Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic variation and structure of complete chloroplast genome in alien monoecious and dioecious *Amaranthus* weeds

Han Xu
Chinese Academy of Inspection and Quarantine
Ning Xiang
Chinese Academy of Inspection and Quarantine
Junhua Zhang
Chinese Academy of Inspection and Quarantine
Yongjiang Zhang (zhangyjpvi@yeah.net)
Chinese Academy of Inspection and Quarantine

Research Article

Keywords: Amaranthus, three subgenera, Chloroplast genome, Genetic variation, Hotspots

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-509003/v1

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

Abstract

Amaranthus is a complex taxon with economic importance as well as harmful weeds. We studied the genetic variation and structure of the chloroplast genomes of 22 samples from 17 species of three subgenera. It was found that the length of the chloroplast genome of *Amaranthus* varied from 149,949 bp of *A. polygonoides* to 150,757 bp of *A. albus*. The frequencies of SNPs and InDels in chloroplast genomes were 1.79 % and 2.86 %, and the variation mainly occurred in the non-coding regions. The longest InDel was 387 bp, which occurred on ycf2, followed by 384 bp InDel on psbM-trnD. Two InDels in ndhE-I on the SSC make the three subgenera clearly distinguished. In LSC, SSC and IRs regions, there were four 30 bp forward and reverse repeats, and the repeats in SSC and LSC were in nearly opposite positions in circular genome structure, and almost divided the circular genome into symmetrical structures. In the topological tree constructed by chloroplast genome, species in subgen. *Amaranthus* and subgen. *Acnida* form monophyletic branches separately and cluster together. *A. albus*, *A. blitoides* and *A. polygonoides* were separated from subgen. *Albersia*, and the rest of subgen. *Albersia* were clustered into a monophyletic branch. The rpoC2, ycf1, ndhF-rpl32 were good at distinguishing most amaranths. The trnk-UUU-atpF, trnT-UGU-atpB, psbE-clpP, rpl14-rps19, and ndhF-D can distinguish several similar species. In general, the chloroplast genome is of certain value for the identification of the similar species of *Amaranthus*, which provides more evidence for clarifying the phylogenetic relationships within the genus.

Introduction

The genus of *Amaranthus* includes 74 species, of which 55 species native to the Americas and the rest originated from the Euraisa, South Africa and Austrilia / Oceania[1-2]. The genus contains pseudocereals crops such as *A. caudatus* L., *A. cruentus* L., and *A. hypochondriacus* L., leaf vegetables *A. tricolor* and *A. blitum*, endangered plants *A. pumulis*, and agricultural weeds[3]. The Flora of China (2003) recorded 15 species and two varieties, all of which were alien except for *A. tricolor*[4]. Since then, some authors have successively found new alien amaranths: *A. bouchonii*[5], *A. tenuifolius*[6], *A. palmeri*[7], *A. standelyansus*[8], *A. powellii*[5], *A. dubius*[9] from the collected specimens in China. In port monitoring regions, *A. tuberculatus*, *A. arenicola*, *A. crispus* etc. were newly intercepted and controlled (Xu, unpublished). Among them, *A. arenicola* and *A. crispus* were transient colonization (Xu, unpublished).

According to inflorescences position, the number of perianth segments and urticle dehiscent / indehiscent[2,10], as well as dioecious or monoecious, are divided into three subgenera: *Amaranthus* subgen. *Amaranthus*, *Amaranthus* subgen. *Acnida* (L.) Aellen ex K.R.Robertson, and *Amaranthus* subgen. *Albersia* (Kunth) Gren. & Godr.[10-11]. Of these, 9 species were listed as “introduced, invasive and noxious plants” in the USDA Plants Database (Southern Weed Science Society 1998), and 21 species as“agricultural weeds” in the Global Compendium of Weeds[12]. The genus is the focus of weed scientific research[13], because of these amaranth weeds posed a certain threat to agricultural ecology in the new habitat. *A. palmeri* and *A. tuberculatus* invade gradually into the new continents out of their origins, and were detected their resistant biotypes[14]. Accurate identification of these species is the basis of weed prevention and control. However, the taxonomy of *Amaranthus* has always been difficult, especially because of the large number of complex taxa which are difficult to define due to the interspecific hybridization and gene introgression.

Many authors have studied on the taxonomy and evolution of the genus. The latest taxonomy revision of monoecious species was completed by Bayón (2015)[2], and the comprehensive dioecious taxonomy was Sauer’s monographs[1]. Waselkov et al. (2018) conducted the phylogenetic analyses of 58 species based on three low-copy nuclear genes and two chloroplast regions[15]. Stetter and Schmid (2017) inferred the phylogeny of 35 amaranths using genotyping by sequencing (GBS)[16]. Xu et al. (2020) analyzed ITS, ALS (domain C, A and D) and ALS (domain B and E) and constructed topological trees[14]. In addition to the part incorrect conclusions drawn by Stetter and Schmid (2017) due to missampling, all of them concluded that the classification boundaries between subgenera are not very clear because of some kinds of amaranths, such as *A. palmeri* and *A. spinosus*, *A. arenicola* and subgen. *Amaranthus*, as well as *A. tuberculatus*. And the monophyletic branches of the subgen. *Albersia* have insufficient bootstrap support value.

Additionally, chloroplast genome-related studies include: Chaney et al. (2016) first reported full chloroplast genomes of *A. hypochondriacus*, *A. cruentus*, *A. caudatus* and their hypothetical wild ancestor species *A. hybridus*, and found 210 single nucleotide polymorphisms (SNPs) and 122 insertion/deletion polymorphisms (InDels) compared to the reference chloroplast genome[17]. Viljoen et al. (2018) studied chloroplast genomes and matK, rbcL, ITS in 59 accessions of 9 species of subgen. *Amaranthus* and 4 species of subgen. *Albersia*, and mainly focused on the genetic relationship between wild and domesticated grain amaranths[18]. At present, studies on the chloroplast genome of *Amaranthus* are mainly focused on the grain amaranths, and there is a lack of overall studies on the three subgenera.

In this paper, combined with the problems existing in the classification and evolution of the genus, and the research need of the genetic variation of alien invasive amaranths, we collected alien species of 3 subgenera, and perform complete analysis of the chloroplast genome, in order to further understand the feature of the chloroplast genomes of *Amaranthus*, and the critical regions of chloroplast genomes used to explain the phylogenetic relationship of the genus, especially the evolution of *A. palmeri* and *A. spinosus*, *A. tuberculatus* and *A. arenicola*, *A. spinosus* and *A. dubius*. The results will provide a new basis for the taxonomic revision, phylogenetic evolutionary, weed evolutionary biology and the development of genetic resources.

Results

Genomic features

The quadripartite structure of 22 samples of 17 species in *Amaranthus* consists of a large single-copy region (LSC with 83, 382 – 84, 062 bp), a small single-copy region (SSC with 17, 937 – 18, 124 bp), and a pair of inverted repeat regions (IRs with 23, 964–24, 357 bp). The full length of the 22 cp genomes ranges from 149,949 bp in *A. polygonoides* to 150, 756 bp in *A. albus* (Table 1). The chloroplast genome sequences were deposited in GenBank (Table 1).

The total GC content was 36.5% to 36.6%, only *A. albus*, *A. blitoides* and *A. polygonoides* have a GC content of 36.5% (Table 1). The chloroplast genome contains a total of 133 genes, including 88 protein-coding genes, 37 tRNA genes, and 8 rRNA genes, 18 of which were duplicated in the inverted repeat regions (Table S2). The gene rps12 was trans-spliced; the 50-end exon was located in the LSC region, whereas the 30- intron and exon were duplicated and located in the inverted repeat regions. The partial duplicate of rps19 and ycf1 genes appeared as pseudogenes as they lost their protein-coding ability. 16 genes have introns.

Variants of cp genomes

The length of the SSC region was conserved among the subgenera by comparing the length of the chloroplast genomes of 22 individuals from 17 species. *A. palmeri*, *A. tuberculatus* and *A. arenicola* in subgen. *Acnida* were 18027 - 18042 bp in length (average value 18038.5 ± 5.3151), the SSC length of 5 species of subgen. *Amaranthus* was 17937-17948 bp (average value 17941 ± 3.3665), and the SSC length of 8 species of subgen. *Albersia* was 18057 - 18124 bp (average value 18076.3 ± 22.6806) (Table 1; Figure 1). At SSC, there were about 77 bp InDels in ndhE-G and 180bp InDels in ndhG-I, which induced the variation of SSC length among subgenera (Figure 2). The frequencies of SNPs and InDels in the chloroplast genomes of the 17 species were 1.79% and 2.86%, respectively (Table 2). The frequencies of SNPs and InDels in the genes were 1.22% and 1.14%, and the frequencies of SNPs and InDels in the intergenic spacer were 3.25% and 7.32%, respectively (Table 2). In general, the variation mainly occurred in the intergenic spacer region, and InDels mainly occurred in the non-coding region (Table 2). The longest InDel was 387 bp, which occurred on ycf2, followed by 384 bp InDel on psbM-trnD.

Repeat and SSR analyses

Each species has 28 to 38 repeats, distributed in 30 locations, including 11 to 14 forward repeats, 11 to 17 palindromic repeats, and 6 to 8 reverse repeats ranging from 30 to 64 bp in length. There were 19 common repeats locations, of which 11 had no variation and 8 had variation in length. The R3, R8, R11 and R13 had the most abundant variation (Figure 3). The R12 (forward and reverse repeats) was distributed in LSC, IRa, SSC and IRb. The R12 on SSC is almost opposite to R12 on LSC, dividing the entire circular genome into two parts of nearly equal length. The repeats on LSC were mainly concentrated near Repeat 12 (loci 29572-46282), loci 8166-8327, loci 29572 and loci 75230. The repeats on IRs are constant within the genus. There were two common repeats in SSC, and one was a palindrome sequence shared by subgen. *Acnida*, subgen. *Amaranthus*, and *A. albus*.

MISA analysis showed that each cp genome of *Amaranthus* contained 29-39 SSRs (Table 3). On average, the number of SSR types from more to less was mono-, tetra-, di-, tri-, penta- and hexa-nucleotides in order (Table 3). About 55.56% of those SSRs were composed of A or T bases. Among all SSRs, most loci located in LSC (77.78 %) and IGS (71.91%). About 12 repeat motifs were shared by all species in the genus while the remaining motifs were species-specific or subgenus-specific (Table 3). Different combinations of SSR markers could distinguish all species except *A. standleynaus* and *A. crispus*, *A. dubius* and *A. spinosus* (Table 3).

Phylogenetic trees of whole chloroplast genomes

The results obtained in this study in limited samples were basically consistent with previous studies based on chloroplast gene sequences. *A. palmeri*, *A. arenicola* and *A. tuberculatus* in subgen. *Acnida* clustered together (BS/PP=100/1) (Figure 4). *A. hybridus*, *A. hypochondriacus*, *A. dubius*, *A. spinosus*, *A. retroflexus* clustered together (BS/PP=100/1) (Figure 4). And the above two clades were very close (BS/PP=100/1) (Figure 4). *A. albus* and *A. blitoides* were clustered (BS/PP=35/0.84) and separated from subgen. *Albersia* and were closely related to subgen. *Amaranthus* and subgen. *Acnida* (BS/PP=58/0.99) (Figure 4). *A. polygonoides* become a single basal branch. The rest of subgen. *Albersia* were clustered into one branch (BS/PP=100/1) (Figure 4).

Hotspots for *Amaranthus*

The partially qualified fragment regions searched by exhaustive method were overlapped, and the overlapped regions were combined together as a hotspot region. Finally, 16 hotspot fragments with a length of 737 to 2818 bp were obtained, and the SNP variation frequency ranged from 0.78% to 1.49% (Table S3). The topological trees constructed by the alignments of these 17 hot fragments and the topological trees constructed by the alignment sequences of each gene and intergenic spacer were consistent with the chloroplast genome topological tree, namely, the hotspots with more than 90% bootstrap value support for the subgen. *Amaranthus*, subgen. *Acnida* and subgen. *Albersia* branch (excluding *A. albus*, *A. polygonoides*, and *A. blitoides*) were ndhF-rpl32, ycf1 and rpoC2 (Figure S1).

In several similar taxa, there were 25 InDels and 11 SNPs between *A. tunetanus* and *A. standleyanus*. *A. crispus* and *A. standleyanus* had no difference. There are 46 SNPs and 144 InDels between *A. arenicola* and *A. tuberculatus*. By sequence alignment and variation analysis, it was found that trnK-UUU-atpF, trnT-UGU-atpB, psbE-clpP, rpl14-rps19, ndhF-D could be used to distinguish *A. tunetanus* from *A. standleyanus*, *A. crispus*, and *A. arenicola* from *A. tuberculatus*.

Discussion And Conclusion

In this study, the SSC length of *A. hybridus* and other species in subgen. *Amaranthus* was basically consistent with previous studies on four grain amaranths[17] (Chaney et al., 2016). Chaney et al. (2016) reported that the chloroplast genome of *Amaranthus* contained 111 genes, while Viljoen et al. (2018) reported that *A. tricolor* contained 140 genes [17-18]. Data in both studies showed minor errors and duplications. After repeated data proofreading in this study, 133 genes were confirmed in the chloroplast genome of *Amaranthus*. In addition, due to the inclusion of more amaranths than the former sample, the number of loci polymorphisms found increased to 2735 SNPs and 4363 InDels.

In addition, Chaney et al. (2016) found 29-37 SSRs in four grain amaranths[17]. In our study, 29 to 39 SSRs were identified. After statistical analysis and labeling of the SSRs from each sample, different combinations of SSR markers were found to be able to distinguish the similar species: *A. arenicola* and *A. tuberculatus*, *A. standleyanus* and *A. tunetanus*. In terms of repeats, Chaney et al. (2016) reported 34 to 37 repeats, including 14 to 16 forward repeats and 20 to 21 palindromes[17]. This study found 28 to 38 repeats, 11 to 14 forward repeats, 11 to 17 palindromic repeats, and 6 to 8 reverse repeats. Moreover, the distribution of repeats on the chloroplast genome of *Amaranthus* is found to be regular, such as the distribution of R12. This symmetrical structure should play an important role in the recombination or stabilization of *Amaranthus* chloroplast genes.

The topological tree constructed from the chloroplast genome is basically consistent with the phylogenetic results of Waselkov et al. (2018) using the chloroplast sequence of matK/trnK-UUU and trnL-UAA[15]. Namely, *A. albus*, *A. blitoides*, *A. polygonoides* from subgen. *Albersia* points out to become a separate branch. *A. palmeri* and *A. spinosus* belong to the original subgenus. The chloroplast capture event speculated to occur in Waselkov et al. (2018) was further confirmed. In combination with nuclear gene studies, the relationship between *A. palmeri* and *A. spinosus*[15-16,19], and *A. palmeri* is one of the few species in the dioecious subgenera (*A. watsonii* and *A. arenicola*) that have the characteristics of five perianth segments, suggestting that the hybridization of a species of subgen. *Acnida* and *A. spinosus* in the earlier stage may have led to the chloroplast capture event, which eventually resulted in the formation of *A. palmeri.*

In combination with previous studies, we found that rpoC2, ycf1 and ndhF-rpl32 sequences can be used for phylogenetic and taxonomic identification of *Amaranthus*, according to the principle of similar topological tree branches with the whole chloroplast genome. However, these three sequences cannot effectively distinguish the similar species. In previous studies on the ITS and chloroplast genes matK/trnK-UUU and trnL-UAA of *A. arenicola* and *A. tuberculatus*, the two species were almost indistinguishable[14-15]. In this study, it was found that there was only one SNP site difference in matK/trnK-UUU between *A. arenicola* and *A. tuberculatus*, while their ITS[14, 20] and trnL-UAA sequences showed no difference[20]. In contrast, there are 46 SNPs and 144 InDels between *A. arenicola* and *A. tuberculatus* on chloroplast genomes. The ITS sequences of *A. crispus* and *A. tunetanus* were the same, with only one base difference from *A. standleyanus*[14]. However, there were 25 InDels and 11 SNPs in the chloroplast genomes of *A. tunetanus* and *A. standleyanus*. The five newly discovered regions, trnK-UUU-atpF, trnT-UGU-atpB, psbE-clpP, rpl14-rps19, and ndhF-D, have enough parsimony information sites to distinguish several similar species.

In conclusion, the chloroplast genome is of some significance to the phylogenetic study of *Amaranthus*. However, the study of interspecific and intraspecific gene variation had better be combined with the morphological characteristics of the samples. For species whose morphology is difficult to define, identification errors often occur in samples, and thus the results of molecular analysis are correspondingly wrong. Additionally, the inconsistency of phylogenetic relationships between the chloroplast genome and the nuclear gene sequence of *Amaranthus* may provide new evidence for the evolution and origin of some species.

Materials And Methods

Plant samples, DNA extraction, and sequencing

In this experiment, 21 samples from 16 species of *Amaranthus* and three species as outgroups were used for chloroplast genome analysis (Table S1). All samples were collected from the wild population around the processing plants, wastelands, wharfs in the port supervision area except *A. deflexus*. *A. deflexus* is a common weed collected from the wasteland near the Spanish fields. The specimen collection team was composed of officials of the National Port Weed Monitoring Office of CIQ. The samples collected were approved by the customs and other plant quarantine authorities, and complied with relevant regulations. Habitat and biodiversity were not damaged, and endangered species were not involved. The specimens were deposited at the plant inspection and quarantine institute of Chinese Academy of Inspection and Quarantine (CAIQ) (Beijing, China). All samples were mature plants with flowers and fruits, and identified according to the classification monographs of *Amaranthus* by Sauer (1972)[21], Mosyakin and Robertson (2003)[3], and Bayón (2015)[2]. Total genomic DNA was extracted from the silica-dried leaf tissues using Plant Genomic DNA Kit (Tiangen Biotech Co., China). Genomic DNA of each individual was indexed by a barcode and then pooled together with other samples for sequencing in one lane of HiSeq 2500 (Illumina) (Novogene, Beijing, China).

Genome assembly and annotation

The paired-end sequencing data (2 × 150 bp) were used to assemble its complete chloroplast genome. Sequencing adapters and barcodes were trimmed and low quality reads with Q value ≤ 30 removed. Trimmed paired end reads were mapped to the chloroplast sequence of *A. hypochondriacus* (GenBank accession: MG 836505), with default parameters. The reads were assembled using the Geneious Prime v. 2020.1.2 (Biomatters, Auckland, New Zealand). The consensus chloroplast sequence of *Amaranthus* spp. was retrieved separately and used as a reference for several rounds of mapping of itself reads in order to validate its consensus chloroplast sequence. All trimmed and quality-filtered sequence reads have been deposited in Genbank of NCBI. Non-mapped reads, which are assumed to be of non-plastid origin, were excluded from further analysis. The complete chloroplast genome sequence was annotated using the Geneious Prime v. 2020.1.2 (Biomatters, Auckland, New Zealand) by comparing with the genome of *A. hypochondriacus* (GenBank accession: MG 836505). The assembled and annotated *Amaranthus* spp. chloroplast genome sequence was deposited at NCBI (Table 1).

Genome comparative analysis

A comparative plot of full alignment with annotations of the 22 chloroplast genomes was produced and the nucleotide variability was calculated by Geneious prime v. 2020.1.2 (Biomatters, Auckland, New Zealand) to analyze the total number of mutations. The comparative analysis included the reference sequence *A. hypochondriacus.*

Characterization of repeat sequences and SSRs

We used REPuter[22] to identify the position and size of repeat sequences, which included forward, palindromic, reverse, and complement repeats in the chloroplast genomes of *Amaranthus*. The sequence identity and minimum length of repeat size was set to > 90% and 30 bp. MISA perl script (Thiel et al. 2003) was used to detect the simple sequence repeats (SSRs) in the chloroplast genomes[23]. The thresholds for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs were 10, 5, 4, 3, 3, and 3 repeat units, respectively.

Phylogenetic trees

All phylogenetic analyses were undertaken by the Geneious Prime v. 2020.1.2 software (Biomatters, Auckland, New Zealand), based on the chloroplast genomes of 25 sequences of 20 species (Table S1), including the reference chloroplast genome *A. hypochondriacus*, and three outgroups, *Celosia trigyna* (Genbank Accession: MN057637), *Alternanthera philoxeroides* (Genbank Accession: MK795965) and *Froelichia latifolia* (Genbank Accession: MH286309). The 25 chloroplast genome sequences were aligned using MAFFT[24]. The DNA substitution model (GTR+I+G model) was chosen using jModelTest 2.1.6[25], and used in maximum likelihood (ML) analysis and Bayesian inference. ML analysis was conducted using RAxML version 8.2.11[26] on the Geneious Prime v. 2020.1.2 (Biomatters, Auckland, New Zealand). Bayesian inference was conducted using MrBayes 3.2.6[27] with Ngen=1 000 000, Samplefreq=200, and Burninfrac=0.25.

Search for hotspots

Two methods were used to select suitable regions: (1) search with SNP sites greater than 10 per 1000bp based on exhaustive method by Microsoft Excel 2010; (2) the gene and gene spacer were analyzed one by one manually. Finally, a topological tree was constructed for the searched region and compared with the chloroplast genome topological tree to test the resolution authenticity of this region. The topology tree construction method is consistent with the method in 2.6.

Declarations

Acknowledgements

This work was supported by the Basic Scientific Research Program of CAIQ (2020JK036, 2014JK010□2017JK038), National agricultural standardization Project (NBFW-14-2018). The work involved in this study is part of the above projects on pest identification techniques.

Author Contributions

Han Xu: study design, data interpretation, manuscript writing. Ning Xiang: study design. Junhua Zhang: study design. Yongjiang Zhang: study design, conducting the study.

Competing Interests

The authors declare no competing financial and/or non-financial interests in relation to the work described.

Data Availability

The data generated and analyzed in this study are available from the authors on request.

Ethics declarations

Not applicable.

Collection statement

The samples collected were approved by the customs and other plant quarantine authorities, and complied with relevant regulations. Habitat and biodiversity were not damaged, and endangered species were not involved.

References

1. Sauer, J. D. Revision of the dioecious amaranths. Madroño 13, 5-46 (1955).
2. Bayón, N. D. Revisión taxonómica de las especies monoicas de *Amaranthus* (Amaranthaceae): *Amaranthus* subg. *Albersia* y *Amaranthus* subg. *Amaranthus*. *Ann. Mo. Bot. Gard.* **101**, 261-383 (2015).
3. Mosyakin, S. & Robertson, K. R. *Amaranthus*. Magnoliophyta: Caryophyllidae, part 1. In: Flora of North America Editorial Committee editors. Flora of North America North of Mexico, Vol. 4. Oxford University Press, New York (2003).
4. Published on the Internet http://www.efloras.org [accessed 6 May 2021] Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA (2021).
5. Xu, H. & Li Z. Y. *Amaranthus powellii* Watson and *A. bouchonii* Thell., two newly naturalized species in China. *Guihaia* **39**(10), 1416-1419 (2019).
6. Li, F. Z., Song, B. H. & Lu, Y. Q. Two new records of plant from China. *Guihaia* **22**(1), 7-8 (2002).
7. Li, Z. Y. *Amaranthus palmeri* Watson, a newly naturalized species in China. *Chinese Bulletin of Botany* **20**(6), 734-735 (2003).
8. Li, Z. Y. *Amaranthus standleyanus* Parodi ex Covas, a newly naturalized plant in China. *Bulletin of Botanical Research* **24**(3), 265-266 (2004).

9. Wang, Q. S., Wang, Y., Yan, X. L., Zeng, X. F., Ma, J. S. & Li, H. Q. *Amaranthus dubius* ex Thell., a newly naturalized plants of mainland of China. *Journal of Tropical and Subtropical Botany* **23**(3), 284-288 (2015).
10. Mosyakin, S. L. & Robertson, K. R. New infrageneric taxa and combinations in *Amaranthus* (Amaranthaceae). *Bot. Fenn.* **33**, 275-281 (1996).
11. Costea, M., Sanders, A. & Waines, G. Preliminary results toward a revision of the *Amaranthus hybridus* complex (Amaranthaceae). *Sida* **19**, 931-974 (2001).
12. Randall, R. P. Global Compendium of Weeds. (Available at: http://www.hear.org/gcw/; Accessed on: 6 May, 2021).
13. Tranel, P. J. & Trucco, F. 21st century weed science: A call for *Amaranthus* Pp. 53-81 in Weedy and Invasive Plant Genomics, ed. C. N. Stewart, Jr. Ames: Blackwell. (2009).
14. Xu, H., Pan, X., Wang, C. Chen, Y., Chen, K. & Zhu, S. F. Species identification, phylogenetic analysis and detection of herbicide-resistant biotypes of *Amaranthus* based on ALS and ITS. *Rep.* **10**, 11735 (2020).
15. Waselkov, K. E., Boleda, A. S. & Olsen, K. M. A Phylogeny of the Genus *Amaranthus* (Amaranthaceae) Based on Several Low-Copy Nuclear Loci and Chloroplast Regions. *Bot.* **43,** 439-458 (2018).
16. Stetter, M. G. & Schmid, K. J. Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Phylogenet. Evol.* **109**, 80-92 (2017).
17. Chaney, L., Mangelson, R., Ramaraj, T., Jellen, E. N. & Maughan, P. J. The Complete Chloroplast Genome Sequences for Four *Amaranthus* Species (Amaranthaceae). *APPS* **4**(9), 1600063 (2016).
18. Viljoen, E., Odeny, D. A., Coetzee, M. P. A., Berger, D. K. & Rees, D. J. G. Application of Chloroplast Phylogenomics to Resolve Species Relationships Within the Plant Genus *Amaranthus*. *Mol. Evol.* **86**, 3-4 (2018).
19. Riggins, C. W., Peng, Y. H., Stewart, Jr C. N. & Tranel, P. J. Characterization of de novo transcriptome for waterhemp (*Amaranthus tuberculatus*) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Manag. Sci.* **66**, 1042-1052 (2010).
20. Murphy, B. P. & Tranel, P. J. Identification and Validation of *Amaranthus* Species-Specific SNPs within the ITS Region: Applications in Quantitative Species Identification. *Crop Sci.* **58**, 304-311 (2018).
21. Sauer, J. D. The dioecious amaranths: a new species name and major range extensions. *Madroño* **21**, 426-434 (1972).
22. Kurtz, S. & Schleiermacher, C. REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**, 426-427 (1999).
23. Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (*Hordeum vulgare* L). *Appl. Genet.* **106**, 411-422 (2003).
24. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *NAR* **30**, 3059-3066 (2002).
25. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Methods* **9**, 772 (2012).
26. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
27. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu, L. & Suchard, M. A., Huelsenbeck JP MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Biol.* **61**, 539-542 (2012).

Tables

Table 1 Summary information for the chloroplast genomes of *Amaranthus.*

No.	Species	Comparison of genome length (bp)				GC content	Genbank Accession
		Total	LSC	SSC	IRs		
1	*Amaranthus retroflexus*	150244	83605	17937	24351	36.6%	MN091971
2	*Amaranthus dubius*	150524	83880	17940	24352	36.6%	MN091972
3	*Amaranthus spinosus* 113	150523	83879	17940	24352	36.6%	MT526784
4	*Amaranthus spinosus* 11902	150524	83880	17940	24352	36.6%	MT526783
5	**Amaranthus hypochondriacus*	150523	83878	17941	24352	36.6%	*MG836505
6	*Amaranthus hybridus*	150690	84062	17948	24340	36.6%	MT559305
7	*Amaranthus palmeri*	150731	84010	18027	24347	36.6%	MN091990
8	*Amaranthus arenicola* JSTZ	150632	83901	18039	24346	36.6%	MN091969
9	*Amaranthus arenicola* HBTS	150630	83899	18039	24346	36.6%	MZ152791
10	*Amaranthus tuberculatus* GZW	150679	83945	18042	24346	36.6%	MT559304
11	*Amaranthus tuberculatus* 11994	150695	83961	18042	24346	36.6%	MN091967
12	*Amaranthus tuberculatus* 12194	150696	83962	18042	24346	36.6%	MN091968
13	*Amaranthus blitum*	150621	83806	18057	24379	36.6%	MT526777
14	*Amaranthus crispus*	150567	83793	18060	24357	36.6%	MT526778
15	*Amaranthus standleyanus* 11960	150567	83793	18060	24357	36.6%	MT526781
16	*Amaranthus standleyanus* 7433	150568	83794	18060	24357	36.6%	MT526782
17	*Amaranthus tunetanus*	150581	83805	18062	24357	36.6%	MT526780
18	*Amaranthus deflexus*	150256	83489	18065	24351	36.6%	MT526776
19	*Amaranthus capensis*	150707	83928	18075	24352	36.6%	MT526779
20	*Amaranthus blitoides*	150667	83878	18089	24350	36.5%	MT526786
21	*Amaranthus albus*	150756	83943	18111	24351	36.5%	MT526785
22	*Amaranthus polygonoides*	149948	83896	18124	23964	36.5%	MT472619

Note: *Chloroplast genomic data for *Amaranthus hypochondriacus* were obtained from GenBank.

Table 2 Variation of the chloroplast genomes in *Amaranthus*.

Region	Length (bp)	SNPs		InDels	
		Numbers	Frequency (%)	Numbers	Frequency (%)
Consensus sequence	152519	2735	0.0179	4363	0.0286
Gene	110128	1354	0.0123	1258	0.0114
CDS	80201	1034	0.0129	862	0.0107
tRNA	2780	9	0.0032	0	0
rRNA	9042	6	0.0007	0	0
Intron	18105	305	0.0168	396	0.0219
IGS	42391	1381	0.0326	3105	0.0732

Table 3-1 Distribution of SSRs in the chloroplast genome of *Amaranthus*.

Region			Locus	SSR type	pol	alb	blo	blu	cap	def	cri	sta 2	sta 7433	sta 11960	ret	hyb	hyp	dub	spi	spi 11902	pal	are	are	tub 11994	tub 12
LSC	IGS	trnK-UUU-rps16	4457-4468	(AAGA)3			*																		
			4719-4730	(AT)6		*		*	*	*	*	*	*	*											
			4743-4806	(ATT)4		*	*														*	*	*	*	*
			4762-4782	(AT)8			*											*	*	*					
			4782-4793	(TTA)4													*	*	*	*					
			4795-4806	(ATT)4	*	*	*																		
			4796-4807	(TTA)4											*	*	*	*	*	*					
			4878-4892	(TAT)5												*									
			4880-4894	(TTA)4,5			*	*		*	*	*	*	*	*		*	*	*	*					
			4882-4911	(ATT)5																	*	*	*	*	*
			4913-4924	(TTA)4											*						*	*	*	*	*
			4929-4943	(TAT)5												*									
	Intron	rps16	5798-5809	(T)12					*								*	*	*	*					
	IGS	rps16-trnQ-UUG	6789-6805	(A)13							*	*	*	*											
	IGS	trnQ-UUG-psbK	7257-7269	(T)12		*																			
	IGS	psbK-psbI	7851-7863	(T)12,13			(T)12		(T)13	(T)12								(T)13	(T)13	(T)13	(T)12	(T)12	(T)12	(T)12	(T)12
	IGS	psbI-trnS-GCU	8172-8184	(T)12,13					(T)12		(T)12		(T)12	(T)12										(T)13	(T)13
	tRNA	trnS-GCU	8214-8226	(A)12					*						*			*	*	*					
	IGS	trnS-GCU-trnG-UCC	8654-8666	(A)12,13		(A)13	(A)13			(A)12															
	Intron	trnG-UCC	9466-9484	(T)12					*																
	IGS	atpA-atpF	11720-11734	(T)13,14,15	(T)14	(T)13	(T)15	(T)15	(T)15	(T)13	(T)13	(T)13	(T)14	(T)13											
	IGS	atpF-atpH	13305-13319	(TTTAT)3												*									
			13459-13470	(GGAA)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	CDS	rpoC2	16683-16696	(T)14				*																	
			18568-18580	(T)13			*																		
	Intron	rpoC1	22993-23008	(A)12,13,14,15,16	(A)15	(A)16		(A)16	(A)14	(A)13	(A)13	(A)16	(A)13	(A)13							(A)13	(A)13	(A)13		
			23015-23028	(T)12		*		*		*					*										
	IGS	rpoB-trnC-GCA	28048-28080	(A)12		*					*	*	*	*											
	IGS	petN-psbM	29550-29563	(A)12,14		(A)14			(A)12	(A)12															
	IGS	psbM-trnD-GUC	30072-30084	(A)12				*								*									
			30127-30318	(T)12			*																		
			30358-30371	(A)12	*											*									
			30742-30753	(A)12						*	*	*	*	*								*	*		

Notes: Except for “blo”, which stands for *A. blitoides*, and “blu”, which stands for *A. blitum*, the other species are represented by the first three letters of their species names. “*” indicates that SSR markers at this locus are the same in different species, and blank indicates that SSR markers at this locus are not present in this species.

Table 3-2 Distribution of SSRs in the chloroplast genome of *Amaranthus*.

Region			Locus	SSR type	pol	alb	blo	blu	cap	def	cri	sta 2	sta 7433	sta 11960	ret	hyb	hyp	dub	spi	spi 11902	pal	are	are	t 1
LSC	IGS	trnD-GUC-trnY-GUA	31288-31299	(T)12					*															
	IGS	trnE-UUC-trnT-GGU	32019-32035	(A)16					*															
	IGS	trnT-GGU-psbD	32638-32653	(T)12,14		(T)14	(T)12			(T)12														
			33128-33139	(TCTT)3	*	*	*	*	*	*	*	*	*	*										
	IGS	trnG-GCC-trnfM-CAU	37604-37615	(T)12			*	*			*	*	*	*										
			37622-37641	(TCAAAA)3											*		*	*	*	*				
			37631-37646	(AAAC)3						*														
	IGS	rps4-trnT-UGU	47256-47276	(TA)6,7,9		(TA)6	(TA)6	(TA)6		(TA)6	(TA)6	(TA)6	(TA)6	(TA)6	(TA)7	(TA)7	(TA)9	(TA)9	(TA)9	(TA)9	(TA)6	(TA)6	(TA)6	(
			47597-47608	(A)12											*	*								
			47604-47619	(AT)6,8	(AT)6			(AT)7	(AT)6	(AT)6							(AT)8	(AT)8	(AT)8	(AT)8	(AT)6			
			47619-47632	(TA)6,7	(TA)6	(TA)7					(TA)7	(TA)7	(TA)7	(TA)7										
	IGS	trnT-UGU-trnL-UAA	48127-48138	(A)12	*																			
	Intron	trnL-UAA	48605-48618	(A)12,13,14		(A)13		(A)13	(A)14	(A)14					(A)13	(A)13	(A)12	(A)12	(A)12	(A)12				
	IGS	ndhK-ndhC	51542-51553	(T)12	*	*	*		*	*	*	*	*	*	*		*	*	*	*	*	*	*	*
	IGS	ndhC-trnV-UAC	52311-52322	(TTTC)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*			
	IGS	trnV-UAC-trnM-CAU	53534-53545	(ATCT)3	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	trnM-CAU-atpE	53706-53721	(T)14,15,16	(T)15	(T)15	(T)15	(T)14	(T)15	(T)15	(T)15	(T)15	(T)15	(T)15	(T)15	(T)16	(T)15	(T)15	(T)15	(T)15	(T)15	(T)15	(T)15	(
	IGS	atpB-rbcL	55745-55757	(A)12,13							(A)12	(A)12	(A)12	(A)12			(A)12	(A)12	(A)12	(A)12	(A)13	(A)12	(A)12	(
	IGS	accD-psaI	60133-60157	(T)12,15,16,17													(T)12				(T)15	(T)15	(T)15	(
			60267-60281	(A)15	*																			
	CDS	psaI	60733-60744	(TTTA)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	petL-petG	66751-66770	(T)16,17,18,19,20	(T)20	(T)19	(T)18	(T)16	(T)17	(T)17	(T)18	(T)18	(T)18	(T)18	(T)18	(T)19	(T)20	(T)20	(T)20	(T)20	(T)18	(T)17	(T)17	(
	IGS	rpl33-rps18	68574-68586	(T)13							*	*	*	*										
			68665-68679	(TATTA)3												*								
			68678-68689	(TA)6							*	*	*	*										
	IGS	rpl20-rps12	69940-69953	(T)14	*																			
			70437-70450	(T)13												*								

Notes: Except for “blo”, which stands for *A. blitoides*, and “blu”, which stands for *A. blitum*, the other species are represented by the first three letters of their species names. “*” indicates that SSR markers at this locus are the same in different species, and blank indicates that SSR markers at this locus are not present in this species.

Table 3-3 Distribution of SSRs in the chloroplast genome of *Amaranthus*.

Region			Locus	SSR type	pol	alb	blo	blu	cap	def	cri	sta 2	sta 7433	sta 11960	ret	hyb	hyp	dub	spi	spi 11902	pal	are	are
LSC	Intron	clpP Intron	71472-71483	(AAAT)3											*								
			71537-71551	(AAAAT)3											*								
			72068-72079	(T)12		*																	
			72227-72244	(A)12,14,15,16	(A)15	(A)16	(A)16	(A)12	(A)16	(A)15	(A)14	(A)14	(A)14	(A)14			(A)12	(A)12	(A)12	(A)12	(A)12		
	IGS	psbH-petB	75854-75865	(TTTC)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	Intron	petD	77815-77828	(AT)6,7																		(AT)6	(AT)6
	CDS	rpoA	78957-78969	(T)13	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	rps11-rpl36	80525-80536	(T)12											*	*	*	*	*	*		*	*
	Intron	rpl16	82886-82910	(T)12,13,14,15			(T)15			(T)12											(T)14	(T)14	(T)14
			82950-82971	(A)12,13,14			(A)13	(A)13	(A)14						(A)12								
	CDS	rpl22	85134-85147	(T)14	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*
IRa	IGS	rps19-rpl2	85622-85633	(T)12			*																
	CDS	ycf2	90647-90658	(CTT)4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	rRNA	rrn23	105401-105412	(AGGT)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	rrn4.5-rrn5	106884-106895	(CCCT)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SSC	IGS	ndhF-rpl32	112060-112074	(T)13			*																
			112936-112948	(AAT)4						*													
	IGS	rpl32-trnL-UAG	113619-113632	(TA)6,7							(TA)6	(TA)7	(TA)6	(TA)6							(TA)6		
			113867-113882	(T)16													*	*	*	*			
			114113-114125	(T)13																		*	*
			114129-114150	(A)13,14,15						(A)13					(A)14	(A)14	(A)13	(A)13	(A)13	(A)13	(A)13	(A)15	(A)15
	IGS	trnL-UAG-ccsA	114635-114650	(T)12,13,14,15,16											(T)12	(T)13	(T)15	(T)14	(T)14	(T)14	(T)14	(T)16	(T)16
	CDS	ndhD	116051-116062	(AATA)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	psaC-ndhE	117765-117782	(TCTAGT)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	CDS	ndhE	118073-118095	(TATT)3		*	*								*	*	*	*	*	*	*	*	*
	Intron	ndhA	121028-121039	(T)12				*															
	IGS	rps15-ycf1	127822-127833	(TCTT)3					*														
IRb	IGS	rrn4.5-rrn5	131038-131049	(AGGG)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	rRNA	rrn23	132517-132528	(CTAC)3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	CDS	ycf2	147273-147284	(AAG)4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	IGS	rpl2-rps19	152298-152309	(A)12			*																

Notes: Except for “blo”, which stands for *A. blitoides*, and “blu”, which stands for *A. blitum*, the other species are represented by the first three letters of their species names. “*” indicates that SSR markers at this locus are the same in different species, and blank indicates that SSR markers at this locus are not present in this species.

Figures

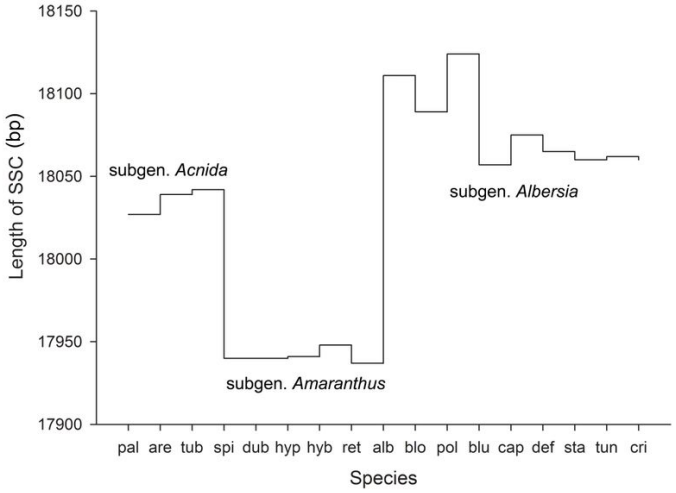

Figure 1

Differences in SSC length of 17 sThe distribution of repeat sequences at 30 loci in Amaranthus. “R” is short for repeat. The red line segment R12 and the black line segments are repeats in all 17 species, the orange line segment represents a repeating sequence in some species. A repeat with only one line segment indicates that there is only one repeat at the site, and vice versa indicates that there are several different repeats at the site.pecies from three subgenera of Amaranthus.

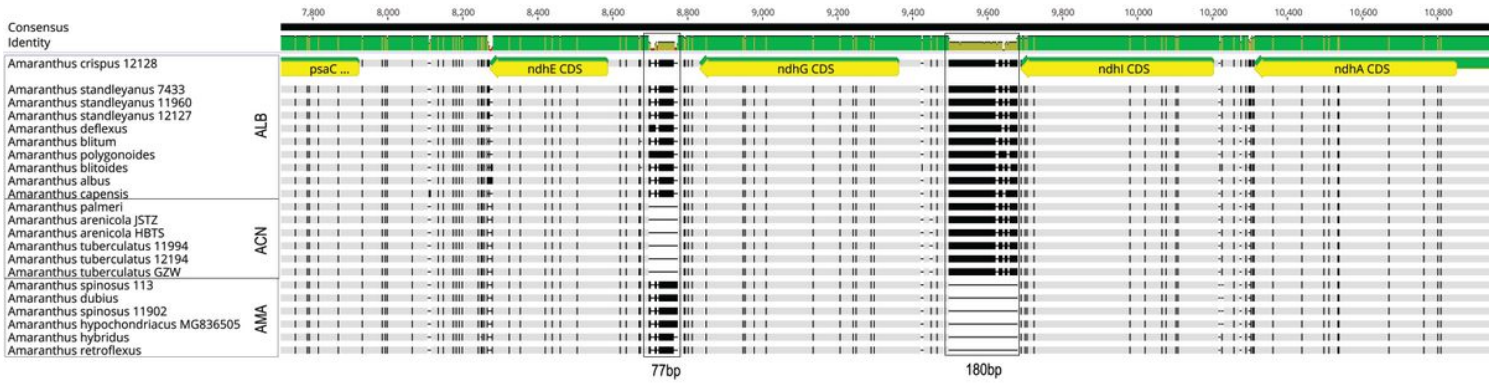

Figure 2

The main insertion and deletion regions on SSC of Amaranthus were ndhE-G (77bp) and ndhG-I (180bp). “AMA” represented the subgen. Amaranthus, “ACD” represented the subgen. Acnida, “ALB” represented the subgen. Albersia.

Figure 3

The distribution of repeat sequences at 30 loci in Amaranthus. “R” is short for repeat. The red line segment R12 and the black line segments are repeats in all 17 species, the orange line segment represents a repeating sequence in some species. A repeat with only one line segment indicates that there is only one repeat at the site, and vice versa indicates that there are several different repeats at the site.

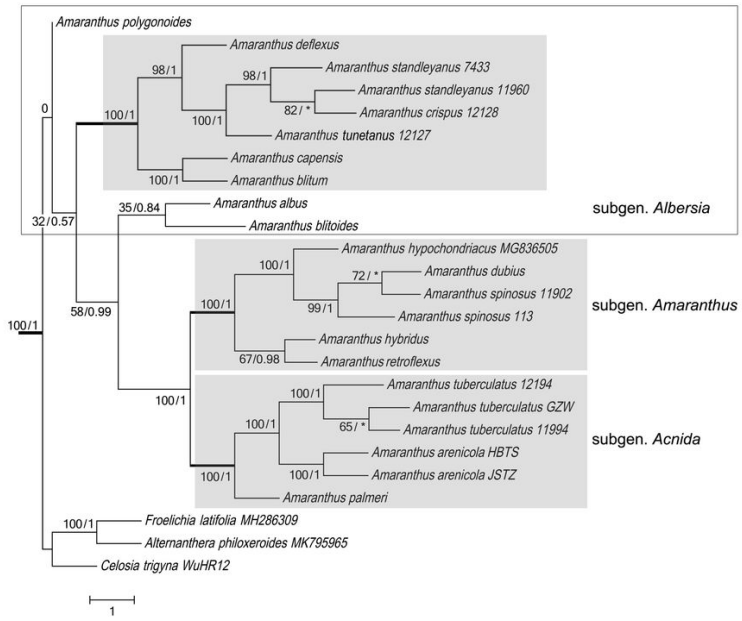

Figure 4

A maximum likelihood topological tree based on chloroplast genome of Amaranthus and three outgroups. Values at each node indicate maximum likelihood bootstrap support (BS) / Bayesian inference posterior probability (PP) value. Individuals marked with grey backgrounds represent major monophyletic branches in the genus.

Supplementary Files

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

- FigureS1ndhFycf1rpoC2.jpg
- SupplementaryinformationTableS14.xlsx