valuable economic trees that have been cultivated across Asia for thousands of years for their aromatic and medicinal properties. The current lack of transcriptome and EST-SSR data for these species represents a constraint with regards to identifying increasingly more diverse cultivars in China. This study was conducted to compare the transcriptome profiles of two *Zanthoxylum* species (*Z. bungeanum* and *Z. armatum*) and develop the EST-SSR markers to assess the genetic diversity in Sichuan pepper. **Results:** A total of 36.76 G high-quality clean data were screened for subsequent analysis. For *Z. bungeanum* and *Z. armatum*, 64,944 and 75,669 unigenes were obtained, respectively. After comparing different databases, we found that the highest number of unigenes (43,3198 and 49,638 for *Z. bungeanum* and *Z. armatum*, respectively) were annotated using the NR database. In addition, six pairs of EST-SSR markers were selected to determine the genetic diversity of 125 samples of three *Zanthoxylum* species. A population clustering dendrogram indicated that there was an average of four samples per population, with the number of allelic genes (N_a) ranging from 1 to 1.4510 (average of 1.1462). The number of effective allelic genes (N_e) ranged from 1 to 1.2744 (average of 1.0971), whereas genetic diversity (N_{ei}) ranged from 0 to 0.1486 (average of 0.05512) and Shannon genetic diversity (H) ranged from 0 to 0.2177 (average 0.08123). In individual clustering dendrograms, the 125 samples were divided into five cohorts. Cohort I contained *Z. bungeanum* and its varieties, cohort II included *Z. armatum* and its varieties, cohort III contained mostly *Z. piperitum* and two other *Z. bungeanum* samples, and cohorts IV and V each contained only two samples, with genetic distances of 0 to 1.14 between them. **Conclusion:** In this study, we analyzed the comparative transcriptome data of two *Zanthoxylum* species, for which we assessed genetic diversity using EST-SSR primers derived from their EST information. The genetic diversity of 125 samples obtained from the three *Zanthoxylum* species *Z. bungeanum*, *Z. armatum* and *Z. piperitum* was relatively low, which can probably be attributed to apomixis.

Background

Species within the genus *Zanthoxylum* (family Rutaceae), commonly referred to as Sichuan pepper, include a wide variety of shrubs and trees. More than 250 species in *Zanthoxylum* have been identified to date, most of which are distributed in tropical and subtropical Asia, Africa, and Oceania. In China, there are 41 species of *Zanthoxylum*, of which 25 are endemic [1], including several well-known species that are cultivated for their aromatic and medicinal properties: *Zanthoxylum bungeanum* Maxim, *Zanthoxylum schinifolium* Sieb. et Zucc, *Zanthoxylum armatum* DC and *Zanthoxylum nitidum* (Roxb.) DC. Sichuan pepper is used as a unique condiment in Asian countries, particularly in China [2, 3]. Currently, *Z. bungeanum* and *Z. armatum* are the main commercial sources of Sichuan peppers in China [4]. However, there are regional variations in the diversity of these two species, and their Chinese names also tend to differ from region to region. For example, “Qinghuajiao” is used as a common name for both *Z. armatum* and *Z. schinifolium* Siebold & Zuccarini. Similarly, “Tengjiao,” “Jinyang Qinghuajiao,” and “Jiuye Qinghuajiao” are generally treated as different species although they are all varieties of *Z. armatum* [5]. As a consequence, such discrepancies can lead to confusion in both practical cultivation and scientific research.

On the basis of conserved sequences, EST-SSR markers have been used to design primers using a combination of SSR and transcriptomes techniques. Using this approach, it is possible to screen samples using more accurate and specific primers [6] and has previously been used in wheat [6], grape [7], rice [8], soybean [9], and orange [10]. To date, however, this methodology has yet to be applied to analyses of *Zanthoxylum*. In this study, we initially examined the two transcriptomes of *Z. bungeanum* and *Z. armatum*, and through screening, developed EST-SSR primers, which were used to assess the genetic diversity of different varieties of Sichuan pepper.

Functional Annotation

2.2.1 GO annotation

GO (Gene Ontology) annotation includes three aspects: cellular component, molecular function, and biological process. For *Z. bungeanum*, we annotated 89,198 isogenes (Additional file 4: Fig. S1). Among these, 57,683, 35,489, and 27,967 isogenes were respectively annotated as biological process, cellular component, and molecular function, respectively. Specifically, 15,590 isogenes were assigned to “metabolic process” which was the most frequently assigned category, and accounted for 27.03% of biological process and 17.48% among all the annotated isogenes. The category “cell” accounted for the highest number of isogenes (7,949) annotated as cellular component, accounting for 22.4% of cellular components and 8.91% of all annotated isogenes. Within the category molecular function, “binding” accounted for most (12,873) of the annotated isogenes and accounted for 46.03% of isogenes assigned to molecular function and 14.43% of all annotated isogenes.

For *Z. armatum*, a total of 137,567 isogenes were annotated with respect to GO terms (Additional file 4: Fig. S2). In detail, 66,368, 31,517, and 17,540 isogenes were annotated as biological process, cellular component, and molecular function, respectively. With regards to biological process, the highest number of isogenes were assigned “metabolic process,” accounting for 26.43% and 12.75% of isogenes annotated as biological process and all annotated isogenes, respectively. For cellular component, the highest number of annotated isogenes (8,840) were assigned to “cell,” accounting for 22.28% and 6.43% of isogenes annotated at cellular component and total annotated isogenes, respectively. In terms of molecular process, 14,252 isogenes were assigned to “binding,” accounting for 45.22% of isogenes annotated as molecular process, and 10.36% of all annotated isogenes.

2.2.2 KEGG pathways

KEGG pathways consist of cellular processes (cohort A), environmental information processing (cohort B), genetic information processing (cohort C), metabolism (cohort D), and organismal systems (cohort E). As shown in Fig. 2a. 25,520 isogenes were annotated to one of these five pathways. Of those isogenes annotated, the highest number (12,085) were assigned to cohort A and accounted for 47.36% all isogenes. A further 4,879 isogenes were mapped to cohort B (19.12%), whereas 3,908 (15.31%), 2,242 (8.79%), and 2,406 (9.43%) isogenes were assigned to cohorts E, C, and D, respectively. Among those isogenes annotated to cohort A, the largest number (4,374, 36.19%) were annotated as “global and overview maps,” whereas for cohort B, the largest number of isogenes (2,345, 48.06%) were annotated as “translation.” In cohort E, “endocrine system” accounted for the highest number of annotated isogenes (826, 21.14%), whereas for cohort C, the highest number of isogenes (1,518, 67.71%) were annotated as “signal transduction.” In cohort D, the highest number of isogenes (942, 39.15%) were annotated as “transport and catabolism.”

Figure 2b shows the results of KEGG pathway analysis for *Z. armatum*. In total, 27,917 isogenes were annotated to the aforementioned five pathways. Of these, the highest number (13,964, 50.02%) were annotated to cohort A, whereas 4,925 isogenes (17.64%) were annotated to cohort B. Cohort E ranked 3rd, with 3,953 (14.16%) of the annotated isogenes, whereas, 611 (9.35%) and 2,464 (8.33%) isogenes were annotated to cohorts D and C, respectively. In detail, 5,137 isogenes were classified in “global and overview maps” which accounted for 36.79% of the isogenes annotated to cohort A. For cohort B, the highest number of isogenes (2,381, 48.35%) were assigned to “translation,” whereas in cohort E, “endocrine system” accounted for the highest number (855, 21.63%) of annotated isogenes. For cohort C, “signal transduction” accounted for the highest number (1,635, 66.36%) of annotated isogenes, whereas the highest number of isogenes in cohort D (1,078, 41.29%) were annotated as “cell growth and death” These findings indicate that *Z. bungeanum* and *Z. armatum* differ only in respect to the highest number of isogenes annotated to pathways in cohort D.

2.2.3 Unigene annotation in different databases

A comparison of 64,944 unigenes using different databases (NR, COG, KEGG, and GO) indicated that, for *Z. bungeanum* (Fig. 2c), the highest number of unigenes (43,318, 66.70% of the total) were annotated using the NR database, whereas the COG databases annotated the fewest unigenes (4,153, 45.83%). Among the total 64,944 unigenes, 2,143 were assigned by all four databases, and 15,700 unigenes were annotated by three databases, amongst which the NR, KEGG, and GO databases accounted for the highest number of unigene annotations (14,201). Furthermore, 14,901 unigenes were annotated by two databases. In detail, 11,401 unigenes were annotated by databases NR and GO, which accounted for the highest percentage. Some unigenes were only annotated by a single database, with the NR and COG databases annotating the highest (43,318) and lowest (4,153) numbers of these unigenes, respectively. For *Z. armatum* (Fig. 2d), 75,669 unigenes were assembled. According to the blast results obtained using the aforementioned four databases, the highest number of unigenes (49,368, 65.60%) were annotated using the NR database, whereas the fewest unigenes (6,130, 8.10%) were annotated using the COG database. Furthermore, 2,658 and 17,193 unigenes were annotated by all four and three databases, respectively. Among these, the largest number of unigenes (15,035) were annotated collectively by NR, KEGG, and GO. We also found that 17,905 unigenes were annotated simultaneously by two databases, NR and GO accounting for the highest number (13,249) of annotated unigenes. Among the unigenes that were only annotated by a single database, the NR and COG databases annotated the highest (12,734) and lowest (6,130) numbers, respectively.

Figure 2. a: KEGG pathway of isogenes for *Zanthoxylum bungeanum*; b: KEGG pathway of isogenes for *Zanthoxylum armatum*; c: Unigenes of *Z. bungeanum* annotated by four databases; d: Unigenes of *Z. armatum* annotated by four databases

2.3 SSR characteristics of *Z. bungeanum* and *Z. armatum*

For *Z. bungeanum*, we assembled 64,944 unigenes with a total length of 54,073,890 bp. For *Z. bungeanum*, a total of 12,746 SSR loci were identified in 10,595 unigenes (Table 1). The percentage incidence of SSRs (number of unigenes with SSR loci/total number of

unigenes) was 16.31%, whereas the percentage occurrence (number of SSRs/total number of unigenes) was 19.63%. The average distance between SSR loci was 4.24 kb. In 1,805 unigenes, more than two SSR loci were detected. In terms of SSR repeat number, mononucleotides accounted for the highest number (6,296), whereas 3,092 and 2,992 SSRs comprised trinucleotides and dinucleotides, respectively. However, we detected tetranucleotide, hexanucleotide, and pentanucleotide SSRs in only 214, 105, and 47 unigenes, respectively.

For *Z. armatum* (Table 1), we assembled 75,669 unigenes, with a total length of 58,975,053. We detected 15,096 SSR loci within 12,612 unigenes. The percentage incidence and occurrence of these SSRs was 16.67% and 19.95%, respectively. The mean distance between two SSR loci was 3.91 kb, and we found that 2,101 unigenes harbored more than two SSR loci. Among all SSR repeats, the highest number of loci (7,830) were mononucleotides, followed in numerical abundance by trinucleotides (3,711), dinucleotides (3,169), tetranucleotides (223), hexanucleotides (112), and pentanucleotides (51). In summary, the number of mononucleotide SSR loci in *Z. armatum* (7,830) exceeded that of all other repeat types combined (7,266) and also the number of mononucleotide loci in *Z. bungeanum*. In addition, the numbers for each type of SSR repeat and SSR loci in *Z. armatum* exceeded those in *Z. bungeanum*.

Statistical analyses of numbers and distribution indicated that the percentage incidence of SSR loci in *Z. bungeanum* was 19.63% and the average distribution distance was 4.24 kb. Among these loci, mononucleotides had the highest incidence (9.69%), followed by trinucleotides (4.76%) and dinucleotides (4.61%). In terms of the average distance between loci, the distribution distance between mononucleotides was the shortest (8.59 kb), followed by trinucleotides (17.49 kb) and dinucleotides (18.07 kb). For *Z. armatum*, the percentage incidence of SSR was 19.95% and the mean distance between loci was 3.91 kb. Mononucleotides were found to have the highest percentage incidence (10.35%), followed by trinucleotides (4.90%) and dinucleotides (4.19%). In terms of the average distance between loci, we found that the shortest mean distance occurred between mononucleotides (7.53 kb), whereas those between trinucleotide and dinucleotides were 15.90 kb and 18.61 kb, respectively. These findings accordingly indicate that the density of SSRs in *Z. armatum* is greater than that in *Z. bungeanum*, although both these species show similar SSR distributions and frequencies. Furthermore, we also compared the SSR distribution distances in the two *Zanthoxylum* species with those in 15 species of woody and herbaceous plants (Additional file 5: Fig. S2). We accordingly found that the lowest average distance between SSRs (1.61 kb) occurred in *Camellia sinensis*, followed by the 1.68 kb in *Rosa roxburghii* and the 2.59 kb in *Hevea brasiliensis*. For *Z. armatum* and *Z. bungeanum*, the average distribution distances of SSRs were 3.91 kb and 4.24 kb, respectively, which places these two species at 4th and 5th among the 17 analyzed species. Consequently, the frequencies of SSR in *Z. armatum* and *Z. bungeanum* were not low; moreover, the SSRs in these two species tend to be abundant, indicating the potential for further discoveries.

2.4 Primer polymorphisms

In our previous study, we randomly selected 120 primer pairs for screening, among which, six pairs were used in the present study to analyze 125 samples of *Zanthoxylum* species based on capillary electrophoresis technology. Four of these primer pairs were derived from *Z. bungeanum*, whereas the other two were from *Z. armatum*. Using these primer pairs, we succeeded in amplifying 51 allelic loci with the mean number of 8.5 polymorphic loci. The average number of amplified allelic genes (N_a) was 2, the average number of effective alleles (N_e) was 1.3093, the average for N_e 's genetic diversity was 0.1916, and the Shannon information index (I) was 0.3037. On the basis of these values, particularly those for N_a and N_e , there initially appeared to be a low genetic diversity among the 125 samples.

Fig. 3. Capillary electrophoresis results of the primer pair ZB23 and ZA51.

2.5 Genetic diversity assessment

2.5.1 Population genetic diversity

Using the UPGMA method for cluster analysis of Sichuan pepper populations, we found that the observed number of allelic genes (N_a) ranged between 1 and 1.45 (average 1.15), and that the effective number of alleles (N_e) ranged from 1 to 1.27 (average 1.10). Genetic diversity (N_e) ranged from 0 to 0.15 (average 0.055), whereas Shannon's information index (I) ranged from 0 to 0.22 (average 0.081). The number of polymorphic loci ranged from 0 to 23 (average 7.46), the percentage polymorphism ranged from 0% to 45.10% (14.60%) (Table 2). In detail, for *Z. bungeanum*, *Z. armatum*, and *Z. piperitum*, average number of N_a was 1.16, 1.17, and 1.04, respectively; the average number of N_e was 1.11, 1.11, and 1.10, respectively; the average number of N_e in each case was 0.06; and the values of Shannon's I were 8.14, 8.44, and 2.00, respectively. These observations tend to indicate that different *Zanthoxylum* species have similar genetic diversity, thereby indicating a relatively low level of genetic diversity among these *Zanthoxylum* species.

On the basis of the population dendrogram shown in Fig. 4, the following information and results can be obtained. Firstly, we successfully clustered the *Zanthoxylum* samples and found that the genetic distances ranged from 0.0050 to 0.4122, with an average distance of 0.1878. Secondly, EST-SSR primers were able to distinguish and identify each of the 125 *Zanthoxylum* samples. Thirdly, the populations were divided into four cohorts. Cohort I almost exclusively comprised *Z. bungeanum* cultivars, whereas cohort II contained *Z. armatum* varieties, cohort III included 'JLHJ-ZB' and 'PTSJ-ZP' (*Z. piperitum*), and cohort IV contained only a single sample 'MXDHP-ZB' (*Z. bungeanum*). Not only were *Z. bungeanum* varieties readily detected, but it was also possible to determine that the samples originated from Longnan City, Gansu Province. Therefore, we suspect that genetic distances relate to this area. However, the members of cohort II were all *Z. armatum* varieties, which, although originating from a range of different localities, show very close genetic distances, and accordingly, do not display regionalism. The results obtained for cohort III and cohort IV were notable in that cohort III contained 'PTSJ-ZP', which is a variety of *Z. piperitum* and would have been expected to have clustered with 'CCSJ-ZP' and 'LJSJ-ZP'. We did, nevertheless, succeed in classifying a number of other varieties, including 'MXDHP-ZB', 'HYDHP-ZB', 'HCDHP-ZB', 'Yuexi Gongjiao', 'TJ-ZA', 'JYQHJ-ZA', and 'JYQHJ-ZA'. In the present study, however, the EST-SSR markers could not unambiguously authenticate *Zanthoxylum* varieties, thereby indicating that this technique is probably more applicable for the identification of species rather than varieties. Thirdly, the results indicated that clusters do not closely reflect regional diversity. Finally, the data obtained tend to indicate that *Z. bungeanum* and *Z. armatum* have low intraspecific variation and genetic differentiation. Accordingly, it would appear that the *Zanthoxylum* species examined in the present study do not show high genetic diversity.

Fig. 4. Population clustering dendrogram based on genetic distance

(Coefficient number: 0 represents the minimum genetic distance)

Further, among the data presented in Fig. 4 and Table 3, there are some interesting results that are worthy of further investigation. Firstly, the genetic distances within populations ranged from 0.0050 to 0.4122, whereas that between populations ranged from 0.1122 to 0.2664, with an average of 0.1886. Further, the coefficient of genetic similarity ranged from 0.7667 to 0.8930 (mean 0.8299). Among the samples collected, 'HYDHP-ZB' has the minimum genetic distance (0.1122) with other populations, whereas 'MXDHP-ZB' has the maximum genetic distance (0.1886). 'MXDHP-ZB' is a variety of *Z. bungeanum* but was classified in a cluster separate from those of cohort I, which contained varieties of *Z. bungeanum*. This seeming discrepancy could be attributable to contingency and errors as a consequence of using only a single sample of 'MXDHP-ZB'. (2) In cultivation 'MXDHP-ZB' has been found to differ morphologically from *Z. bungeanum* species, and the present result confirm these differences at the molecular level. 'CCSJ-ZP', 'LJSJ-ZP', and 'PTSJ-ZP' are cultivars of *Z. piperitum* and theoretically should be cluster together, although we observed that only 'CCSJ-ZP' and 'LJSJ-ZP' clustered. The genetic distance between 'CCSJ-ZP' and 'PTSJ-ZP' was 0.2134, whereas that between 'CCSJ-ZP' and 'LJSJ-ZP' was 0.2439, and that between 'PTSJ-ZP' and 'LJSJ-ZP' was 0.2429. All of these values exceeded the average genetic distance of 0.1884, thereby indicating that the genetic distances among these cultivars were relatively larger. Moreover, based on molecular marker analysis, 'Putao Shanjiao' probably differs from 'Chaochang Shanjiao' and 'Liujin Shanjiao', which should be investigated further.

Sichuan pepper is generally characterized by producing only female flowers and thus reproduction occurs almost entirely through apomixis. In contrast, however, sample 'X-ZB' produces only male flowers, whereas three 'HCDHP-ZB' samples have been exposed to irradiation in space. According to dendrogram constructed in the present study, these samples were assigned into cohort I but did not show any consistent patterns. Although a wild population of *Z. armatum* (ZA-WP) and the cultivar 'TJ-ZA' clustered together in cohort II, there was a large genetic distance between the wild and cultivated populations, indicating their dissimilarity. Finally, 'YXGJ-ZB' which clustered in cohort I, appeared to be relatively distinct from other *Z. bungeanum* varieties. To date, there have been no reports on the differentiation between 'YXGJ-ZB' and other *Z. bungeanum* varieties, and therefore it is conceivable that there are indeed differences between them which could be confirmed on the basis of morphology evidence.

2.5.2 Individual genetic diversity assessment of whole samples

Using the UPGMA method to cluster the 125 *Zanthoxylum* samples (Fig. 5), we found that the genetic distances among these 125 samples ranged from 0 to 2.3811, with an average of 0.5915. The samples could be clustered into five cohorts. Cohort I contained *Z. bungeanum* and its varieties. Cohort II included *Z. armatum* and its varieties. Cohort III mainly contained *Z. piperitum* in addition to two *Z. bungeanum* samples. Cohort IV and cohort V each contained only two samples respectively, with cohort IV containing sample 59 ['Laiwu Huajiao' (*Z. bungeanum*)] and sample 68 ['Hancheng Dahongpao' (*Z. bungeanum*)], and Cohort V containing sample 10 ['Yuexi Gongjiao' (*Z. bungeanum*)] and sample 115 ['Putao Shanjiao' (*Z. piperitum*)]. The construction of individual dendrograms produced the following results and associated problems. Firstly, on an individual basis, this clustering was successful, with respect to the separation

of *Z. bungeanum*, *Z. armatum*, and *Z. piperitum* and their varieties. The results generally indicate that the genetic diversity of *Zanthoxylum* species is not high, which is possibly related to the apomictic mode of reproduction, which generally confers stable phenotypes. Additionally, we found EST-SSR primers were more effective in differentiating species than varieties, because intraspecies of *Z. bungeanum* and *Z. armatum* showed very small genetic distances, indicating little difference among varieties. For instance, samples 73 ('Maowen Dahongpao'), 87 ('Wuxuan No.1'), 89 ('Wuxuan No.2'), and 92 ('Wuxuan No.3') in cohort I; and samples 12 ('Jinyang Qinghuajiao'), 19 ('Jiuye Qinghuajiao'), 34 [*Z. armatum* 'Tengjiao' (Hongya)], 35 [*Z. armatum* 'Tengjiao' (E'mei)], 43 [*Z. armatum* 'Tengjiao' (E'mei)], 44 [*Z. armatum* 'Tengjiao' (Danleng)], 51 ('Jinyang Qinghuajiao'), and 53 ('Jinyang Qinghuajiao') in cohort II. However, although the EST-SSR primers used in this study were able to identify different *Zanthoxylum* species, they were unable to clearly distinguish varieties clearly. Secondly, *Z. armatum* 'Tengjiao' in cohort II had small genetic distances and, given that these samples were collected from different areas, this could indicate that the genes of *Z. armatum* 'Tengjiao' are stable. Thirdly, samples 123-125 ('Hancheng Dahongpao', which has been exposed to irradiation in space) had small genetic distances and showed no obvious differences compared with unirradiated 'Hancheng Dahongpao', indicating that at the gene level, the samples that had been exposed to irradiation could not be distinguished based on molecular marker analysis using EST-SSRs. Moreover, cohort IV contained only two samples, namely samples 10 ('Yuexi Gongjiao') and 115 ('Putao Shanjiao'). Theoretically, 'Yuexi Gongjiao' should be classified in cohort I and 'Putao Shanjiao' should be classified in cohort III because they differ in term of morphological characteristics. Although we found that 'Jinyang Qinghuajiao' exhibited obvious differences from *Z. armatum*, they are considered very similar to each other in practical cultivation. Finally, we found that wild *Z. armatum* samples from Ya'an City and Mian Yang City in Sichuan Province have small genetic distances, which indicates that they are very similar, whereas genetic distances between these samples and *Z. armatum* 'Tengjiao' were relatively larger, indicating that cultivated varieties *Z. armatum* show certain genetic differences compare with wild *Z. armatum*.

Fig. 5 Individual clustering dendrogram based on genetic distance (Coefficient number: 0 represents the minimum genetic distance)

Discussion

3.1 Transcriptome sequencing

Although species of *Zanthoxylum* are economically important trees in China [15], there have been relatively few studies on the transcriptomes and genetic diversity of these species [16]. Moreover, to date, there have been limited whole-genome sequencing studies, and as such, reference genes are lacking. Hence, the development of analytic methods based on functional gene, SSR, and SNP techniques would be highly desirable, both from the perspective of advancing scientific knowledge and from the potential practical implications. Given the current paucity of relevant data, in the present study, we used the Illumina-Hiseq sequencing platform to sequence the genomes of *Z. bungeanum* and *Z. armatum*, and accordingly obtained a large quantity of precise data that provided an excellent basis for the screening using SSR markers and detection of genetic diversity. Following quality control of the raw data, we obtained in excess of 0.245 billion high-quality sequences. In total, we generated 36.76 G of high-quality clean data, with accuracy Q20 and Q30 accuracies exceeding 97% and 93%, respectively. Additionally, regardless of the source of raw data or filterable data, data sizes per sequencing sample were greater than 30 M. Overall, we found that transcriptome results for each sample were reliable and very good, with all samples providing valuable data for subsequent screening using SSR markers and genetic diversity analyses. On the basis of analysis using GO, KEGG, and other databases, we annotated the unigenes of *Z. bungeanum* and *Z. armatum* according to different functions and metabolic pathways. For *Z. bungeanum*, 89,198 unigenes were mapped using the GO database, with percentages relative to the total number of unigenes ranging from 8.91% to 17.48%. Specifically, 57,683 unigenes were assigned to biological process with the highest percentage of 17.48%. In *Z. armatum*, 13,7567 unigenes were annotated using the GO database, with percentages relative to the total number of unigenes of between 6.43% and 12.75%. Among these, 66,368 were mapped to biological process with the highest percentage of 12.75%. On the basis of KEGG database analyses, a total of 25,520 unigenes were assigned to different pathways, with percentages ranging from 9.43% to 47.36%. Among these, 12,520 unigenes were annotated to cohort A (cellular processes) with percentages reaching 47.36%. For *Z. armatum*, 27,917 unigenes were assigned to different KEGG pathways with percentages ranging from 8.83% to 50.02%. Among these, the highest number (13,964) of unigenes were mapped to cohort A. Annotation using the GO database is used to indicate the functions of gene [17], whereas KEGG-based analysis can be used to assess gene product functions and associated metabolic pathways [18]. In the present study, we found the numbers of annotated unigenes and annotation percentage were similar for *Z. bungeanum* and *Z. armatum*, in which these data proved to have consanguinity and similarity.

3.2 EST-SSR validation

One advantage of using the RNA-seq technique to obtain EST-SSR primers is that these markers possess better polymorphism and specificity than do common primers. For classification and identity, it is preferable to utilize EST-SSR due to their cDNA origin and conserved nature [19], which are considered desirable properties of suitable molecular markers. However, there have been few previous studies that examined the transcriptome of *Zanthoxylum*. Recently, in China, Zhao et al. reported the results of transcriptome-based analysis of sprouting 'Chaocang Shanjiao' (*Z. piperitum*) [20]. However, although the origin and evolution of *Zanthoxylum* have been studied extensively and some of EST-SSRs have been discovered [21], Zhao et al. focused on only one species of *Zanthoxylum*. In the present study, we performed transcriptome analyses for two of the main commercially cultivated *Zanthoxylum* species in China, namely, *Z. bungeanum* and *Z. armatum*. On the basis of pre-screening for EST-SSR markers, we selected four and two EST-SSR pairs for detection of genetic diversity in *Z. bungeanum* and *Z. armatum*, respectively. We accordingly found that all of six pairs of SSRs exhibited excellent polymorphism in at least three species of *Zanthoxylum*, which is crucial with respect further characterization of *Zanthoxylum*.

3.3 Genetic diversity assessment

At the species level, EST-SSRs have been shown to be accurate and produce clear results with respect to the identity and genetic diversity of crop plants such as *Triticum* [22], *Hordeum vulgare* [23], and *Citrus* [24]. However, at the cultivar level, this technique often fails to yield clear results, as exemplified by the study on *Zanthoxylum* by Zhao et al. [25]. Nevertheless, the findings of the present study indicate that at the level of both species and variety, there appears to be relatively low genetic diversity among the three species *Z. bungeanum*, *Z. armatum*, and *Z. piperitum*. Moreover, we were unable to detect any substantial regional variation, which can probably be attributed to the parthenocarpic mode of reproduction, which is a notable characteristic of these species and has been widely investigated in other species [26]. Apomixis is considered a primary factor contributing to this category of result, and because reproduction in most *Zanthoxylum* species occurs via apomixis, this is probably a key factor associated gene stability and low genetic diversity. According to characteristics of genetic diversity and market requirements of *Zanthoxylum* in China, the range of requirements has increased, as has the scales of cultivation. As a consequence of differences in climatic conditions in China, it has been found that differences in *Z. bungeanum* cultivated in north and *Z. armatum* cultivated in south is becoming increasing more obvious [4]. Hence, as *Z. bungeanum*, *Z. armatum*, and their cultivars are the main commercially grown Sichuan pepper in China, considerable efforts are being made to maintain the abundant diversity of *Zanthoxylum* with respect to characteristic such as aroma, stingless and pungent.

In the population dendrogram constructed in this study, we found that 'HYDHP-ZB' showed the lowest average genetic distance with other populations (0.1122), whereas in contrast, 'MXDHP-ZB' has the highest average genetic distances with other populations (0.2664). *Zanthoxylum* is cultivated in numerous areas of China, and the main and traditional production areas where there has been a long history of *Zanthoxylum* cultivation are assumed to be sites of cultivation origin. As reported in the literature, the utilization of *Zanthoxylum* in China can be traced back to the 11th century B.C., at which time these trees were planted in Sichuan Province (27). Thereafter, species of *Zanthoxylum* were introduced into other areas. Therefore, the genome of these species has retained similarity with those in the north, as the species has spread to other regions and different cultivars have been bred. Consequently, this may provide a suitable explanation for the minimum mean genetic distances of 'HYDHP-ZB' with other *Zanthoxylum* samples. 'PTSJ-ZP', which is a cultivar of *Z. piperitum*, not only clustered with 'CCSJ-ZP' and 'LJSJ-ZP' in the population dendrogram but also in the individual dendrogram. Although many *Zanthoxylum* species reproduce by vegetative propagation, cross-pollination has been observed in *Z. piperitum*. Therefore, although 'Putao Sanjiao,' like 'CCSJ-ZP' and 'LJSJ-ZP', is a cultivar of *Z. piperitum*, it may show a certain degree of genetic variation from these other cultivars. Furthermore, given that only a single sample of 'PTSJ-ZP' was collected in the present study, whereas we collected three samples each of 'CCSJ-ZP' and 'LJSJ-ZP', further studies will be necessary to confirm the findings in this regard. Although both 'HCDHP-ZB' and 'X-ZB' (which produces only male flowers) were exposed to irradiation in space, they did not show any abnormal clustering compared with the other *Z. bungeanum* samples. Consequently, it would seem that neither extraterrestrial radiation nor the production of male flowers has a significant impact on the genetic structure of these species, or simply that we were unable to detect differences based on SSR analysis. With respect differences between wild and cultivated populations of *Zanthoxylum*, we found that wild *Z. armatum* and the *Z. armatum* cultivar 'TJ-ZA' show a relatively high genetic distance, indicating clear genetic differences between wild and cultivated populations of the same species. Moreover, 'JLHJ-ZB', which is a cultivar of *Z. bungeanum*, should theoretically cluster in the same cohort as *Z. bungeanum*. However, in both population and individual clustering dendrograms, this cultivar clustered separately from the *Z. bungeanum* branch. This could be explained by the fact that 'JLHJ-ZB' is only planted in Jiulong Prefecture, Sichuan Province, which is characterized by a mountainous terrain that would possibly provide a barrier to gene flow.

Conclusions

In this study, we obtained transcriptome data for *Z. bungeanum* and *Z. armatum*. More than 300 million reads per sample, a total of 0.258 billion reads, in excess of 4 G clean bases and base pair reads of up to 38.76 G were obtained regardless of whether *Z. bungeanum* or *Z. armatum* was used as a source of raw data. After quality control, 0.245 billion reads of high quality were maintained, and 36.76 G high-quality clean data were used for subsequent analyses. We used six pairs of EST-SSR primers to determine the genetic diversity of 125 samples. Our results indicated that these EST-SSR primers were able to identify *Z. bungeanum*, *Z. armatum*, and *Z. piperitum*. According to the population dendrogram constructed in this study, there were an average of four samples per population. The number of allelic genes (N_a) ranged from 1 to 1.4510 (average 1.1462), the number of effective allelic genes (N_e) ranged from 1 to 1.2744 (average 1.0971), genetic diversity (N_e/I) ranged from 0 to 0.1486 (average 0.05512), and Shannon's genetic diversity (I) ranged from 0 to 0.2177 (average 0.08123). Generally, the 125 samples were clustered successfully, and overall we found that the *Zanthoxylum* samples examined in this study show a relatively low level of genetic diversity. In line with expectations, different populations and individuals were divided according to species, although they did not appear to show any regional variation. Notably, however, the EST-SSR primers used in this study were not effective in distinguishing between the various cultivars of *Zanthoxylum*.

Methods

5.1 Materials for RNA-sequencing

Materials for RNA-sequencing were *Z. bungeanum* 'Hanyuan' (Accession: Sichuan Province, S-SV-ZB-00302012) and *Z. armatum* 'Tengjiao' (Accession: Sichuan Province, S-SV-ZA-001-2014). Both of these varieties collected at College of Forestry, Sichuan Agricultural University, Wenjiang District, Chengdu City, Sichuan Province, in September, 2016. All specimens were selected randomly, three fresh and healthy leaves collected from three different trees and immediately frozen in liquid nitrogen. After returning to laboratory, all samples were storing in -80°C.

5.2 Materials for genetic diversity identity

A total amount of 125 *Zanthoxylum* samples for genetic diversity were collected from Sichuan Province, Shanxi Province and Gansu Province in China during 2015-2017 (Fig. 1). All the samples included three species: *Z. bungeanum* (70), *Z. armatum* (48) and *Z. piperitum* (7) (Additional file 1: Table S1). In addition, 6 pairs of EST-SSR primer which have been screened by our research[11] (Additional file 2: Table S2).

Fig. 1. Map showing the location of sites at which 125 *Zanthoxylum* samples were collected.

5.3 Equipment and reagents

Each of the following items was obtained from Beijing Solarbio Science & Technology Co., Ltd: an Illumina Hiseq 2000 platform sequencer, a SANYO SIM-F14 automatic snow ice machine, an LDZX-40B vertical automatic electric pressure steam sterilizer, a TGL-16G high speed refrigerated centrifuge, a GELI-DOC2000 gel imaging analyzer, a Bio-Rad polymerase chain reaction device, a 3730 DNA Analyzer, an RNAprep Pure Plant Kit (DP441, TIANGEN BIOTECH (Beijing) CO. LTD), a DNA secure Plant Kit (TIANGEN BIOTECH (Beijing) CO. LTD, DP320), double-distilled water, liquid nitrogen, agarose, absolute ethanol, and Golden View (color developing agent).

5.4 RNA isolation and transcriptome sequencing

RNAs were isolated from of three fresh leaves of *Z. bungeanum* and *Z. armatum* according to the instructions of an RNAprep Pure Plant Kit. Samples of leaves (50–100 mg) were ground to a powder in liquid nitrogen. Following the addition of β -mercaptoethanol, 500 μ L of lysate was placed in tubes and rotated to mix. The lysates were then centrifuged at 12,000 rpm for 2 min and the resulting supernatants transferred to a fresh RNase-free tube, to which was added a 0.4 volumes of absolute ethanol. Following precipitation, the solution was transferred to an adsorption column followed by centrifugation at 12,000 rpm for 15 s. The resulting supernatant was discarded and 350 μ L of deproteinized solution was added to the pellet followed by centrifugation for 15 s at 12,000 rpm. Thereafter, 80 μ L of DNase I was added to the column and left to stand for 15 min. A 350- μ L volume of deproteinized solution was then added to the column, followed by centrifugation for 15 s at 12,000 and the resulting supernatant was discarded. The column was then washed with 500 μ L of buffer (in ethanol), followed by centrifugation for 15 s at 12,000 rpm. After a further centrifugation for 2 min at 12,000 rpm, the column was placed in a fresh RNase-free tube, followed by the addition of 50 μ L of RNase-free H₂O and centrifuged for 1 min at 12,000 rpm to obtain the RNA solution. The RNA product was checked by 1% agarose gel electrophoresis at a voltage of 120 V for 15

min. The concentration and purity of the isolated RNA were determined using a BioTek-Epoch microplate reader (BioTek Instruments, Inc., Beijing), and samples were sent to the Biozeron Company (Beijing) for sequencing.

Trinity software was used to assemble reads based on the following steps: (1) discard adaptor sequences in reads and non-relevant reads; (2) discard low-quality bases at the at 3' terminus; (3) discard reads with an N percentage greater than 10%; and (4) discard reads with lengths of less than 70 bp after the removal of adapters.

5.5 DNA isolation

Fresh leaf tissues (100 mg) of 125 *Zanthoxylum* samples were placed in mortar with liquid nitrogen and ground to a powder. The powder was placed in centrifuge tubes, to which buffer LP1 and 6 µL RNase A were added, followed by vortex mixing for 1 min. Thereafter, 130 µL of buffer LP2 was added followed by centrifugation for 5 min at 12,000 rpm. The resulting supernatant was mixed with 1.5 volumes of buffer LP3 followed by rotation mixing for 15 s. The solution and sediment were subsequently transferred to an adsorption column, which was centrifuged at 12,000 rpm for 30 s, after which the supernatant was discarded. Thereafter, 600 µL of buffer PW was added to the column, followed by centrifugation for 30 s at 12,000 rpm. The resulting supernatant was discarded, and the process repeated. Following a further centrifugation at 12,000 rpm for 2 min, the supernatant was discarded, and the column was dried at room temperature for 2 min. Having dried, the column was placed into a fresh 1.5-mL tube, followed by the addition of 75 µL TE eluant. After standing at room temperature (20-30°C) for 2 min, the column was centrifuged for 2 min at 12,000 rpm to obtain a DNA solution.

5.6 PCR amplification using SSR primers

The 20 µL reaction mixtures used for PCR amplification (Biorad C1000 Touch) comprised the following components: 14.8 µL ddH₂O, 0.4 µL dNTPs, 2 µL buffer, 0.3 µL forward primer (20 µM), 0.3 µL reverse primer (20 µM), 0.2 µL DNA, and 0.2 µL Taq polymerase. The PCR program used was as follows: 94°C pre-denaturation, 5 min; followed by 35 cycles of 94°C denaturation, 30 s; 55°C renaturation, 35 s; and 72°C extension, 40 s; and a final extension at 72°C, 3 min; followed by storage at 4°C.

The Msatcommander, software program MISA (<http://pgrc.ikp-gatersleben.de/misa>) was used to detect SSR loci and primers were designed using Primer 5.0. Initially, primers were designed from the 0–1000 bp sequences upstream and downstream of the detected SSR. Secondly, SSRs sequences repeated at least 11 for mononucleotides and six times for dinucleotides to assign, respectively, whereas SSRs sequences at least five repeats to assign for trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides. Finally, the length of amplification products should be 100–280 bp and annealing temperatures for the designed primers ranged from 55°C to 65°C.

5.7 Assessment of capillary electrophoresis fluorescence labeling data

In this study, we used a 3730XL sequenator for capillary electrophoresis fluorescence labeling. Formamide and markers were initially mixed at ratio of 100:1 by volume, and 9 µL of PCR product was transferred into capillary electrophoresis fluorescence detection system. The PCR product was then diluted 10 times and 1 µL added. Fragment (plant) analysis software was used to assess the raw data through comparing the molecular weights of PCR products molecular markers sizes.

5.8 Treatment of genetic diversity data

Raw data were analyzed using Gene Marker 2.2. Peak sizes were compared with markers to determine the sizes of fragments. SSRs are dominant markers that enable the sizes of products to be established. If different samples contain the same products, they are assumed to be homologous. For all data, the results were converted into a data matrix comprising "0" and "1" indicating no production and production, respectively.

NTSYSpc 2.11 software was used to calculate the genetic distance of populations and UPGMA was used to analyze clusters and construct evolutionary trees. Moreover, indicators of genetic diversity, including Percentage of Polymorphic Loci (PPL); Nei's Gene Diversity (H) [12], and Shannon's information index (H') [13], were determined using POPGENE32. Nei's standard genetic distance was determined using the following formula:

$$D_s = -L_n [J_{xy} / (J_x J_y)^{1/2}]$$

X_{ij}, Y_{ij}: Gene frequency of No. K allele on loci No. r in population X and population Y

Declarations

Additional Files:

Additional file 1: Table S1. Information on 125 *Zanthoxylum* samples

Additional file 2: Table S2. Information on six pairs of EST-SSR primers and genetic diversity results

Additional file 3: Table S3. Transcriptome raw and clean data

Additional file 4: Figure S1. GO (Gene Ontology) annotation for *Zanthoxylum bungeanum* and *Zanthoxylum armatum*

Additional file 5: Figure S2. Comparison of SSR distances between different species

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interest

The author declares that they have no competing interests.

Funding

The research was supported by the Program Great Varieties Cultivation of Traditional Chinese Medicine under grant number 2018SZ0064.

Authors' contributions

Yangchuan Deng and Li Xiang as joint-first authors contributed equally to this research. Yangchuan Deng performed the main the trial, conducted data analyses, and drafted the manuscript, Li Xiang revised the manuscript and provided important advice. Meng Ye and Shilin Chen as joint-corresponding authors came up with some key points, ideas, and financial support and apparatus. Prof. Wei Gong provided some important suggestions for the revision of this manuscript, Tingyu Ma and Li Yang assisted with some trial work, and Xiangxiao Meng provided software assistance. All the authors read and modified this paper.

Author details

¹Yangchuan Deng is a master at the Sichuan Agricultural University of China.

²Li Xiang is an [associate researcher](#) working at the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences

¹Wei Gong is a professor at the Sichuan Agricultural University of China.

²Tingyu Ma is a doctor at the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences.

¹Li Yang is a master at the Sichuan Agricultural University of China.

²Xiangxiao Meng is a reserach assistant at the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences.

²Shilin Chen is the Director of the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences.

¹Meng Ye is the professor at the Sichuan Agricultural University of China.

¹College of Forestry, Sichuan Agricultural University, Huimin Road, Wenjiang District, Chengdu, China. ²Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Dongcheng District, Beijing, China.

References

- [1] Zhang DX, Thomas G, Hartley. Flora of China. Beijing: Science Press. 2008,11.
- [2] Bryant BP, Mezzine I. Alkylamides that produce tingling paresthesia activate tactile and thermal trigeminal neurons. Brain Res. 1999;842(2):452-460.
- [3] Naito H. The herbal medicine Dai-Kenchu-Tou stimulates upper gut motility through cholinergic and 5-hydroxytryptamine 3 receptors in conscious dogs. Surgery. 1999;126(5):918-924.
- [4] Zhang H, Ye M. Research Status on the Taxonomic and Component of Green Zanthoxylum bungeanum Maxim. Nor. Horticult. 2010;(14):918-924.
- [5] Zhu L, Chen X, Li JH, Jiang WD, Zhong CP, Zhou J, Wu CJ. Textual research on origin of Zanthoxylum armatum and consideration of its definition. Chin. Tradi. Herbs. Drugs. 2018;(4):987-992.
- [6] Timothy AH, John TC, Linda M, Natalie H, Robert JH. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol. Breeding. 2002;9(2):63-71.
- [7] Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. 2000;100(5):723-726.
- [8] Ramesh V, Kantety, Mauricio LR, David E, Matthews, Mark E, Sorrells. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 2002;48(5-6):501-510.
- [9] Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 2004;109(1):122-128.
- [10] Chun XC, Ping Z, Young AC, Shu H, Fred G. Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. 2006;112(7):1248-1257.
- [11] Deng YC, Xiang L, Su YY, Yang L, Li HY, Luo XM, Sun W, Ye M. Development of EST-SSR markers of Zanthoxylum species based on RNA-sequencing. J. Northwest A&F Univ. 2019;(4):16-24.
- [12] Nei M. Genetic Distance between Populations. AM. Nat. 1972;106(949):283-292.
- [13] Lewontin RC. Testing the Theory of Natural Selection. Nature, 1972;236(5343):181-182.
- [14] Li ZL, Pan KW, Tariq A, Sun F, Wang SZ, Zhang LS, Sun XM, Wu XG, Song DG. Legume plants may facilitate Zanthoxylum bungeanum tolerance to extreme rainfall. Sci. Rep-UK. 2018;8(1).
- [15] Choi, Hwa J. Evaluation of Antiviral Activity of Zanthoxylum Species Against Picornaviruses. Osong Public Health and Research Perspectives, 2016;7(6):400-403.
- [16] Alam F, Us Saqib QN. Evaluation of Zanthoxylum armatum Roxb for in vitro biological activities. J. Tradit. Chin. Med. 2017:S2225411017300111.

- [17] Matthew DY, Matthew JW, Gordon KS, Alicia O. Gene ontology analysis for RNA-SEQ: Accounting for selection bias. *Genome Biol.* 2010;11(2):R14.
- [18] Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2007;36:D480-D484.
- [19] Varshney RK, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular & Molecular Biology Letters.* 2002;7(2A):537.
- [20] Zhao YC, Liu XW, Li Y, Zhao DG. Transcriptome Data Assembly and Gene Function Annotation of *Zanthoxylum piperitum* DC var. *inermis* Makino. *Genom. Appl. Biol.* 2016;7(7):1805-1819.
- [21] Feng SJ, Zhao LL, Liu ZS, Liu YL, Yang TX, Wei AZ. De novo transcriptome assembly of *Zanthoxylum bungeanum* using Illumina sequencing for evolutionary analysis and simple sequence repeat marker development. *Sci. Rep-UK.* 2017;7(1):16754.
- [22] Eujayl I, Sorrells M, Baum M, Wolters P, Powell W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRs. *Euphytica*, 2001, 119(1-2):39-43.
- [23] Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 2003;106(3):411-422.
- [24] Dong J, Guang YZ. Analysis of Microsatellites in *Citrus Unigenes*. *Act. Genet. Sin.* 2006; 33(4):345-353.
- [25] Zhao LL, Feng SJ, Tian JY, Wei AZ, Yang TX. Internal transcribed spacer 2 (ITS2) barcodes: A useful tool for identifying Chinese *Zanthoxylum*. *Appl. Plant. Sci.* 2018;6(6):e01157.
- [26] Zhan X, Li P, Hui WK, Deng YW, Gan SM, Sun Y, Zhao XH, Chen XY, Deng XM. Genetic diversity and population structure of *Toona ciliata* revealed by simple sequence repeat markers. *Biotechnol. Biotec. EQ.* 2019;(5):1-9.
- [27] Lin HR. Preliminary research on history of *Zanthoxylum bungeanum* Maxim. [Agr. Hist. Chin.](#) 1985(2):63-67.

Tables

Table 1. Comparison of the characteristics of *Zanthoxylum bungeanum* and *Zanthoxylum armatum*

| Species | SSR | Repeat times | | | | | | | Number | Ratio /% |
|---------------------|-----------------|--------------|-------|-------|------|------|------|-------|--------|-------------|
| | | 5 | 6 | 7 | 8 | 9 | 10 | >10 | | |
| <i>Z. bungeanum</i> | Type of SSR | | | | | | | | | |
| <i>Z. armatum</i> | | | | | | | | | | |
| | Mononucleotide | - | - | - | - | - | - | 6 296 | 6,296 | 49.40 |
| | Dinucleotide | - | 1,062 | 606 | 434 | 286 | 199 | 405 | 2,992 | 23.47 |
| | Trinucleotide | 1,498 | 708 | 404 | 288 | 34 | 58 | 102 | 3,092 | 24.26 |
| | Tetranucleotide | 150 | 49 | 5 | 5 | 3 | - | 2 | 214 | 1.68 |
| | Pentanucleotide | 41 | 4 | 1 | 1 | - | - | - | 47 | 0.37 |
| | Hexanucleotide | 56 | 21 | 15 | 8 | 2 | 2 | 1 | 105 | 0.82 |
| | Total | 1,745 | 1,844 | 1,031 | 736 | 325 | 259 | 6,806 | 12,746 | |
| | Ratio/% | 13.69 | 14.47 | 8.09 | 5.77 | 2.55 | 2.03 | 53.40 | | |
| | Mononucleotide | - | - | - | - | - | - | 7,830 | 7,830 | 51.87 |
| | Dinucleotide | - | 1,255 | 668 | 454 | 304 | 209 | 279 | 3,169 | 20.99 |
| | Trinucleotide | 1,754 | 966 | 555 | 297 | 32 | 25 | 82 | 3,711 | 24.58 |
| | Tetranucleotide | 172 | 39 | 1 | 3 | 2 | 4 | 2 | 223 | 1.48 |
| | Pentanucleotide | 43 | 5 | 3 | - | - | - | - | 51 | 0.34 |
| | Hexanucleotide | 57 | 27 | 16 | 7 | 2 | 1 | 2 | 112 | 0.74 |
| | Total | 2,026 | 2,292 | 1,243 | 761 | 340 | 239 | 8,195 | 15,096 | |
| | Ratio/% | 13.69 | 14.47 | 8.09 | 5.77 | 2.55 | 2.03 | 53.40 | | |

Table 2. Data for genetic diversity of species in the population dendrogram

| Species | Population | Abbreviation | Sample number | Observed number of alleles (<i>N_a</i>) | Effective number of alleles (<i>N_e</i>) | Genetic diversity (<i>N_e</i>) | Shannon's Information index (<i>I</i>) | Polymorphic loci |
|---------------------------|--|---------------------|---------------|--|---|---|---|------------------|
| <i>Z. bungeanum</i> | 'X' | X-ZB | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| <i>Z. bungeanum</i> Z. | <i>Z. bungeanum</i> (male flowers only) | (male flowers only) | | | | | | |
| | 'Bayuejiao' | BYJ-ZB | 3 | 1.1176 | 1.0812 | 0.0463 | 0.0681 | 6 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Baishajiao' | BSJ-ZB | 2 | 1.0784 | 1.0555 | 0.0325 | 0.0474 | 4 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Dahongpao' | DHP-ZB | 3 | 1.0196 | 1.0187 | 0.0096 | 0.0134 | 1 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Doujiao' | DJ-ZB | 3 | 1.1765 | 1.1270 | 0.0713 | 0.1042 | 9 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Erhongpao' | EHP-ZB | 3 | 1.1569 | 1.1083 | 0.0618 | 0.0908 | 8 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Fengjiao' | FJ-ZB | 3 | 1.1569 | 1.1289 | 0.0692 | 0.0988 | 8 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Hancheng Dahongpao' | HCDHP-ZB | 3 | 1.3725 | 1.2624 | 0.1486 | 0.2177 | 19 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Hanyuan Dahongpao' | HYDHP-ZB | 3 | 1.3333 | 1.2044 | 0.1220 | 0.1829 | 17 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Hebei Wuci' | HBWC-ZB | 2 | 1.0784 | 1.0555 | 0.0325 | 0.0474 | 4 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Honghuajiao' | HHJ-ZB | 3 | 1.1373 | 1.1102 | 0.0596 | 0.0855 | 7 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Laiwu' | LW-ZB | 9 | 1.4510 | 1.2229 | 0.1355 | 0.2089 | 23 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Linzhou Dahongpao' | LZDHP-ZB | 3 | 1.2941 | 1.1980 | 0.1140 | 0.1683 | 15 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Jiulong Huajiao' | JLHJ-ZB | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Maowen Dahongpao' | MWDHP-ZB | 3 | 1.1373 | 1.1102 | 0.0596 | 0.0855 | 7 |
| | <i>Z. bungeanum</i> | | | | | | | |
| | 'Maoxian Dahongpao' | MXDHP-ZB | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |

| | | | | | | | | |
|----------------------|-----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Meihujiao' | MHJ-ZB | 3 | 1.1373 | 1.1205 | 0.0633 | 0.0895 | 7 |
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Mianjiao' | MJ-ZB | 3 | 1.0784 | 1.0439 | 0.0272 | 0.0414 | 4 |
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Qin An No.1' | QA1-ZB | 3 | 1.2157 | 1.1644 | 0.0905 | 0.1309 | 11 |
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Hancheng Dahongpao' | HCDHP(I)-ZB | 3 | 1.2745 | 1.2102 | 0.1155 | 0.1669 | 14 |
| | (space irradiation) | | | | | | | |
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Wuxuan' | WX-ZB | 9 | 1.2157 | 1.0831 | 0.0507 | 0.0811 | 11 |
| | <i>Z.bungeanum</i> | | | | | | | |
| | 'Youjiao' | YJ-ZB | 3 | 1.0784 | 1.0542 | 0.0309 | 0.0454 | 4 |
| | <i>Z.bungeanum</i> | | | | | | | |
| | Average | \ | 3.18 | 1.16 | 1.11 | 0.06 | 0.09 | 8.14 |
| AA. <i>armatum</i> | <i>Z.artmatum</i> | TJ(DL)-ZA | 10 | 1.2353 | 1.1283 | 0.0738 | 0.1113 | 12 |
| <i>Z.armatum</i> | 'Tengjiao' | | | | | | | |
| AB. <i>Z.armatum</i> | (Danleng) | | | | | | | |
| | <i>Z.artmatum</i> | TJ(HY)-ZA | 3 | 1.0392 | 1.0374 | 0.0191 | 0.0267 | 2 |
| | 'Tengjiao' | | | | | | | |
| | (Hongya) | | | | | | | |
| | <i>Z.artmatum</i> | TJ(EM)-ZA | 9 | 1.1176 | 1.0499 | 0.0305 | 0.0481 | 6 |
| | 'Tengjiao' | | | | | | | |
| | (Emei) | | | | | | | |
| | 'Erwai huajiao' | EWHJ-ZA | 3 | 1.2941 | 1.1980 | 0.1140 | 0.1683 | 15 |
| | <i>Z.armatum</i> | | | | | | | |
| | 'Jinyang Qinghuajiao' | JYQHJ-ZA | 8 | 1.1765 | 1.1253 | 0.0685 | 0.0998 | 9 |
| | <i>Z.armatum</i> | | | | | | | |
| | <i>Z.artmatum</i> | | | | | | | |
| | 'Qinghuajiao' | QHJ-ZA | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| | <i>Z.armatum</i> | | | | | | | |
| | 'Zhaotong Huajiao' | ZTHJ-ZA | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| | <i>Z.armatum</i> | | | | | | | |
| | 'Jiuye Qinghuajiao' | JYQHJ-ZA | 6 | 1.3529 | 1.2744 | 0.1486 | 0.2138 | 18 |
| | <i>Z.armatum</i> | | | | | | | |
| | <i>Z.armatum</i> | ZA(WP) | 6 | 1.2745 | 1.1355 | 0.0840 | 0.1301 | 14 |

| | | (wild population) | | | | | | |
|---------------------|---------------------|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Average | | \ | 5.22 | 1.17 | 1.11 | 0.06 | 0.09 | 8.44 |
| <i>Z. pipertum</i> | ChaocangShanjiao | CCSJ-ZP | 3 | 1.1176 | 1.0915 | 0.0500 | 0.0721 | 6 |
| <i>Z. piperitum</i> | <i>Z. piperitum</i> | | | | | | | |
| | 'Liu jin Shanjiao' | LJSJ-ZP | 3 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| | <i>Z. Piperitum</i> | | | | | | | |
| | 'Putao Shanjiao' | PTSJ-ZP | 1 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 0 |
| | <i>Z. piperitum</i> | | | | | | | |
| Average | | \ | 2.33 | 1.04 | 1.03 | 0.02 | 0.02 | 2.00 |
| Average | \ | 4 | 1.15 | 1.10 | 0.06 | 0.08 | 7.46 | |

Table 3. The maximum and minimum genetic distances between different populations

| Name of populaton | Average genetic distance with other population | Population with maxmum genetic distance | Population with minimum genetic distance |
|-----------------------------|--|---|--|
| X-ZB (male flowers only) | 0.2312 | 0.3483 MXDHP-ZB | 0.0878 BYJ-ZB |
| BYJ-ZB | 0.1857 | 0.3411 MXDHP-ZB | 0.0878 X-ZB (male flowers) |
| BSJ-ZB | 0.1993 | 0.3543 ZTHJ-ZA | 0.1115 HYDHP0-ZB |
| CCSJ-ZP | 0.2062 | 0.3516 QHJ-ZA | 0.0922 HYDHP-ZB |
| DHP-ZB | 0.1974 | 0.4122 JYQHJ-ZA | 0.0229 WX-ZB |
| TJ(DL)-ZA | 0.1857 | 0.3557 EHP-ZB | 0.0050 JYQHJ-ZA |
| DJ-ZB | 0.1890 | 0.3545 ZTHJ-ZA | 0.0550 HCDHP-ZB |
| TJ(EM)-ZA | 0.2125 | 0.3557 EHP-ZB | 0.0079 TJ(HY)-ZA |
| EHP-ZB | 0.1898 | 0.3968 QHJ-ZA | 0.0561 DJ-ZB |
| EWJH-ZA | 0.1630 | 0.2779 MXDHP-ZB | 0.0468 JYQHJ-ZA |
| FJ-ZB | 0.1501 | 0.2776 MXDHP-ZB | 0.0360 MJ-ZB |
| HCDHP-ZB | 0.1512 | 0.2819 QHJ-ZA | 0.0540 HYDHP-ZB |
| HYDHP-ZB | 0.1122 | 0.1723 MXDHP-ZB | 0.0476 JYQHJ-ZA |
| HBWC-ZB | 0.2172 | 0.3426 QHJ-ZA | 0.0774 DJ-ZB |
| HHJ-ZB | 0.1809 | 0.2937 X-ZB (male flowers) | 0.0762 HCDHP(I)-ZB |
| TJ(HY)-ZA | 0.2325 | 0.3953 EHP-ZB | 0.0079 TJ(EM)-ZA |

| | | | |
|-------------|--------|-----------|-----------|
| JYQHJ-ZA | 0.1855 | 0.3141 | 0.0050 |
| | | DJ-ZB | TJ(DL)-ZA |
| JLHJ-ZB | 0.2613 | 0.4122 | 0.1252 |
| | | DHP-ZB | PTSJ-ZP |
| LW-ZB | 0.1349 | 0.2855 | 0.0494 |
| | | MXDHP-ZB | HYDHP-ZB |
| LZDHP-ZB | 0.1491 | 0.2647 | 0.0478 |
| | | ZTHJ-ZA | HYDHP-ZB |
| LJSJ-ZP | 0.2171 | 0.3483 | 0.0166 |
| | | QHJ-ZA | CCSJ-ZP |
| MWDHP-ZB | 0.1524 | 0.2911 | 0.0335 |
| | | JLHJ-ZB | QA1-ZB |
| MXDHP-ZB | 0.2664 | 0.3765 | 0.1723 |
| | | YXGJ-ZB | HYDHP-ZB |
| MHJ-ZB | 0.1489 | 0.2864 | 0.0077 |
| | | ZTHJ-ZA | WX-ZB |
| MJ-ZB | 0.1894 | 0.3310 | 0.0252 |
| | | JLHJ-ZB | QA1-ZB |
| PTSJ-ZP | 0.2498 | 0.3765 | 0.1260 |
| | | QHJ-ZA | JYQHJ-ZA |
| QA1-ZB | 0.1327 | 0.2665 | 0.0136 |
| | | JLHJ-ZB | MHJ-ZB |
| QHJ-ZA | 0.2207 | 0.3968 | 0.0082 |
| | | EHP-ZB | TJ(HY)-ZA |
| HCDHP(I)-ZB | 0.1561 | 0.2528 | 0.0762 |
| | | TJ(HY)-ZA | HHJ-ZB |
| WX-ZB | 0.2041 | 0.3594 | 0.0229 |
| | | JLHJ-ZB | DHP-ZB |
| YJ-ZB | 0.1970 | 0.3078 | 0.0776 |
| | | YXGJ-ZB | HBWC-ZB |
| YXGJ-ZB | 0.2270 | 0.3765 | 0.1188 |
| | | MXDHP-ZB | EHP-ZB |
| ZTHJ-ZA | 0.1639 | 0.3765 | 0.0288 |
| | | MXDHP-ZB | TJ(DL)-ZA |
| JYQHJ-ZA | 0.1467 | 0.2625 | 0.0447 |
| | | MXDHP-ZB | TJ(DL)-ZA |

| | | | |
|---------|--------|--------|---------|
| ZA(WP) | 0.1937 | 0.2963 | 0.0555 |
| | | BSJ-ZB | EWJH-ZA |
| Average | 0.1886 | 0.3268 | 0.0550 |

Figures

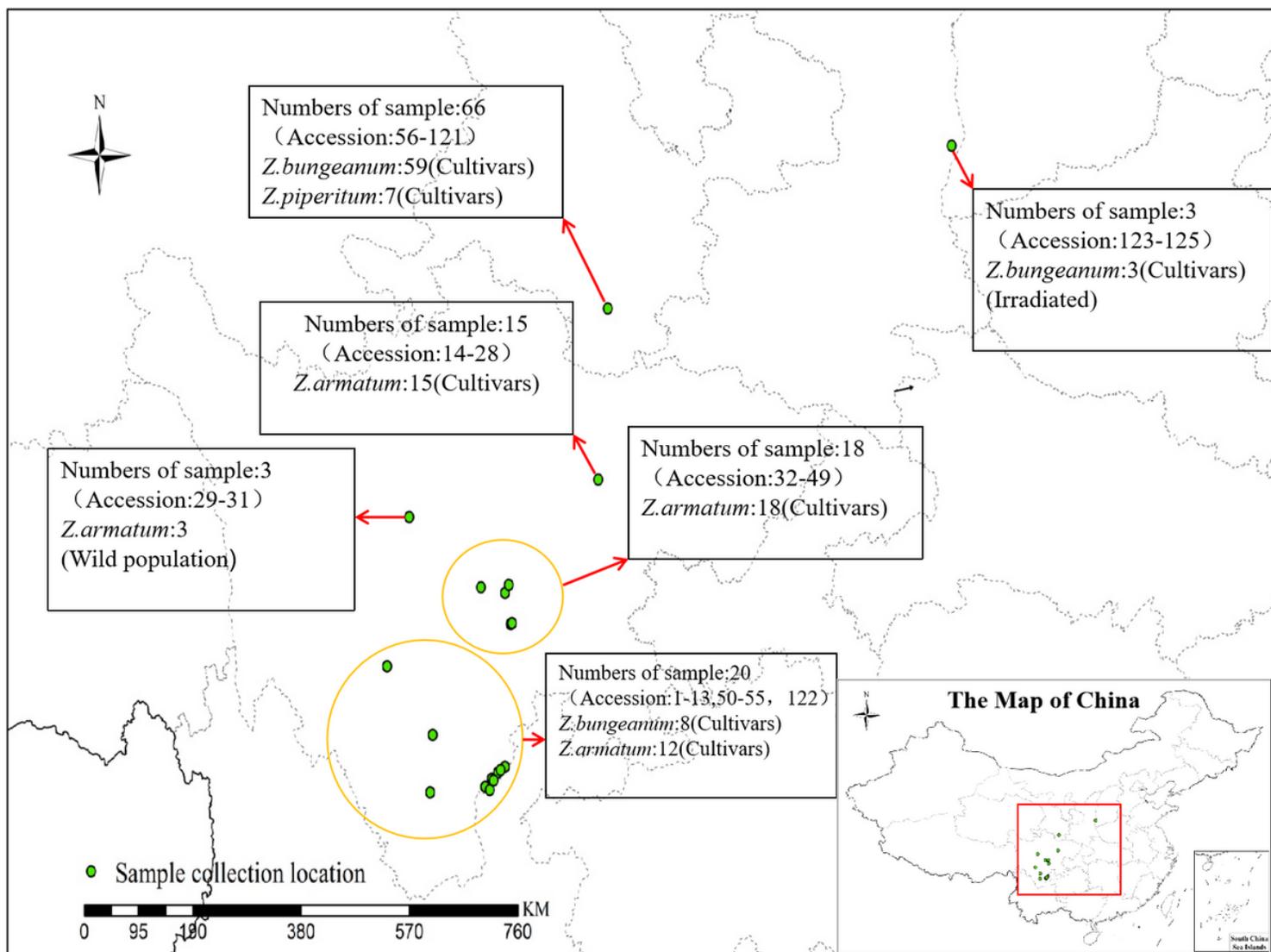


Figure 1

Map showing the location of sites at which 125 *Zanthoxylum* samples were collected. 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.

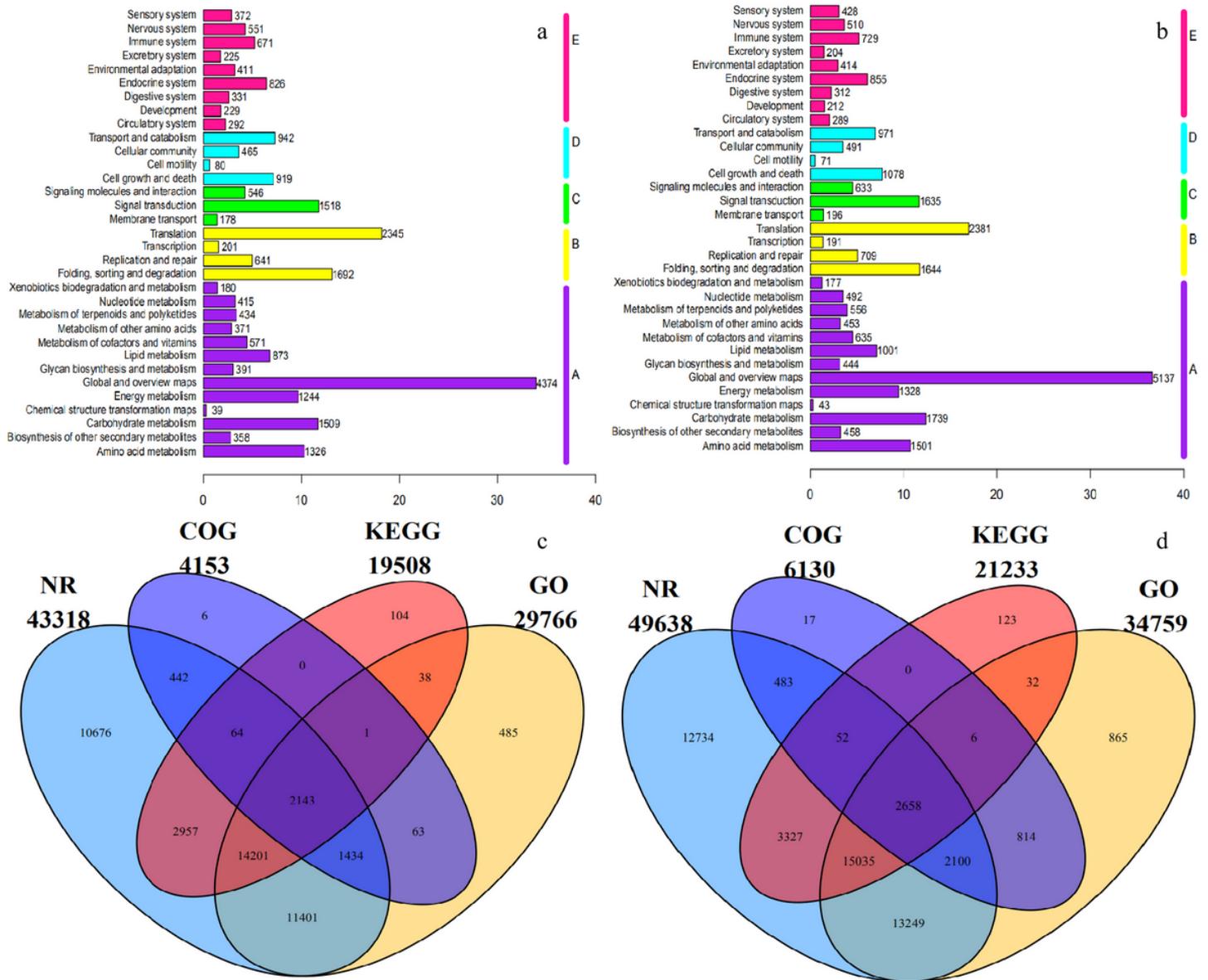


Figure 2

a: KEGG pathway of isogenes for *Zanthoxylum bungeanum*; b: KEGG pathway of isogenes for *Zanthoxylum armatum*; c: Unigenes of *Z. bungeanum* annotated by four databases; d: Unigenes of *Z. armatum* annotated by four databases

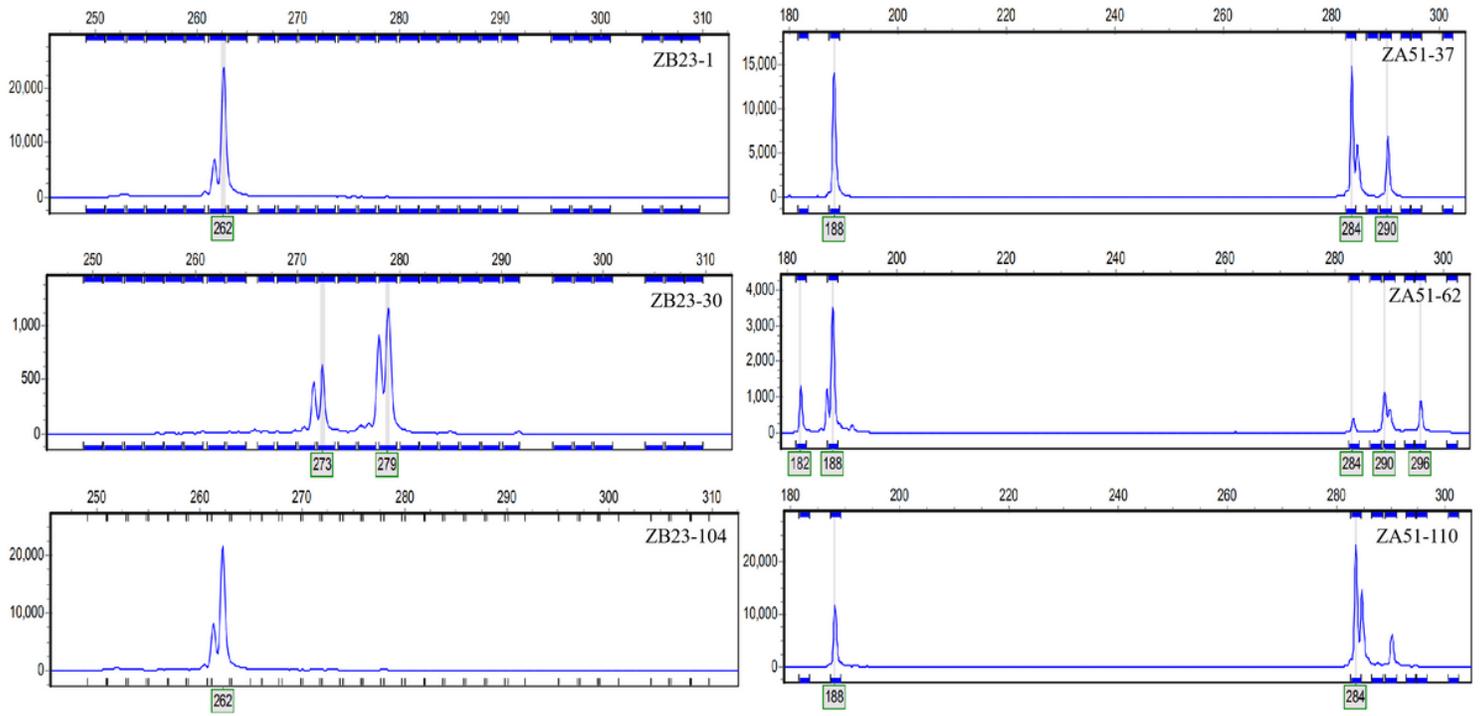


Figure 3

Capillary electrophoresis results of the primer pair ZB23 and ZA51.

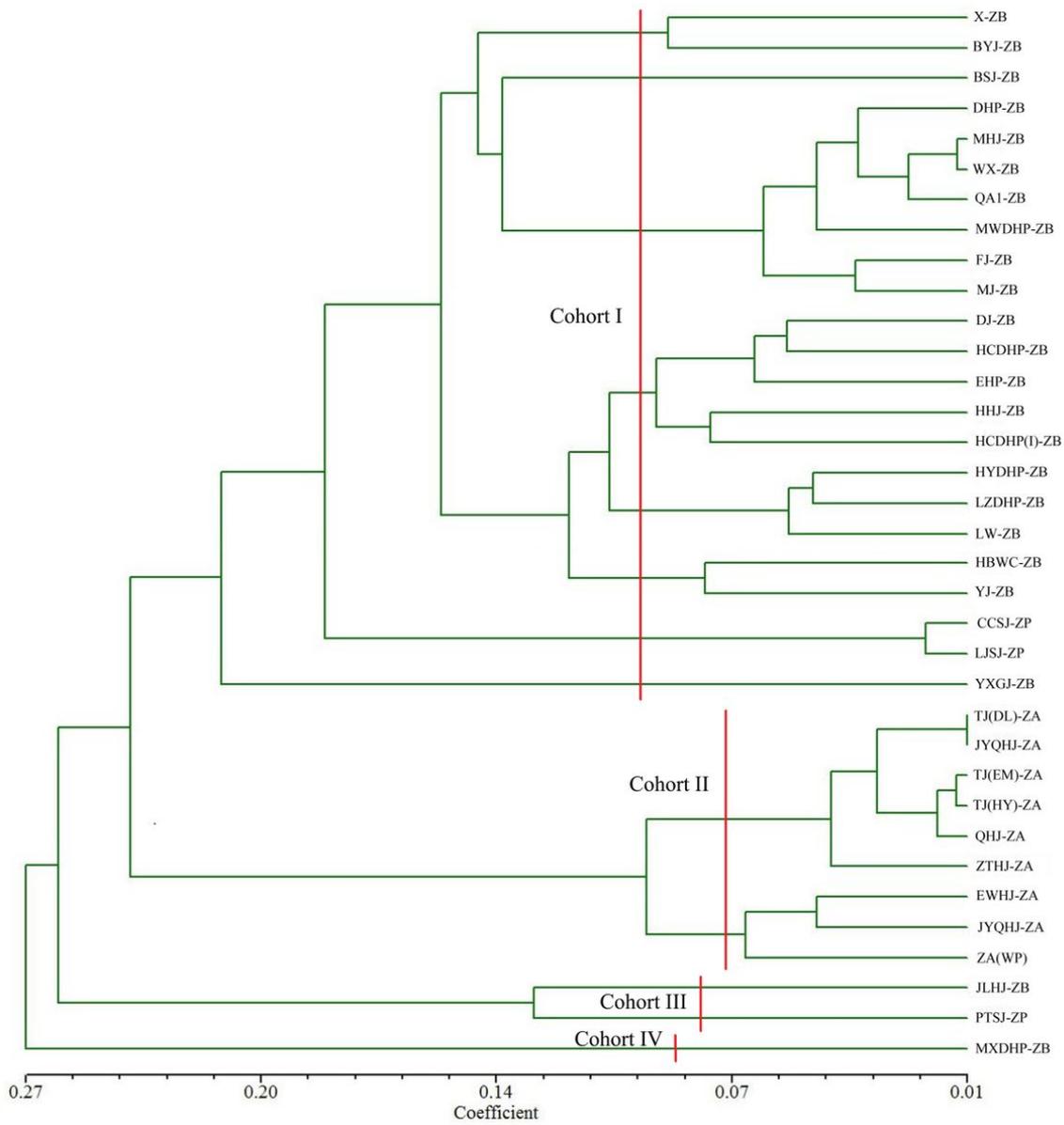


Figure 4

Population clustering dendrogram based on genetic distance (Coefficient number: 0 represents the minimum genetic distance)

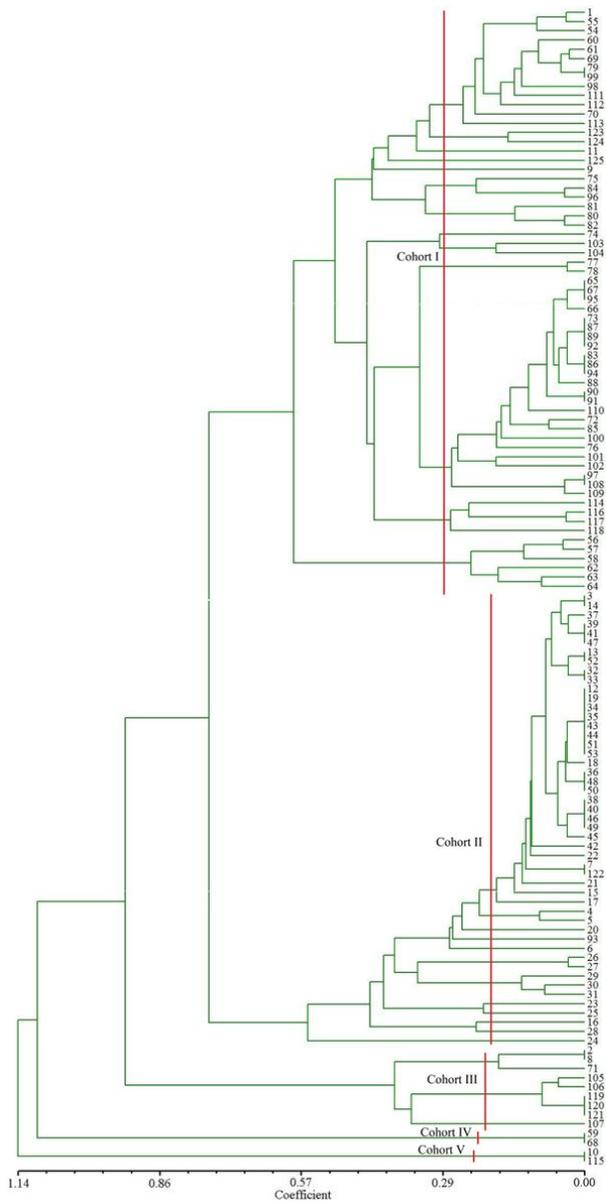


Figure 5

Individual clustering dendrogram based on genetic distance (Coefficient number: 0 represents the minimum genetic distance)

Supplementary Files

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

- [Additionalfile4FigureS1.png](#)
- [Additionalfile3TableS3.doc](#)
- [Additionalfile1TableS1.doc](#)
- [Additionalfile5FigureS2.jpg](#)
- [Additionalfile2TableS2.doc](#)