

The Complete Chloroplast Genome Sequences for Four Amaranthus Species (Amaranthaceae)

Authors: Chaney, Lindsay, Mangelson, Ryan, Ramaraj, Thiruvarangan, Jellen, Eric N., and Maughan, Peter J.

Source: Applications in Plant Sciences, 4(9)

Published By: Botanical Society of America

URL: <https://doi.org/10.3732/apps.1600063>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

THE COMPLETE CHLOROPLAST GENOME SEQUENCES FOR FOUR *AMARANTHUS* SPECIES (AMARANTHACEAE)¹

LINDSAY CHANEY^{2,4}, RYAN MANGELSON², THIRUVARANGAN RAMARAJ³, ERIC N. JELLEN²,
AND PETER J. MAUGHAN^{2,5}

²Department of Plant and Wildlife Sciences, Brigham Young University, 5144 Life Sciences Building, Provo, Utah 84602 USA;
and ³The National Center for Genomic Resources, Santa Fe, New Mexico 87505 USA

- **Premise of the study:** The amaranth genus contains many important grain and weedy species. We further our understanding of the genus through the development of a complete reference chloroplast genome.
- **Methods and Results:** A high-quality *Amaranthus hypochondriacus* (Amaranthaceae) chloroplast genome assembly was developed using long-read technology. This reference genome was used to reconstruct the chloroplast genomes for two closely related grain species (*A. cruentus* and *A. caudatus*) and their putative progenitor (*A. hybridus*). The reference genome was 150,518 bp and possesses a circular structure of two inverted repeats (24,352 bp) separated by small (17,941 bp) and large (83,873 bp) single-copy regions; it encodes 111 genes, 72 for proteins. Relative to the reference chloroplast genome, an average of 210 single-nucleotide polymorphisms (SNPs) and 122 insertion/deletion polymorphisms (indels) were identified across the analyzed genomes.
- **Conclusions:** This reference chloroplast genome, along with the reported simple sequence repeats, SNPs, and indels, is an invaluable genetic resource for studying the phylogeny and genetic diversity within the amaranth genus.

Key words: Amaranthaceae; *Amaranthus hypochondriacus*; chloroplast genome; grain amaranths; next-generation sequencing; pigweed.

The genus *Amaranthus* L. (Caryophyllales: Amaranthaceae) has a worldwide distribution with approximately 70 herbaceous plant species that range from aggressive weeds (pigweeds) to cultivated ornamentals, vegetables, and nutritious pseudocereals. The grain amaranths (*A. hypochondriacus* L., *A. caudatus* L., and *A. cruentus* L.) are ancient crops native to the New World that were likely domesticated multiple times from putative progenitor species (*A. hybridus* L., *A. quitensis* Kunth, and *A. powellii* S. Watson; Mallory et al., 2008; Maughan et al., 2011). Amaranths represent a regionally important staple food crop that also played an important cultural role in pre-Columbian civilizations; indeed, the Spaniards recorded that the Aztec emperor Moctezuma II required an annual tribute of amaranth nearly equal to the annual tribute of maize (Sauer, 1967). After the Spanish conquest of the New World, the arriving Europeans actively discouraged the cultivation of the grain amaranth due to its deeply rooted use in indigenous religious ceremonies (Iturbide and Gispert, 1994). Recently, however, the grain amaranths have

acquired renewed attention due to the nutritional quality of their seed, which is high in crude protein (reaching 22.5% dry matter basis), fiber (8%), and the essential amino acid lysine (0.73–0.84%) (Bressani et al., 1992).

The primary function of chloroplasts in plants is photosynthetic carbon fixation. Chloroplasts are thought to have originated from a free-living cyanobacteria-like ancestor through an ancient endosymbiotic relationship (Raven and Allen, 2003). The chloroplast genome has evolved into a circular molecule with a quadripartite structure consisting of a small single copy (SSC), a large single copy (LSC), and two copies of inverted repeat (IR) regions. In addition to this conserved quadripartite genome structure, chloroplast genomes have highly conserved gene content and a nearly collinear gene order among most land plants (Jansen et al., 2005). Because of their compact size, lack of recombination, and maternal inheritance (Birky, 2001), chloroplast genomes have served as important DNA sequences for plant identification and the reconstruction of plant phylogenies (Parks et al., 2009; Moore et al., 2010).

In this study, we report a high-quality chloroplast genome assembly of an agronomically important grain amaranth cultivar (*A. hypochondriacus* ‘Plainsman’), which we use as a reference sequence to reconstruct the chloroplast genome of four additional *A. hypochondriacus* accessions, as well as one accession each of *A. caudatus*, *A. cruentus*, and *A. hybridus*. We report the genome size, structure, and gene content of the chloroplast genome for each of these accessions and mine their genomes for simple sequence repeats (SSRs), repetitive sequences, single-nucleotide polymorphisms (SNPs), and insertion/deletion polymorphisms (indels). This is the first report of a chloroplast genome published in the genus *Amaranthus* and will serve as an important

¹Manuscript received 27 May 2016; revision accepted 27 July 2016.

This research was supported by a grant from the Office of Research and Creative Activities at Brigham Young University. The authors gratefully acknowledge D. Brenner (U.S. Department of Agriculture National Plant Germplasm System [USDA-NPGS], Iowa State University, Ames, Iowa, USA) and D. Baltensperger (Texas A&M University, College Station, Texas, USA) for providing seed samples.

⁴Current address: Division of Natural Science and Mathematics, Snow College, 150 E. College Avenue, Science Building 205, Ephraim, Utah 84627 USA

⁵Author for correspondence: jeff_maughan@byu.edu

doi:10.3732/apps.1600063

Applications in Plant Sciences 2016 4(9): 1600063; <http://www.bioone.org/loi/apps> © 2016 Chaney et al. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).

reference resource for future phylogenetic studies in the Amaranthaceae family (Schmitz-Linneweber et al., 2001; Li et al., 2014).

METHODS AND RESULTS

Seeds representing all three grain amaranth species and their putative progenitor species (*A. hybridus*) were obtained from the publicly available U.S. Department of Agriculture National Plant Germplasm System (USDA-NPGS) *Amaranthus* germplasm collection (<http://www.ars-grin.gov/>; Table 1). All plant materials were grown in the Life Sciences greenhouses at Brigham Young University (Provo, Utah, USA) in 12-cm pots with one plant per pot using Sunshine Mix II (Sun Gro Horticulture, Agawam, Massachusetts, USA).

To develop a high-quality reference chloroplast genome for *Amaranthus*, high-molecular-weight DNA from the cultivar *A. hypochondriacus* ‘Plainsman’ (PI 558499) was extracted from fresh leaf tissue by Amplicon Express (Pullman, Washington, USA). Whole-genome DNA sequencing was performed using PacBio RS II (Pacific Biosciences, Menlo Park, California, USA) long-read technology at the National Center for Genome Resources (Santa Fe, New Mexico, USA). A total of four libraries were constructed using the PacBio 20-kbp protocol (Kim et al., 2014). These libraries were loaded onto 20 single-molecule real-time sequencing (SMRT) cells and sequenced using P6 polymerase and C4 chemistry with 4-h movie times. Sequencing yielded 2,359,408 filtered subreads (13,578,880,181 bp) with a mean read length of 5706 bp. These reads were assembled using FALCON version 0.4 (<https://github.com/PacificBiosciences/FALCON>), which produced 9494 contigs, of which five showed high BLASTN (*E*-value = 0) homology and spanned nearly the entire length of chloroplast genome from other plant species within the family Amaranthaceae, specifically *Spinacia oleracea* L. (NC_002202.1) and *Beta vulgaris* L. subsp. *vulgaris* (KR230391.1). These contigs were assembled manually relative to the chloroplast sequence of *S. oleracea*, with putative gaps filled with Ns. Using default parameters, the PacBio-filtered subreads were then mapped back to the draft amaranth chloroplast and locally realigned using the map reads to reference and local realignment tools found in the NGS Core Tools of the CLC Genomic Workbench (version 8.0.3; QIAGEN, Valencia, California, USA). A total of 20,042 (0.8%) reads were aligned to the draft genome, from which a consensus sequence for the complete amaranth chloroplast genome was extracted. This reference sequence was validated by resequencing 14 chloroplast gene sequences (Table 2) using amplicon sequencing and by mapping Illumina short reads (see below) from ‘Plainsman’ to the PacBio reference assembly. Both the short-read mapping consensus as well as the amplicon resequencing showed 100% identity to the PacBio-assembled reference genome. The complete *A. hypochondriacus* reference chloroplast genome has been deposited into GenBank (accession no. KX279888).

To study genetic diversity, total genomic DNA for eight *Amaranthus* accessions, including ‘Plainsman’ and four additional *A. hypochondriacus* accessions, and one accession each of *A. caudatus*, *A. cruentus*, and *A. hybridus* (Table 1), was extracted from 30 mg of freeze-dried leaf tissue according to the protocol reported by Todd and Vodkin (1996). Mate-pair libraries with insert sizes of 800 bp were sequenced on the Illumina HiSeq platform (Illumina, San Diego, California, USA) to generate 2 × 100 bp reads (Beijing Genomic Institute, Hong Kong, China). The generated reads were quality trimmed and six million paired reads were randomly subsampled from each accession and mapped to the ‘Plainsman’ reference chloroplast genome followed by local realignment using the NGS core toolbox in the CLC Genomics Workbench with default parameters. Consensus sequences for each accession are provided in Appendix S1. Across all accessions, the mean depth of coverage and percentage of total reads that mapped to the ‘Plainsman’ reference was 254× and 6.6%, respectively (Table 1).

The consensus chloroplast genomes for each accession were annotated using CpGAVAS (Liu et al., 2012) using default settings with *S. oleracea* (Caryophyllales: Amaranthaceae) set as the reference species (Schmitz-Linneweber et al., 2001). Any problematic annotations were corrected using the Apollo genome editor (Lewis et al., 2002). All eight chloroplast genomes were composed of a single circular double-stranded DNA molecule with the typical quadripartite structure (Fig. 1). Not unexpectedly, all of the chloroplast genomes are similar in length to the ‘Plainsman’ reference genome (150,518 bp, with a 60-bp range among all accessions; Table 1). This is similar to other chloroplast genomes in the Amaranthaceae family (*S. oleracea*: 150,725 and *B. vulgaris* subsp. *vulgaris*: 149,637; Schmitz-Linneweber et al., 2001; Li et al., 2014). The ‘Plainsman’ reference chloroplast genome consisted of an 83,873-bp LSC (69-bp range among all accessions), a 17,941-bp SSC (16-bp range among all accessions), and a pair of 24,352-bp IRs (51-bp range among all accessions). The chloroplast genome had

TABLE 1. Information on the sequenced amaranth chloroplast genomes and genome features.

| Accession no. | Species | Collection origin ^a | Genome size (bp) | Coverage depth (×) | % of reads mapping | LSC length (bp) | SSC length (bp) | IR length (bp) | GC content (%) |
|------------------------|------------------------------|--------------------------------|------------------|--------------------|--------------------|-----------------|-----------------|----------------|----------------|
| PI 558499 ^b | <i>A. hypochondriacus</i> L. | Nebraska, USA | 150,518 | 146 | 3.8 | 83,873 | 17,941 | 24,352 | 36.61 |
| PI 477913 | <i>A. cruentus</i> L. | Mexico | 150,555 | 302 | 7.9 | 83,908 | 17,945 | 24,351 | 36.60 |
| PI 481125 ^c | <i>A. hypochondriacus</i> | India | 150,556 | 283 | 7.4 | 83,908 | 17,946 | 24,361 | 36.60 |
| PI 5111731 | <i>A. hypochondriacus</i> | Mexico | 150,555 | 166 | 4.4 | 83,906 | 17,947 | 24,351 | 36.60 |
| PI 540446 | <i>A. hypochondriacus</i> | Pakistan | 150,499 | 208 | 5.5 | 83,850 | 17,947 | 24,351 | 36.61 |
| PI 605351 | <i>A. hybridus</i> L. | Greece | 150,552 | 239 | 6.2 | 83,903 | 17,947 | 24,351 | 36.60 |
| PI 619259 | <i>A. hypochondriacus</i> | Nepal | 150,559 | 194 | 5.1 | 83,910 | 17,947 | 24,391 | 36.60 |
| PI 642741 | <i>A. caudatus</i> L. | Bolivia | 150,556 | 495 | 12.9 | 83,919 | 17,957 | 24,340 | 36.60 |

Note: IR = inverted repeat; LSC = large single copy; SSC = small single copy.

^aAll origin information is derived from the Germplasm Resources Information Network (<http://www.ars-grin.gov/npgs/>). Several accessions were collected in the Old World, although all originated in the Americas according to Sauer (1967).

^b‘Plainsman’.

^cClassified as *A. hypochondriacus* based on Maughan et al. (2011).

TABLE 2. Genes contained in the sequenced amaranth chloroplast genomes.

| Category | Gene group | Gene names |
|------------------|------------------------------------|---|
| Photosynthesis | Subunits of ATP synthase | <i>atpA, atpB, atpE, atpF^a, atpH, atpI</i> |
| | Subunits of NADH-dehydrogenase | <i>ndhA^{a,b}, ndhB^{a,b} (×2), ndhC, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK^b</i> |
| | Subunits of cytochrome b/f complex | <i>petA, petG, petL, petN^b</i> |
| | Subunits of photosystem I | <i>psaA, psab, psac, psal, psaj</i> |
| | Subunits of photosystem II | <i>psbA, psbB, psbC^b, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ</i> |
| Self-replication | Subunit of RuBisCO | <i>rbcL</i> |
| | Large subunit of ribosome | <i>rpl2 (×2), rpl4, rpl6^b, rp20, rpl22, rpl32, rpl33</i> |
| | DNA-dependent RNA polymerase | <i>rpoA, rpoB, rpoC1^{a,b}, rpoC2^b</i> |
| | Small subunit of ribosome | <i>rps2, rps3, rps4, rps7 (×2), rsp8, rps11, rps12^b, rps14, rps15, rps16^a, rsp18</i> |
| | rRNA genes | <i>rrn4.5S (×2), rrn5S (×2), rrn16S (×2), rrn23S (×2)</i> |
| | tRNA genes | <i>trnA-TGC^a (×2), trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-CAT, trnG-GCC, trnH-GTG, trnI-CAT (×2), trnL-CAA (×2), trnL-TAG, trnM-CAT, trnN-GTT (×2), trnP-TGG, trnQ-TTG, trnR-ACG (×2), trnR-TCT, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT, trnV-GAC (×2), trnW-CCA, trnY-GTA</i> |
| Other genes | Subunit of acetyl-CoA-carboxylase | <i>accD</i> |
| | c-type cytochrome synthesis gene | <i>ccsA</i> |
| | Envelope membrane protein | <i>cemA</i> |
| | Protease | <i>clpP^{a,b}</i> |
| | Translational initiation factor | <i>infA</i> |
| Unknown function | Maturase | <i>matK</i> |
| | Conserved open reading frames | <i>ycf4</i> |

^aGenes containing one or more introns.

^bGenes used to validated the reference assembly using amplicon resequencing.

a GC content of 37.6% for all eight accessions (Table 1) and consisted of a total of 111 genes, including 72 protein-coding genes, 31 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes (Table 2, Appendix S2). Other published chloroplast genomes in the Amaranthaceae family have similar gene counts: *S. oleracea* is reported to have at least 108 genes and *B. vulgaris* subsp. *vulgaris* is reported to have 113 genes (Schmitz-Linneweber et al., 2001; Li et al., 2014). There were 13 genes that were repeated in the IR region, one is involved in photosynthesis and the others have self-replication functions (Table 2). There were seven genes with introns, one of which has two introns (Table 2). In the reference genome, 111 genes constituted a total of 72,598 bp and were composed of 18,299 codons, excluding the stop codons (Appendix S2). Similar to other land plant chloroplast genomes, the most prevalent amino acid was leucine (10.6%) and the least common was cysteine (1.1%) (Huang et al., 2013, 2014; Qian et al., 2013), with a strong bias toward an A or T at the third codon position (Clegg et al., 1994).

Microsatellites, or simple sequence repeats (SSRs), were identified using MISA (Thiel et al., 2003). The SSRs usually have a higher mutation rate, are easily genotyped using PCR, and are frequently used as markers for phylogenetic analysis and in marker-assisted breeding when located on nuclear chromosomes. We followed standard chloroplast thresholds for identifying SSRs (Sablok et al., 2015), specifically a minimum stretch of 12 for mono-, six for di-, four for tri-, and three for tetra-, penta-, and hexanucleotide repeats and a minimum distance of 100 bp between compound SSRs. We identified 37 SSRs in the reference genome, and 29 to 33 in the other sequenced *Amaranthus* chloroplast genomes (Table 3). The SSRs were most commonly found in the LSC region of the chloroplast genomes (60–63% of SSRs). The most abundant motifs were the runs of the mononucleotide A or T. Analysis of the consensus chloroplast genome sequences of the eight sequenced *Amaranthus* accessions identified repeat number polymorphisms for five, two, and one of the mono-, di-, and tetranucleotide SSRs, respectively. The complete list of the SSRs identified and their positions in the amaranth chloroplast genomes is provided in Appendix S3.

Repetitive sequences in chloroplast genomes are considered to play an important role in the rearrangement and stabilization of chloroplast genomes (Milligan et al., 1989; Maréchal et al., 2009). Repeated elements in the *Amaranthus* chloroplast genomes were identified using REPuter (Kurtz et al., 2001). A minimum repeat size of 30 bp and a sequence identity of ≥90% were used to identify forward and palindrome repeats (no reverse or complement repeats were present). Only a single IR region was included in the analysis to avoid inherent redundancy. The number and size of repeats were conserved between amaranth chloroplast genomes (Table 4). A total of 34 repeats were found in the *A. hypochondriacus* ‘Plainsman’ chloroplast genome, including 14 forward repeats and 20 palindromic repeats ranging from 30 to 64 bp in length (Appendix S4).

SNPs and indels were identified from binary alignment map (BAM) files generated by aligning reads to the ‘Plainsman’ reference genome using

BWA-MEM (Li, 2013) for each of the seven resequenced amaranth accessions. SNPs were identified using SAMtools mpileup (Li et al., 2009), and indels were identified using the Genome Analysis Tool Kit HaplotypeCaller (GATK-HC) (McKenna et al., 2010). Relative to the ‘Plainsman’ reference chloroplast genome, an average of 210 SNPs and 122 indels were identified across the accessions (Table 5; Appendix S5, S6). Indels ranged in size from 2 bp to 75 bp. T/G base substitutions accounted for the highest percentage of all SNPs (18.6%), while G/C base substitutions were the least frequent (2.9%) (Appendix S5).

CONCLUSIONS

We report the assembly and annotation of the first reference-quality chloroplast genome for the genus *Amaranthus*. Using this reference genome, we reconstructed the chloroplast genome for representatives of the grain amaranths—*A. hypochondriacus*, *A. cruentus*, and *A. caudatus*—and their putative wild ancestor, *A. hybridus*. The amaranth chloroplast genome retains the quadripartite structure so highly conserved among land plants and is highly conserved at the nucleotide level among the grain amaranths. Nonetheless, several SNPs, indels, and polymorphic SSRs were identified that will serve as invaluable genetic markers for the future genetic improvements of these emerging crop species. The development of a reference chloroplast genome for *Amaranthus* will also be invaluable for broader comparative studies within Caryophyllales—an order of increasing importance as it contains many halophytic and drought-tolerant species. The availability of a reference chloroplast genome will also open the door to biotechnology applications within the genus such as chloroplast gene transformation. We note that the genus contains not only important grain species, but also several well-known weedy species (e.g., *A. retroflexus* L., *A. viridis* L., *A. tuberculatus* (Moq.) J. D. Sauer, and *A. spinosus* L.) that are collectively referred to as “pigweeds” and are among the most economically damaging weeds in the United States (Tranel and Trucco, 2009). This reference chloroplast genome will provide an important framework for the accurate assembly of chloroplast genomes of these weedy species, which will be of particular value for monitoring spread and interspecies hybridizations.

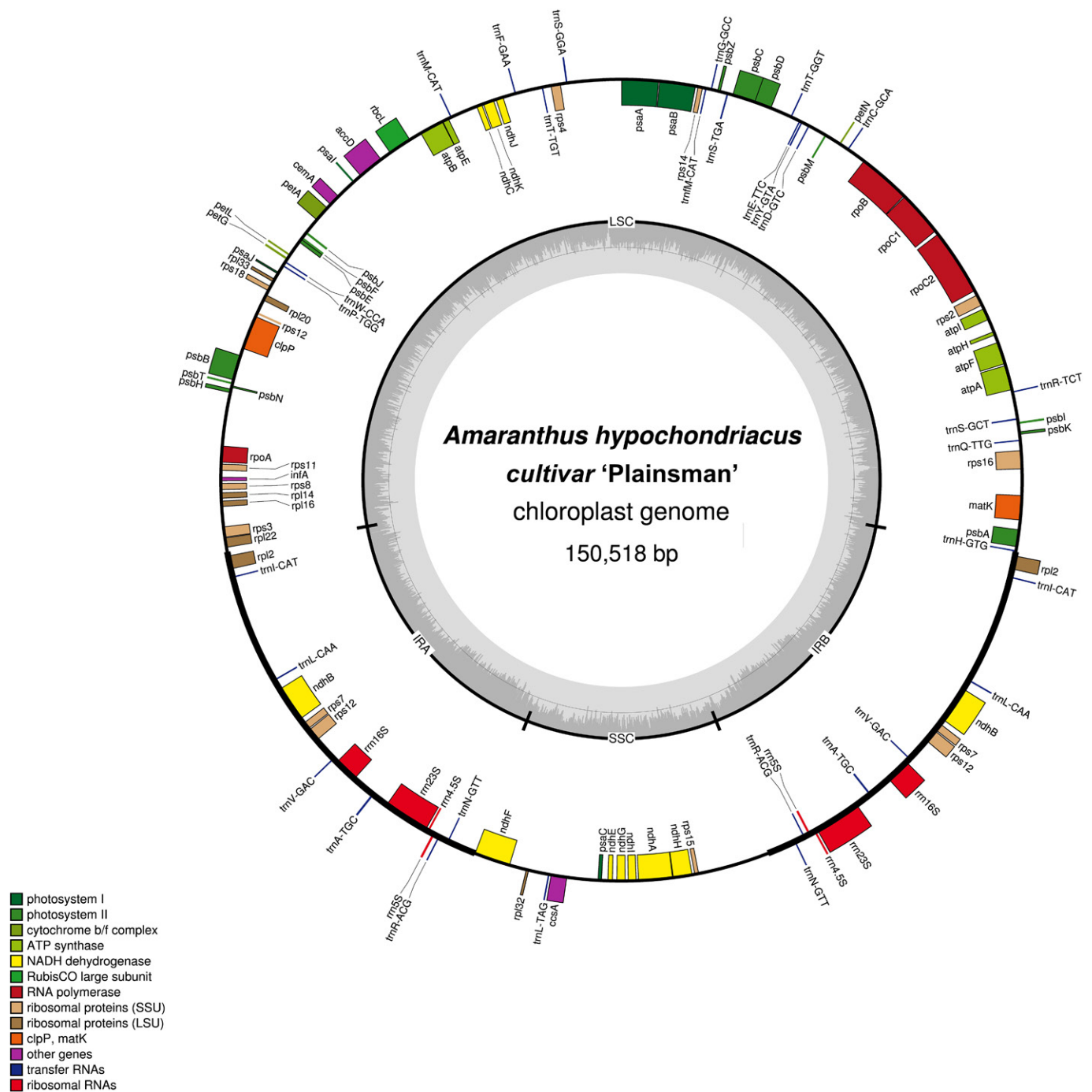


Fig. 1. Chloroplast reference genome of *Amaranthus hypochondriacus*. Genes shown inside the outer circle are transcribed counterclockwise and those outside are transcribed clockwise. Genes belonging to different functional groups are color coded. Gray area in the inner circle indicates the GC content. Figure produced using OGDRAW (Lohse et al., 2013).

LITERATURE CITED

BIRKY, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148.

BRESSANI, R., A. SÁNCHEZ-MARROQUÍN, AND E. MORALES. 1992. Chemical composition of grain amaranth cultivars and effects of processing on their nutritional quality. *Food Reviews International* 8: 23–49.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

HUANG, H., C. SHI, Y. LIU, S.-Y. MAO, AND L.-Z. GAO. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14: 151.

HUANG, Y.-Y., A. J. M. MATZKE, AND M. MATZKE. 2013. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS ONE* 8: e74736.

ITURBIDE, G. A., AND M. GISPERT. 1994. Grain amaranths (*Amaranthus* spp.). In J. E. Hernandez-Bermejo and J. Leon [eds.], *Neglected crops*, 93–101. FAO, Rome, Italy.

TABLE 3. Simple sequence repeats (SSRs) in the eight amaranth chloroplast genomes.

| Species (Genome) | Mono | Di | Tri | Tetra | Penta | Hexa | Compound | Total |
|---------------------------------------|------|----|-----|-------|-------|------|----------|-------|
| <i>A. hypochondriacus</i> (PI 558499) | 15 | 2 | 5 | 11 | 0 | 2 | 2 | 37 |
| <i>A. cruentus</i> (PI 477913) | 11 | 2 | 3 | 11 | 1 | 1 | 0 | 29 |
| <i>A. hypochondriacus</i> (PI 481125) | 11 | 2 | 3 | 11 | 1 | 1 | 0 | 29 |
| <i>A. hypochondriacus</i> (PI 511731) | 11 | 2 | 2 | 11 | 1 | 1 | 0 | 28 |
| <i>A. hypochondriacus</i> (PI 540446) | 11 | 2 | 2 | 12 | 1 | 1 | 0 | 29 |
| <i>A. hybridus</i> (PI 605351) | 11 | 2 | 3 | 11 | 1 | 1 | 0 | 29 |
| <i>A. hypochondriacus</i> (PI 619259) | 11 | 2 | 5 | 11 | 1 | 1 | 1 | 32 |
| <i>A. caudatus</i> (PI 642741) | 13 | 1 | 4 | 11 | 1 | 2 | 1 | 33 |

TABLE 4. Number of repeated sequences by repeat type in the eight amaranth chloroplast genomes.

| Species | Genome | Forward | Palindrome | Total |
|---------------------------|-----------|---------|------------|-------|
| <i>A. hypochondriacus</i> | PI 558499 | 14 | 20 | 34 |
| <i>A. cruentus</i> | PI 477913 | 16 | 21 | 37 |
| <i>A. hypochondriacus</i> | PI 481125 | 15 | 21 | 36 |
| <i>A. hypochondriacus</i> | PI 511731 | 15 | 21 | 36 |
| <i>A. hypochondriacus</i> | PI 540446 | 16 | 21 | 37 |
| <i>A. hybridus</i> | PI 605351 | 15 | 21 | 36 |
| <i>A. hypochondriacus</i> | PI 619259 | 15 | 21 | 36 |
| <i>A. caudatus</i> | PI 642741 | 15 | 21 | 36 |

TABLE 5. Number of single-nucleotide polymorphisms (SNPs) and insertions/deletions (indel) in the seven amaranth chloroplast genomes in comparison to the reference chloroplast genome, *A. hypochondriacus* (PI 558499).

| Polymorphism | <i>A. cruentus</i> (PI 477913) | <i>A. hypochondriacus</i> (PI 481125) | <i>A. hypochondriacus</i> (PI 511731) | <i>A. hypochondriacus</i> (PI 540446) | <i>A. hybridus</i> (PI 605351) | <i>A. hypochondriacus</i> (PI 619259) | <i>A. caudatus</i> (PI 642741) |
|--------------|-----------------------------------|--|--|--|-----------------------------------|--|-----------------------------------|
| SNP | 201 | 201 | 234 | 212 | 200 | 206 | 217 |
| indel | 106 | 127 | 138 | 130 | 123 | 133 | 99 |

JANSEN, R. K., L. A. RAUBESON, J. L. BOORE, C. W. DEPAMPHILIS, T. W. CHUMLEY, R. C. HABERLE, S. K. WYMAN, ET AL. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. In E. A. Zimmer, T. J. White, R. L. Cann, and A. C. Wilson [eds.], *Methods in enzymology*, vol. 224: Molecular evolution: Producing the biochemical data, 348–384. Academic Press, San Diego, California, USA.

KIM, K. E., P. PELUSO, P. BABAYAN, P. J. YEADON, C. YU, W. W. FISHER, C.-S. CHIN, ET AL. 2014. Long-read, whole-genome shotgun sequence data for five model organisms. *Scientific Data* 1: 140045.

KURTZ, S., J. V. CHOUDHURI, E. OHLEBUSCH, C. SCHLEIERMACHER, J. STOYE, AND R. GIEGERICH. 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29: 4633–4642.

LEWIS, S. E., S. M. J. SEARLE, N. HARRIS, M. GIBSON, V. IYER, J. RICHTER, C. WIEL, ET AL. 2002. Apollo: A sequence annotation editor. *Genome Biology* 3: research0082.1.

LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Website <http://arxiv.org/abs/1303.3997> [accessed 31 August 2016].

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNEL, J. RUAN, N. HOMER, G. MARTH, ET AL. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25: 2078–2079.

LI, H., H. CAO, Y.-F. CAI, J.-H. WANG, S.-P. QU, AND X.-Q. HUANG. 2014. The complete chloroplast genome sequence of sugar beet (*Beta vulgaris* ssp. *vulgaris*). *Mitochondrial DNA* 25: 209–211.

LIU, C., L. SHI, Y. ZHU, H. CHEN, J. ZHANG, X. LIN, AND X. GUAN. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13: 715.

LOHSE, M., O. DREHSEL, S. KAHLAU, AND R. BOCK. 2013. OrganelleGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41: doi:10.1093/nar/gkt289.

MALLORY, M. A., R. V. HALL, A. R. MCNABB, D. B. PRATT, E. N. JELLEN, AND P. J. MAUGHAN. 2008. Development and characterization of microsatellite markers for the grain amaranths. *Crop Science* 48: 1098.

MARÉCHAL, A., J.-S. PARENT, F. VÉRONNEAU-LAFORTUNE, A. JOYEUX, B. F. LANG, AND N. BRISSON. 2009. Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* 106: 14693–14698.

MAUGHAN, P. J., S. M. SMITH, D. J. FAIRBANKS, AND E. N. JELLEN. 2011. Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). *Plant Genome* 4: 92–101.

McKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS, A. KERNYTSKY, K. GARIMELLA, ET AL. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.

MILLIGAN, B. G., J. N. HAMPTON, AND J. D. PALMER. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Molecular Biology and Evolution* 6: 355–368.

MOORE, M. J., P. S. SOLTIS, C. D. BELL, J. G. BURLEIGH, AND D. E. SOLTIS. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

PARKS, M., R. CRONN, AND A. LISTON. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.

QIAN, J., J. SONG, H. GAO, Y. ZHU, J. XU, X. PANG, H. YAO, ET AL. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8: e57607.

RAVEN, J. A., AND J. F. ALLEN. 2003. Genomics and chloroplast evolution: What did cyanobacteria do for plants? *Genome Biology* 4: 209.

- SABLOK, G., G. P. RAJU, S. B. MUDUNURI, R. PRABHA, D. P. SINGH, V. BAEV, G. YAHUBYAN, ET AL. 2015. ChloroMitoSSRDB 2.00: More genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database* 2015: bav084.
- SAUER, J. D. 1967. The grain amaranths and their relatives: A revised taxonomic and geographic survey. *Annals of the Missouri Botanical Garden* 54: 103–137.
- SCHMITZ-LINNEWEBER, C., R. M. MAIER, J. P. ALCARAZ, A. COTTET, R. G. HERRMANN, AND R. MACHE. 2001. The plastid chromosome of spinach (*Spinacia oleracea*): Complete nucleotide sequence and gene organization. *Plant Molecular Biology* 45: 307–315.
- THIEL, T., W. MICHALEK, R. K. VARSHNEY, AND A. GRANER. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 10: 411–422.
- TODD, J. J., AND L. O. VODKIN. 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* 8: 687–699.
- TRANEL, P. J., AND F. TRUCCO. 2009. 21st-century weed science: A call for *Amaranthus* genomics. In C. N. Stewart Jr. [ed.], *Weedy and invasive plant genomics*, 53–81. Wiley-Blackwell, Ames, Iowa, USA.